Verbally Assisted Haptic-Graph Comprehension: Multi-Modal Empirical Research Towards a Human Computer Interface

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. nat.
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Universität Hamburg

Eingereicht beim Fachbereich Informatik von

ÖZGE ALAÇAM

aus Ankara (Türkei)

Dezember 2015

Gutachter;

Prof. Dr. Christopher Habel

Prof. Dr. Stefan Kopp

Tag der Disputation: 21.12.2015

Zusammenfassung

Statistische Diagramme, u.a. Liniendiagramme, werden in vielen multimodalen Kommunikationssituationen eingesetzt, sowohl in Wissenschaft und Ausbildung, als auch in der Wirtschaft und den generellen Medien, da durch die Visualisierung von Daten und Information Denken und Problemlösen unterstützt werden kann. Liniengraphen präsentieren primär Datenpunkte. Doch die visuelle Wahrnehmung des Menschen ermöglicht aus der Linie des Graphen weitere Entitäten und Relationen zweiter Ordnung (wie Extremwerte, Trends oder Veränderungen in den Trends) zu erkennen, die auf der den Formeigenschaften des Graphlinie beruhen.

Um Blinden oder stark Sehbehinderten einen Zugang zu graphischen Darstellungen zu ermöglichen, ist die sensorische Substitution mithilfe der haptischen und auditiven Wahrnehmung ein vielversprechender Ansatz. Haptische Präsentationen von Graphen sind ein geeigneter Ersatz zu visuellen Diagrammen, um diese Kenntnisse als Ersatz, um visuelle Diagramme zu erwerben, insbesondere, wenn sie in Hybrid-Systemen integriert sind. Allerdings bringt die Gestaltung eines effizienten Umfeldes für diesen Zweck ihre eigenen Anforderungen mit. Aufgrund von modalitätsbedingten Unterschieden zwischen der visuellen und der haptischen Wahrnehmung ist eine Äquivalenz der aus den Graphen ableitbaren Informationen nicht erreichbar. Der Fokus dieser Dissertation liegt darauf, sehbehinderten Benutzern während der haptischen Wahrnehmung eines Graphen zusätzlich verbale Unterstützung als begleitende Modalität zu bieten mit dem Ziel, die Lücke zwischen visuellem und haptischem Graphverstehen zu schließen und kohärente Information bereitzustellen. Das verbal unterstützte Verstehen von haptischen Graphen ist eine aufgabenorientierte gemeinsame Aktivität zwischen zwei Agenten: Einem Menschen, der den Graphen exploriert, d.h. der die Grafik aktiv mittels der haptischen Modalität untersucht, und einem verbale Unterstützung leistenden Agenten, der diesen Prozess für den Entdecker auf einfache Art und Weise erleichtert. Die erfolgreiche Konzipierung eines solchen Systems kann nur aus einer interdisziplinäre Perspektive erreicht werden, die mehrere Mainstream-Forschungsfelder kombiniert, und zwar Graphverstehen, Ereignissegmentierung, Produktion von referentiellen Ausdrücken durch Menschen und automatisierte Systeme, Dynamik innerhalb einer gemeinsamen kooperativen Aktivität und HCI-Aspekte für Effizienz und Effektivität des Systems.

Zu diesem Zweck wurde eine Reihe von empirischen Studien zur visuellen und zur haptischen Wahrnehmung von Graphen, zum Verstehen dieser Graphen und zur Kommunikation über Graphen durchgeführt. Das Verstehen haptischer Graphen und die Affordanzen der kollaborativen Umgebung wurden sowohl durch Einzelbenutzer-Experimente als auch durch Zweipersonen-Experimente (Benutzer + menschlicher Kooperationspartner) unter verschiedenen Forschungsfragen untersucht. Eine systematische Untersuchung der Wechselwirkung zwischen sensorischen, repräsentationellen und kommunikativen Modalitäten beim Graphverstehen spielt die zentrale Rolle für die Realisierung eines automatischen verbalen Assistenzsystems, das erfolgreiche Unterstützung bei der haptischen Erkundung und beim Verstehen von Graphen bereitstellt. To that end, various methods including the analyses of linguistic data, speech-accompanying gestures, sketches, haptic exploration movements and post-exploration questionnaires were utilized.

Die wichtigsten empirischen Resultate sind:

- Ohne verbale Unterstützung, hatten die Benutzer von haptischen Graphen häufig Schwierigkeiten beim Graphverstehen aufgrund unvollständigen Wissens in ihren Graph-Schemata, die eine Schlüsselrolle bei der Schaffung einer geeigneter Zuordnung zwischen den graphischen Funktionen und der begrifflichen Ereignis spielt.

- Amodale geometrischen Eigenschaften (wie Form, Größe, Ausrichtung etc.) haben eine starke signifikante Wirkung auf die Event Segmentierung und Beschreibung.
- Die sensorische Modalität der Graphwahrnehmung hat einen Einfluss auf die Produktion von Gesten.
- Eigeninitiative der Explorierenden bei der Anforderung von Assistenz und eine angemessene durch sprachlichen Modifikatoren angereicherte verbale Unterstützung, durch die Assistenten, erweist sich als hilfreiche Kombination für eine gemeinsame Aktivität, die erfolgreiches Graphverstehen ermöglicht.
- Haptische Explorationsmuster (die Geschwindigkeit, die Anzahl der Hin-und-Her-Aktionen auf der Graphlinie, die qualitativen Angaben tragen etc.) können verwendet werden, um den Bedarf von Unterstützung zu erkennen.
- Abgleich auf der Situationsebene und Wahl eines geeigneten Referenzrahmens in Bezug auf das bestehende kommunikative Ziel sind von entscheidender Bedeutung.
- Die Analyse von Gesten hat sich HCI-Forschungsparadigma als sehr effizientes Werkzeug bewiesen, unter anderem bei der Auflösung von Mehrdeutigkeiten in verbalen Daten.

Das aus der Gesamtheit der empirischen Untersuchungen gewonnene und integrierte Wissen konnte für die Entwicklung von Design-Richtlinien genutzt werden, spezifisch für verbal unterstütztes Verstehen haptischer Graphen und im Allgemeinen, für den Entwurf formbasierter haptischer Darstellungen und für Anforderungen an Konstellationen verbaler Assistenz.

Abstract

Statistical graphs, such as line graphs, are widely used elements of multimodal communicational settings as it is the case with the news media, economy bulletins or educational settings due to their facilitating influence on crucial cognitive processes, such as reasoning, problem solving and memory. At the bottom line, they present data points. However by means of human pattern perception processing (i.e. if some Gestalt principles are satisfied), the continuous line of the graphs allows easy extraction of second order entities and relations (such as *extreme values*, *trends*, or *changes in trends*), which are closely linked to shape properties of the graphs.

A promising approach to provide blind people with access to external representations is sensory substitution using haptic and audio modalities. Haptic presentations of graphs provide a suitable means to acquire such knowledge as a substitute to visual graphs, in particular when they are integrated in hybrid systems. However designing efficient environment for this purpose has its own challenges. Due to perceptual differences between visual and haptic modalities, informational or functional inequivalences are inevitable. This dissertation focuses on providing visually impaired users with a verbal assistance as an accompanying modality to their haptic graph exploration with the aim of bridging this gap and presenting coherent information. The verbally assisted haptic graph comprehension is a task oriented joint activity between two agents; a human explorer who perceives the graph haptically through active exploration and a verbal assistance agent that provides helps to the explorers in a simple but facilitating manner. Successful design of such system can be achieved only with an interdisciplinary perspective that incorporate several mainstream research fields such as “graph comprehension”, “event segmentation”, “referring expression production by humans and generation by automated systems”, “dynamics of a joint activity” and “HCI aspects for efficient and effective system design”.

For this purpose, a series of cognitive science oriented empirical studies concerning perception, comprehension and communicational aspects was performed on both visual and haptic modalities. Haptic graph comprehension and the affordances of the collaborative environment were investigated through single-user experimental paradigms as well as through human-human joint activity settings focusing on various research questions. Such multi-modal interface involves various sensory and communicational modalities. A systematic investigation of the interaction between them plays an important role in the realization of the system. To that end, various methods including the analyses of linguistic data, speech-accompanying gestures, sketches, haptic exploration movements and post-exploration questionnaires were utilized.

In brief, the empirical findings indicated that

- Without assistance, the haptic explorers have difficulties in conceptualization of the graphs due to having incomplete knowledge in their graph schemata, which has a key role in establishing appropriate mapping between the graphical features and the conceptual event.
- Amodal geometric properties (such as shape, size, orientation etc.) have a strong significant effect on event segmentation and description.
- The sensory modality of graph reading has an effect on gesture production.
- Taking initiative in requesting help and having adequate verbal assistance enriched by modifiers in a response, seems a superb combination for a successful joint activity that enhances graph comprehension.

- Haptic exploration patterns (the speed, the number of back-and-forth actions on the graph line, the qualitative ascriptions etc.) can be used to detect assistance need in an instantaneous and automatic manner.
- Coming into alignment at the situation level and the appropriate choice of frame of reference with respect to the communicative goal at hand is crucial.
- The analysis of gestures as a HCI research paradigm has been proved to be very efficient tool, i.e. in resolving ambiguities in verbal data.
- The integrated knowledge obtained from all empirical investigations was used to constitute design guidelines, in specific for verbally assisted haptic graph comprehension and in general for shape based haptic representations and verbal assistance constellations.

Thesis Related Publication List

2015

Alaçam, Ö., Habel, C. and Acartürk, C. 2015. Switching Reference-frame Preferences during Verbally Assisted Haptic-graph Comprehension. In the Proceedings of 6th International Conference on Spatial Cognition (*Cognitive Processing*: DOI :10.1007/s10339-015-0730-9)

Alaçam, Ö., Acartürk, C. and Habel, C. 2015. Haptic exploration patterns in virtual line-graph comprehension. In A. Marcus (Ed.): Design, User Experience and Usability (DUXU/HCII 2015), Berlin: Springer-Verlag.

2014

Alaçam, Ö., Acartürk, C. and Habel, C. (2014). Referring Expressions in Discourse about Haptic Line Graphs . In Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue. (pp. 7-17). SemDial 2014 – DialWatt. Verena Rieser & Phillipe Muller (eds.)

Acartürk, C., Alaçam, Ö., and Habel, C. (2014). Developing a Verbal Assistance System for Line Graph Comprehension. In A. Marcus (Ed.): Design, User Experience and Usability (DUXU/HCII 2014), Part II, LNCS 8518. (pp. 373–382). Berlin: Springer-Verlag.

Kerzel, M., Alaçam, Ö., Habel, C., and Acartürk, C. (2014). Producing Verbal Descriptions for Haptic Line-Graph Explorations. In Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue. (pp. 205-208). SemDial 2014 – DialWatt. Verena Rieser & Phillipe Muller (eds.)

Alaçam, Ö., Habel, C. and Acartürk, C. (2014). Verbally Assisted Haptic Graph Comprehension: The Role of Taking Initiative in a Joint Activity. In the Proceedings of the 2st European Symposium on Multimodal Communication, University of Tartu, Estonia, August 6-8, 2014.

Alaçam, Ö., Habel, C. and Acartürk, C. (2014). Perspective Alignment during the Course of Verbally Assisted Haptic Graph Comprehension. In the online Proceedings of the RefNet Workshop 2014 on Psychological and Computational Models of Reference Comprehension and Production. Edinburgh, UK. 31 August 2014.

2013

Alaçam, Ö., Habel, C., and Acartürk, C. (2013) Investigation of Haptic Line-Graph Comprehension Through Co-Production of Gesture and Language. Tilburg Gesture Research Meeting, Tilburg

Alaçam, Ö., Habel, C., and Acartürk, C. (2013) Towards designing audio assistance for comprehending haptic graphs: A Multimodal Perspective. In C. Stephanidis and M. Antona (Eds.), Universal Access in Human Computer Interaction. Heidelberg: Springer

Habel, C., Alaçam, Ö., and Acartürk, C. (2013). Verbally assisted comprehension of haptic line graphs: referring expressions in a collaborative activity. In Proceedings of the CogSci 2013 Workshop on Production of Referring Expressions, Berlin.

Alaçam, Ö., Habel, C., and Acartürk, C. (2013). Gesture and Language Production in Communication through Bar Graphs. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th Annual Conference of the

Cognitive Science Society, Austin, TX. pp. 1714–1719. Cognitive Science Society

2007-2012

Acartürk, C., and Alaçam, Ö. (2012). Gestures in Communication through Line Graphs. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society, Austin, TX. pp. 66–71. Cognitive Science Society.

Alaçam, Ö., and Acartürk, C. (2012) Work in Progress: Annotating Gestures in Communication through Graphs. Proceedings of the Graduate Symposium of 7th International Conference on the Theory and Application of Diagrams, Canterbury, UK. pp. 5–7.

Alaçam Ö., Hohenberger A., and Çağıltay, K. (2010). Effects of Graph Type in the Comprehension of Cyclic Events. In A. Goel, M. Jamnik, N. Narayanan (Eds.), Diagrammatic Representation and Inference. Lecture Notes in Computer Science 6170 (pp. 279-281): Springer Berlin / Heidelberg. Presented at Graduate Symposium and Poster Session at Diagram 2010, Portland, USA.

Alaçam Ö. (2010). Comparison Strategies of Different Event Types in Different Graph Types (MS. Thesis – The Department of Cognitive Science, Informatics Institute, Middle East Technical University, Ankara)

Alaçam, Ö., Hohenberger, A. E. and Çağıltay, K. (2009). Comparison Strategies of Different Event Types in Different Graph Types. In *Proceedings of 15th European Conference on Eye Movements, p.94* , Southampton, UK, 2009. University of Southampton.

Acartürk,C., Habel, C., Cagiltay, K., and Alaçam, Ö. (2008). Multimodal comprehension of language and graphics: Graphs with and without annotations. *Journal of Eye Movement Research*, 1(3):2, 1-15. Also presented as a poster at ECEM 2007, 14th European Conference on Eye Movements,19-23 2007, Potsdam, Germany.

Acartürk, C., Habel, C., Cagiltay, K., & Alaçam, Ö. (2008). Learning from text and graphs: The role of annotations and sensory modality. In A. Maes & S. Ainsworth (Eds.), *Proceedings EARLI Special Interest Group Text and Graphics. Exploiting the Opportunities Learning with Textual, Graphical and Multimodal Representations* (pp. 13-16).

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Christopher Habel for the continuous support, motivation, encouragement and immense knowledge. The countless inspiring discussions with him enlightened me and profoundly changed the way how I think.

I would also like to thank Cengiz Acartürk for his encouragement and insightful comments that greatly assisted the research.

I am grateful to Stephan Kopp for accepting to review this thesis and for his helpful comments and questions.

I would especially like to thank Sibel Toprak for her valuable comments during the final stages of the publication. I give special thanks to Matthias Kerzel for the technical support with the haptic representations. I would like to thank the research students Leena Suresh, Gregor Gawellek, Lena Andreeßen and Neele Stoeckler for their valuable effort.

The work reported here has been financially supported by the German Science Foundation (DFG) IRTG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS). I thank Jianwei Zhang, who is the spokesperson of CINACS.

I would like to show my gratitude to my colleagues in the Knowledge and Language Processing Group (WSV) and in the Natural Language System Groups (NATS) for sharing their knowledge and for the detailed support. Carola Eschenbach deserves grateful thanks for her valuable comments. I would like to thank Hildegard Westermann, Wolfgang Menzel, Melania Duma, Mirela-Stafania Duma, Junlei Yu, Felix Lindner, Timo Baumann, Christine and Arne Köhn, and Sebastian Beschke. I thank Tatjana Lu Tetsis and Wiebke Eggers for all the support with the administrative tasks.

Special mention goes to my old colleagues from METU Computer Center. I am thankful for their help and support during the time that I performed empirical studies at the METU HCI Research and Application Laboratory.

I would be remiss without mentioning my friends who were always there for me, gave me the motivation and the distraction that I deeply appreciate; Nergis Ayşe Gürel Köybaşı, Berrin Doğusoy, Demet Kırbulut, Selda Eren Kanat, Yusuf Şafak Bayram, Zuhal Kavacık, Dilşah Cırıl, Onur Kerimoğlu, Sezgi Sözen and Annalisa Morticelli.

Finally, but by no means least, I would like to thank my family for showing belief in me and supporting me throughout writing this thesis and my life in general. I dedicate this thesis to them.

To each of the above, I extend my deepest appreciation.

CONTENTS

SECTION – I:

The State of the Art of "Multimodal Communication and Graph Comprehension"

SECTION - II

Haptic Graph Comprehension

SECTION – III
Verbally Assisted Haptic Graph Comprehension System

SECTION – I

The State of the Art of

"Multimodal Communication and

Graph Comprehension"

Introduction

Purpose of the Study

Presenting and representing information in visuo-spatial formats, such as graphs, maps or diagrams, is important as well as successful, for thinking, problem solving and communication. Statistical graphs, such as line graphs, are widely used in multimodal communicational settings, as it is the case with the news media and economy bulletins. These are also crucial elements in learning environments: for example, they enhance the understanding of mathematical information and highlight the comprehension of relational data from natural and social sciences (e.g. Hegarty, 2011; Shah and Hoeffner, 2002) because they have a facilitating influence on crucial cognitive processes underlying the learning process. Graphs are mostly not used alone but in combination with language and also with gestures, forming multimodal communication settings. In particular, this is the case in documents (e.g. web pages) and in agent-to-agent communicational settings. Graphs can be used for the extraction and comparison of information. As for statistical graphs, they do not only present data, they also allow for the extraction of second order entities; these could be extreme values, trends, or changes in trends. In particular, the properties of the line shape make it possible to distinguish a global maximum from a set of local maxima, or to detect inflection points depicting trend changes. Moreover, graphs can be used to extract information that may not even be explicitly contained in them (i.e. predicting the near-future based on the trend).

Acquiring such knowledge is crucial also for blind or visually impaired people. One of the most common ways of making graphs accessible to them is to provide a summarized verbal description of the graph message via screen text readers (e.g. Ferres, Lindgaard, Sumegi and Tsuji, 2013). On the other hand, haptic substitution has been a successful mode of sensory substitution as well (e.g. Moll and Psyander, 2013; Oliveira, Quek, Cowan and Fang, 2012). However, due to perceptual differences, informational or functional (computational) inequivalence are inevitable. For instance, the haptic modality has a lower bandwidth compared to the visual one (Loomis, Klatzky and Lederman, 1991). Besides, haptic perception is sequential and local. Meanwhile, both local and global information can be extracted from a graph at one glance with visual perception. Visual representations can therefore be considered as superior in the amount of conveyed information.

There have been continuous efforts to include blind and visually impaired people, where these visio-spatial interfaces are used. In order to achieve this, to bridge the gap caused by the above-mentioned perceptual differences and to present coherent information to haptic graph readers, haptic graphs should be accompanied by alternative modalities. The existing research and methods to that end are elaborated in Chapter 2. In short, the multimodal interfaces that incorporate haptic and audio interfaces that provide perceptual access to graphical representations seem to constitute an effective tool (Yu and Brewster, 2003; Zhao, Plaisant, Shneiderman and Lazar, 2008; Abu Doush, Pontelli, Simon, Son and Ma, 2010). Over the past decade, haptic-audio interfaces have been developed to provide perceptual access to spatial representations to the visually impaired and, thus, to facilitate their comprehension of spatial displays. There still remains much need for both further development of specific types of haptic-spatial interfaces and research that focuses on peculiar characteristics of the interface design. For example, in haptic line graphs, the local and sequential character of perception is based on exploration processes, i.e., hand movements following the graph-line for the purpose of gathering information about its geometrical properties. Users can, for example, explore haptic graphs by hand movements (manually) following graph lines engraved in a (real) physical plane (Figure 1.b) or by using a force-feedback device like the Phantom Omni® (recently renamed as GeoMagic® TouchTM, see Figure 0-1.c), to explore virtual graph lines.

(a) (b) (c)

Figure 0-1 (a) exploration of a physical haptic map, (b) Phantom Omni device, and (c) visualization in a geometry domain (see, Kerzel and Habel, 2013, Fig.1)

This dissertation focuses on providing visually impaired users with a verbal assistance as an accompanying modality to their haptic graph exploration. The design of the proposed verbal assistance system, involves two main components, namely one component responsible for providing verbal-assistance (by the system) and another one for the active haptic exploration by the explorer. While the active exploration component allows the user to discover the embodied graph shape and its details (such as concavities and convexities, as well as maxima and minima), the verbal assistance component helps the user to label those explored regions. The latter has critical benefits for both online comprehension and recall for later use (this issue is discussed in detail in Section-II).

In classical natural-language generation systems, the task of providing information is evaluated under three aspects; "*what to say*", "*how to say*" and "*when to say*". From the HCI perspective this task is known as "saying the right thing at the right time in the right way" (Fischer, 2001). The "*what to say*" task is also referred to as the *conceptualization* task (Levelt, 1989; Habel and Tappe, 1999). In this stage, the content of the utterances to be produced by the system is determined. The "*how to say*" task, also called as the *formulation* task, comprises the steps regarding sentence planning, surface generation, morphology and formatting. In a nutshell, the surface form of some content is constructed from semantic representations obtained in the previous step (Reiter, 1994; Reiter and Dale, 2000; van Deemter, Krahmer, and Theune, 2005). Apart from deciding which information ("*what to say*") should be provided in which way ("*how to say*"), the timing of the provided assistance ("*when to say*") is another important issue that needs to be touched upon for developing a successful interactive system that reacts automatically to the user's action (e.g. Lohmann, 2012).

Such an assistive system involves many aspects that need to be handled thoroughly; such as the decisions regarding verbal content, the dynamics of the interaction between the user and the system and also the technical issues concerning natural language generation and the design of the haptic graphs. In this dissertation, my main focus is on the first two aspects. For this purpose, empirical studies are needed to understand the principles underlying the conceptualization of graphs, the communication about graphs and the haptic exploration of graphs. What is especially challenging about designing haptic graphs: It is necessary to determine which concepts depicted by a graph—or by segments thereof—are appreciated as important. This challenge becomes all the more significant when designing haptic line graphs that involve indistinct perceptual representations for conceptually distinct entities. An example for this is a graph with several local minima with values that are close to each other, as opposed to a simple graph line with a distinct global minimum, see the illustrations of two graphs that exhibit different characteristics of global minimum in Figure 0-2. Figure 0-2(a) has two local minima and to distinguish the global minimum is easy task for both visual and haptic graph comprehender (both terms "*reader*" for visual comprehender and "*explorer*" for haptic comprehender are used to define the person who perceive and comprehend the graph). Figure 0-2(b) exhibits challenging situation, which is described above. Human perceptual visual apparatus provides appropriate tools to overcome this problem, on the other hand for a haptic explorer who explores the graph with sequential exploration, deciding which one is global minimum is almost impossible. Within the scope of this dissertation, it is proposed that verbal assistance may facilitate overcoming these kinds of problems by providing necessary information through the auditory channel.

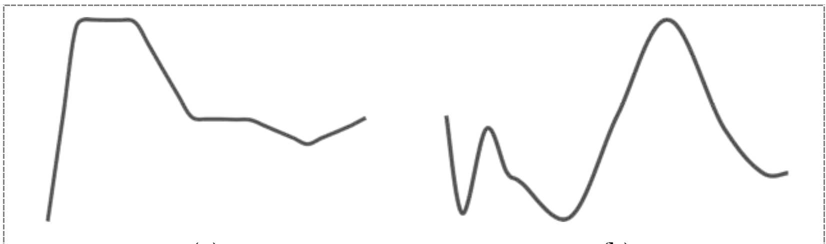


Figure 0-2 The illustration of the challenges in haptic graph comprehension due to sequential perception

In the investigation of these abovementioned issues, two experimental paradigms that employ a single-user haptic exploration and a haptic exploration in joint activity (dual-user) were conducted. The single-user haptic exploration paradigm (Paradigm-I) is used to investigate (i) modality-dependent characteristics of haptic graph comprehension, (ii) the effect of graph shape (the saliency of the shape landmarks or segments etc.) and (iii) the effect of linguistically coded content in the comprehension of second order entities, such as general and temporally restricted trends based on the recognition of global and local curvature landmarks. Secondly, an experimental setting, in which two participants perform joint activity for graph exploration, thus performing verbally assistant haptic graph exploration (Paradigm-II) was employed in order to investigate (i) the circumstances under which verbal assistance facilitates haptic comprehension of graphs and (ii) the underlying aspects of successful communication, Throughout the dissertation, the terms "interlocutors" or "partners" are used interchangeably for the same meaning.

My approach for studying these issues is to incorporate ideas from Human Computer Interaction and Cognitive Science, since I believe neglecting cognitive processes in human computer interaction design would lead to an impractical and ineffective system design. As depicted in Figure 0-3, a successful system design can be achieved by taking verbally assisted haptic graph exploration as a task-oriented collaborative activity (Clark, 1996) between two partners: a (visually-impaired) explorer (*E*) of a haptic graph and an observing assistant (*A*) providing verbal assistance. The latter could be a human or automated system.

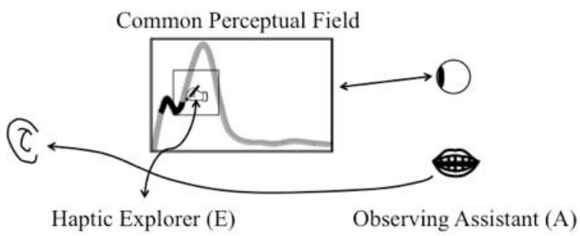


Figure 0-3 Assisted haptic graph exploration, a joint activity

Successful communication through graphs and language usually requires the integration of information contributed by both graphical entities and verbal entities so that the graph reader arrives at integrated conceptual and spatial representations. *A* and *E* share a common field of perception, namely the line graph, but their perception and comprehension processes differ significantly. In addition to having different sensory abilities, the participants have different activity roles (explorer vs. assistant).

The success of the task-oriented joint activity of the explorer the assistant depend on the alignment of the interlocutor's internal models, especially on building common ground (Garrod and Pickering, 2004). *E*'s internal model of the activity space results from haptic and motor sensations, whereas *A*'s internal model of the same space is built up by visual perception. Therefore similarities and differences in their conceptualization play a central role in aligning at the situation-model level. This topic is discussed in detail in Chapter 2.3 and Chapter 9.

To be really considered assistive, *A* should verbally provide E with content which is difficult to acquire haptically. This—haptically difficult to acquire—content has to be coordinated in a timely manner with the haptically explored content in the same sentence (or phrase) to fulfill the given-new contract (Clark and Haviland, 1977). All in all, for successful and non-redundant communication, the verbal assistance system is expected to provide most helpful and relevant information for haptic explorer at that particular moment, and this content should be selected among all-possible information that can be derived from the representation considering haptic explorer´s previous actions on the graph and previous utterances. The motivation that underlies this expectation is that the content of the verbal assistance has the potential to influence the alignment process, thus leading to a better or worse comprehension of the haptic graphs.

A systematic investigation of the interaction between modalities in communication through graphs plays an important role in the realization of an automatic verbal assistance system that provides instantaneous support for haptic explorers. This investigation has the potential to contribute to identifying design principles to achieve efficiency and effectivity in multimodal communication settings. In this dissertation, various methods were utilized in order to get insights in human conceptualization of haptic graphs: (1) an analysis of referring expression production (during the joint activity as part of the either communication or post-exploration verbal descriptions of the graphs), (2) a combined analysis of verbal descriptions and speech-accompanying gestures and (3) a combined analysis of sketches and linguistic content produced as a part of the dialogue between the explorers and the assistants, (4) the haptic exploration movements of the explorer, and (5) post-exploration questionnaires[1].

Research Questions

To sum up, I am looking for answers for numerous research questions that can be categorized under two topics: First, how does a graph explorer conceptualize the abstract events presented through haptic modality? In the investigation of this topic, (i) the effect

[1] The eye movements of the visual graph readers were also recorded, but this analysis is beyond the scope of this dissertation.

[2] such as event segmentation and description, dynamics of joint activity, design guidelines for HCI

of the sensory modalities, (ii) the effect of the presence of data labels in visual graphs and (iii) the effect of amodal geometric properties on event conceptualization, segmentation and description were explored in a series of experiments (Section-II). These investigations also provide systematic analyses to understand which information is easy to grasp or hard to encode. Furthermore, incorporating different modalities with the haptic modality makes new research questions arise: What kind of content should be made accessible through the haptic modality? What kind of content should be communicated using language (speech) or sounds (sonification)? One of the main claims in this dissertation is that verbal assistance systems may facilitate graph comprehension processes. Here, the specific focus lies on verbal assistance provided for the shape of graph line.

The second topic is comprised of questions concerning the suitability of a verbal assistance system for collaborative activity. For the investigation of this topic, the dynamics of joint activity, such as the selection of a reference frame, the role of taking initiative, and the alignment between the interlocutors were examined (Section-III).

The findings obtained from these investigations are evaluated and discussed from a theoretical as well as an applied perspective.

Furthermore, although both recognition and generation of gestures constitute a very commonly investigated topic from a human-computer interaction perspective (specifically, human – robot or human-virtual agent perspective), their use as a method in the investigation of conceptualization, is not prevalent in the existing literature. These investigations are also useful to answer the question of whether gesture analysis is a useful and influential method in the domain of HCI.

Contribution of the study

To my knowledge, there are only very few instances of haptic assistive systems that take the shape properties of graphical representations into account in the design process (e.g. see Ferres et al., 2013; Wu, Carberry, Elzer and Chester, 2010). There is also still a lack of research on the role of shape comprehension in haptic graph exploration. Furthermore, the available literature on event segmentation studies is focused on segmentation with respect to the visual or auditory modality, leaving out the haptic one.

Quite recently, considerable attention has been paid to haptic studies, since haptic devices are becoming widespread: They are not only becoming relevant for visually impaired community, but also for use in multi-modal educational design, game technologies, and medical operations (some important sub-domains are the following: touch enabled computer applications, skill training, collaborative human-robot interaction, haptic control systems). Although the research here was not conducted to address this line of research, the results are also applicable to them. Current state-of-art haptic graph systems would benefit from this research with regard to extensive multimodal behavioral data analysis as well as the intertwined approach to investigating multiple research domains[2] that have importance in graph comprehension domain. There is a huge amount of literature about each of these research areas. However, only a few studies exist regarding the haptic modality. The investigation of the haptic modality from the perspectives of both HCI and theoretical research may be uniquely useful and may even open up new research areas.

Moreover, the contribution of gesture research to HCI design combined with the results coming from speech-accompanying gesture analysis also provides valuable insights about the relations among gesture, language and space that could be discussed from a theoretical perspective.

[2] such as event segmentation and description, dynamics of joint activity, design guidelines for HCI systems for visually impaired people and gesture-language-space relations.

Organization of the Dissertation

This dissertation consists of three main sections that elaborate on the issues mentioned above. The first section presents the "state of the art" of multi-modal communication, with special focus on haptic graphs. Congruency of the information (the event represented by the graph) to its carrier (i.e. graph) is the fundamental research topic, which has been investigated for many decades. Chapter-1 will present general concepts and existing research focusing on this issue. Chapter-2 aims to introduce the overview of verbal assistance systems designed for visually impaired people. The use of gesture analysis as a research paradigm in HCI design is discussed in Chapter-3. The final chapter of Section-I (Chapter-4) aims to give a coherent overview of the experimental paradigms, the details about the experiments, and the basic coding schemes employed in this dissertation.

Section-II is dedicated to the topic of haptic graph comprehension, mostly taking only the comprehender (the graph reader) perspective into account (leaving the role and the contribution of the assistance system aside for now). The first chapter of this section, Chapter-5, presents fundamental principles of haptic exploration, the differences between two sensory modalities (*visual* and *haptic*) and it investigates how the perceptual and conceptual factors influence the haptic graph comprehension. Apart from the differences, also the similarities between the different modalities are elaborated on, with evidence provided from the empirical evaluations. As mentioned previously, haptic exploration of a graph invites explorers to focus on the shape of the graph. Therefore anchoring hard-to access abstract representations to concrete entities is possible through shape. Chapter-6 will focus on the effect of the shape properties (of the graph line) on the segmentation of events into sub-events and also on referring to those events. In Chapter-7, the empirical evidence presented in all of Section-II will be discussed. The final chapter of this section is dedicated to the interpretation of the results concerning the gesture modality from both a HCI-oriented applied research and a theoretical perspective, by focusing on the possible contributions of this study into gesture-language research.

Section-III focuses on haptic comprehension as a joint activity between two partners (the explorer and the observing assistant). The first chapter (Chapter-9) discusses some key elements of successful communication, which include the alignment of internal mental representations between the partners at different range of levels, the role of taking initiative for requesting help, and how information should be presented. The detection of the time when information is needed is discussed in Chapter-9 as well. The next chapter (Chapter-10) of this section provides heuristics and design guidelines for how the results obtained in these experiments can be implemented and how they can be useful in the design of effective and efficient verbally assisted system for visually-impaired people. In the final chapter (Chapter-11), I conclude with overall discussion and discuss several short-term and long-term future studies.

CHAPTER 1

(Multi-Modal) Graphical Communication

1.1 Graphs and Graph Properties

Due to their efficiency in presenting and allowing the comprehension of quantitative information and relations, graphs are widespread among non-graphic specialists and in most professional areas (Trickett and Trafton, 2006; Fischer, 2000; Kosslyn, 1989; Larkin and Simon, 1987). The primary goal of visualizing data is to (re-)present them in a format more suitable for thinking, problem solving and communication. This view is taken implicitly or explicitly in most seminal publications on graphs and on visualization in general over the last few decades (see, e.g., Hegarty, 2011; Kosslyn, 2006, 1989; Tufte, 1983). Graphs constitute a successful means to present data in a way that is suited for the task of analyzing the data and the task of communicating data analysis results. Communicating visualized data using line graphs is used extensively in scientific publications, textbooks, magazines and news-papers; Zacks, Levy, Tversky, and Schiano's (2002) study on the use of graphs in the print media shows that line graphs are one of the most frequently used types of graphs in addressing non-experts. In addition to text-graphics documents, graphs, spoken language, and often gestures, accompany each other forming multimodal communication in many professional communication as well as classroom settings.

Graphs can be used to extract a single piece of information from the graph, or to compare two or more pieces of information. The primary gain in using graphs, however, is not to make individual data points visible, as tables would successfully carry this information as a textual format. Tables enable us to have access vast amount of quantitative data presented in a structured way. However, when a reader aims to look for the relations that require integration of several values, accessing this knowledge from tables[3] would be effortful compared to graphs, which can provide easy visual access to relations between data points ('x1-y1 has a larger y-value than x2-y2') and second-order entities, such as *trends*, *local* and *global maxima*. Second order properties (e.g., *strength of an upward trend*) and second-order relations (*crossing of lines*) can be also easily detected because of the preattentional processes taking place in the human visual system (Habel and Acartürk, 2012). For instance, the properties of the line shape allow distinguishing global points (maximum and minimum) from a set of local points (maxima and minima), or detecting inflection points that depict trend changes. This advantage can be ascribed to pattern perception processes in the human brain, such as visual chunking

[3] It should be noted that tables exhibit particular disadvantage when it comes to haptic/tactile representation. Due to sequential characteristics of the modality, finding relevant piece of information, keeping it in mind, then searching for other relevant information one by one is a challenging and thus cognitively difficult procedure.

(see Shah, Mayer and Hegarty, 1999). Line graphs enhance the detection of such concepts, which are closely linked to shape properties of the graphs. Besides, according to Landau and Jackendoff (1993), the shape is the dominating criteria in labeling shape-based representations and this holds also for the line graphs, in which the shape is the dominant saliency attribute for identifying entities using referring expressions. Furthermore, comprehension of haptic line graphs is based on exploration processes, i.e. hand movements following the line, with the goal to collect information provided by the geometrical properties of the line explored. In particular, shape properties are detected first, and they are utilized in anchoring hard-to access abstract representations to concrete graph entities. Therefore, the shapes of line representations, to be more precise the amodal geometric shape properties of the line graph, are considered as a key concept for haptic graph comprehension and also for providing multimodal assistance that accompanies to haptic exploration.

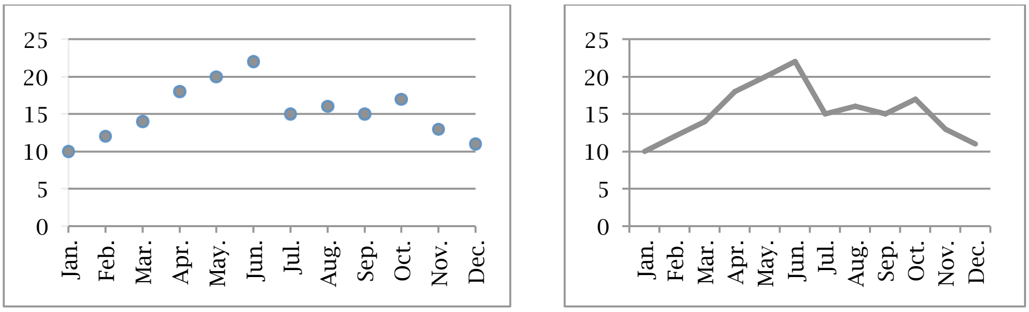
In order to represent information for the different aims given above, different graph designs (i.e. line graph, bar chart or pie chart) are being used in general. Different graph designs differ in emphasizing specific details about the data (Renshaw, Finlay, Tyfa and Ward, 2004; Ratwani and Trafton, 2008). According to Pinker (1990), "different types of graphs are not easier or more difficult in general, but are easier or more difficult depending on the particular class of information that is to be extracted". As stated by Kosslyn (1994, p.271) "a good graph forces the reader to see the information the designer wanted to convey". Lots of studies, whose common findings on graph perception and operations indicate that different graph types are better suited for different judgment and problem-solving tasks, support Pinker's view (Casner, 1991; Cleveland, 1990, 1985; Gillan and Lewis, 1994; Hollands and Spence, 1992; Lohse, 1993; Shah and Carpenter, 1995; Simkin and Hastie, 1987; Tan and Bensbasat, 1990, 1993; Hollands and Spence, 1998). In other words, successful interpretation of the information depicted in the graph is heavily dependent on the selection of correct design (Renshaw et. al, 2004).

Graphs are considered as one type of symbolic diagrams, which are widely used means of communication in scientific and technical areas. Graphs and charts are more formalized visualization with language-like conventions compared to other types of pictorial diagrams such as maps, plans and so on (Schmidt-Weigand, 2006). Typically, a statistical graph consists of a co-ordinate system with two main axes and their respective labels. Then there is also a data region, which contains graphical components (such as points, lines or bars) that form a spatio-temporal representation (Fischer, 2000). By these properties, graphs have syntactic, semantic and pragmatic levels like language (Kosslyn, 1989; Schnotz, 2002) and they are useful for performing linguistic analyses on those levels (Hegarty, Carpenter and Just, 1991). Kosslyn states that a syntactic analysis focuses on the properties of the lines and regions themselves; they are not interpreted in terms of what they convey or refer to. On the other hand, the semantic analysis focuses on the meanings of the configurations of lines and what they demonstrate (e.g. axes labels, etc.). The semantic analysis can be considered as the literal reading of each of the components of a chart or graph and the literal meaning that arises from the relationship between these components. And finally, he defines the pragmatic analysis as finding the meaningful symbols conveying information above and beyond their direct semantic interpretation. In addition, pragmatic considerations govern the relationship between the information in a display and the readers' purposes and needs. Like textual representations, graphs also express relations and properties of objects (Gurr, 1999). For example, Zacks and Tversky (1999) showed that when subjects see bar graphs, they describe discrete contrasts (like exact values, maximum points, and higher/ lower/ greater than/ less than-relationships) in the data; when they see line graphs, they describe trends (rising, falling, increasing, decreasing) (also see Fischer, Dewulf and Hill, 2005; Meyer, Shinar and Leiser, 1997; Zacks and Tversky, 1999; Kosslyn, 1993; Levy, Zacks, Tversky and Shiano, 1996). For example, if the graph reader's intention in reading a graph involves information integration operations, then the performance is better when the graph design incorporates features that maximize the integration and extraction of information. This is achieved through providing suitable and compatible perceptual organization (by employing *Gestalt principles* i.e. spatial proximity, similarity of color, shape and size) with the intended message of the graph. Wickens and Carswell (1995) suggest that that such design principles promote parallel processing and/or assist in the viewer's information integration task. This reduces demands on working memory and, subsequently, enhances task performance. On the other hand, if the task requires that individual entities be processed separately, then their arrangement should be best organized to allow this through perceptual separation. This difference in the interpretation of information from different graphs, seem to be based on principles of cognitive "naturalness" which is supported by the Gestalt principles that underlie figural perception. According to these principles, bars are convenient for conveying categorical information and lines are convenient for conveying ordinal or interval data. In bar graph design, each value is represented with a separate bar, and this helps readers separate entities or categories. Meanwhile, values are connected by a single line indicating that to all the values belonging to the same entity in line graph design (Ali and Peebles, 2013; Trickett and Trafton, 2006; Zacks and Tversky, 1999; Lohse, 1993; Pinker, 1990).

In spite of the bar graph's advantageous on discrete comparison tasks (Shah and Hoeffner, 2002; Zacks and Tversky, 1999), comparing two separate dots and reasoning about, for example about higher/lower relation, would be problematic with the haptic representations of bar graphs. In the haptic modality, the connectedness represented by lines also helps to extract relational inferences (e.g. "lower" or "higher"). There is a bunch of studies that investigate whether Gestalt principles are applicable to the haptic modality (Frings and Spence, 2013; Gallace and Spence, 2011). In a nutshell, it seems that the principles of proximity, similarity, common fate, good continuation and closure have also an influence on tactile perception. However, the influence might exhibit itself differently in different modalities: As reported by Frings and Spence (2013) for instance, the Gestalt "*principle of proximity*" may differ between visual and haptic modalities. Their results illustrate that if two dots in the haptic space are too close to each other, they might be encoded as a single entity, instead of being recognized as two entities of one group. However, these studies focus on the exploration through finger and hand (involving perception through both cutaneous and kinesthetic receptors). The indirect perception through a stylus (employing only kinesthetic perception) may exhibit further differences. But it is safe to assume that Gestalt principles are not working in exactly same manner in the two modalities.

A standard starting point for generating graphs are tables or, their computer-science counterpart, relational data-bases (Mackinley, 1986). But, exclusively visualizing data points is suitable only in certain cases. For example, Figure 1-1a depicts average daily maximal temperature at some city "X" in a form of data point graph. From the perspective of graph comprehension, statistical line graphs are different from graphs of mathematical functions in that the line (proper) serves different purposes. In particular, in a function line graph, the line represents value mappings between the two axes, whereas in a statistical line graph, the line usually serves the purpose of creating a visual continuum. Therefore the lines are usually not veridical representations of the data. Under specific conditions (e.g., if some Gestalt principles are satisfied), human visual processing leads, pre-attentively, to the visual impression of a linear whole, namely a line. Figure 1-1(b) depicts a line graph that relieves the perceptual and cognitive load by making the line explicit. These characteristics make line graphs an appropriate means of reasoning and communication about events represented by them. Given their capabilities of representing events in temporal order, line graphs constitute the most frequently used type of statistical graphs (Zacks et al., 2002).

The contrast between data-point graphs and line graphs exemplifies how substantially the human perceptual system determines the comprehension of data visualizations. Blind and visually impaired people might have more critical problems in exploring data-point graphs haptically (as depicted in Figure 1-1) using a Phantom Omni device, since Gestalt constitution is not supported in this case.



(a) (b)

Figure 1-1 Averaged daily maximal temperature distribution represented by a (a) data-point graph and a (b) line graph

From the huge amount of research on graph comprehension, it can be concluded that both the graph type and the conceptual domain conveyed by the graph message affect readers' descriptions. A graph designer/producer may have a specific message to deliver or the reader may have a specific intention for looking at the graph. In either case, the graph needs to carry the message well. In other words, graph comprehension processes are based on an interaction of bottom-up and top-down (the selection of task-relevant information) activation of cognitive schemata (e.g. Shah and Freedman, 2011; Hegarty, 2005; Schnotz, 2002). To that end, the graph and the represented concept should be in line with each other, and this alignment should be taken into consideration by the verbal assistant system.

Furthermore, in order to compare the comprehensibility of two or more different visual representations - graph designs or structures - their informational and computational equivalence should be also considered (Larkin and Simon, 1987; Palmer, 1978; Schnotz, 2002). Being *informationally equivalent* corresponds to having the same information content expressed in two different representations. For example, if the textual and visual representations of a statement "X is bigger than Y" have the same variables and same relations, then they are considered informationally equivalent. Additionally, in order to be called *computationally equivalent*, the effort for retrieving information from each of the informationally equivalent representations should be about the same. For the given example above, the statement represented by the graph should be retrieved as easily as the same statement from the textual representation.

Considering the state-of-the-art with respect to haptic graphs, providing haptic data labels is a design challenge. That is why this issue is one of the reasons for not having informationally isomorphic representations. Within the scope of this thesis, the effect of data labels are investigated by comparing visual graphs with and without them (see Figure 1-2 (a-b), systematically controlling informational difference between the two graph types. On the other hand, the effect of the sensory modality regarding the computational equivalence was investigated by comparing visual and haptic graphs without data labels (see Figure 1-2 (b-c). In the latter case, no data labels were provided to keep the information carried by both modalities similar.

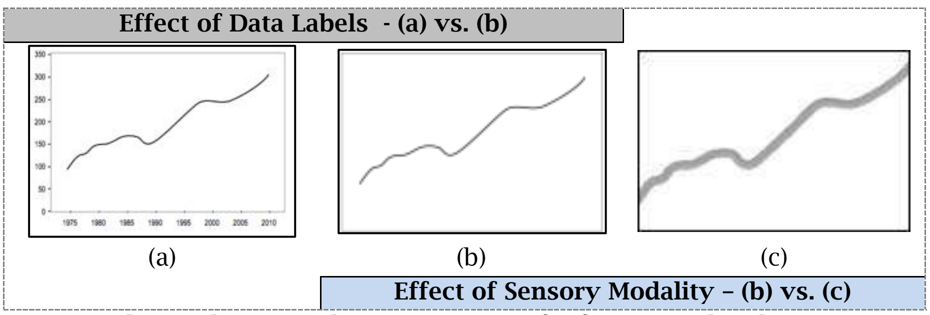


Figure 1-2 The conditions in the investigation of informational and computational equivalence

1.2 Event Representation and Representational Modalities

Graphs are used to represent abstract concepts, which unfold over time (such as change in the bird population or change in temperature). In other words, they depict events. They transform these abstract concepts into static concrete objects by employing graphical entities such as graph shape, axes, data labels etc. And through (simultaneous) visual perception we can obtain information about an event represented by the graph similar to perception of a static object. On the other hand, although haptic graphs are also static objects that represent same abstract event, they are perceived in a sequential manner through active exploration. Therefore perception of haptic graphs resembles to event perception. Regarding haptic perception, we can refer to events in two different layers and they should be clearly distinguished: (i) the conceptual event layer (the event represented by the graph such as the distribution of the bird population over the years) and (ii) event-like perception of haptic graph shape. The matters of event segmentation and event description are elaborated on in Chapter 6.

Comprehending graphs and communicating about them requires the involvement of various sensory (representational) and communicational modalities. Sensory modalities are employed to perceive the graph that depicts conceptual and abstract event in a concrete way. Graph perception and comprehension are mostly dependent on representational (sensory) modalities, such as visual, haptic and verbal. On the other hand, communicating over graph involves communicational modalities such as language, gestures, etc. For example, imagine a speaker verbally describing a population of a bird species to her audience, while producing deictic and iconic gestures to point to the corresponding graphical entities on a graph. For the audience, the resulting representation makes use of multiple sensory channels that carry language, gesture and graphical information. Therefore as well as the characteristics of the abstract event and the perceptual differences, the characteristics of the representational (*visual and haptic graphs, language*) and communicational modalities (*language and gesture*) also need to be touched upon. The modalities can be classified into two categories; *analog/continuous* representations and *digital/symbolic/discrete* representations. Pictorial (diagrammatic or depictive) representations are forms of "*analog representations*", which display features in a continuous scale. The visual and haptic graphs can be considered in this category. The speech-accompanying gestures also display analog properties. On the other hand, in *symbolic representations*, objects are represented by arbitrary symbols, i.e. prepositional representations in language are considered as one type of symbolic representations that exhibit categorical property (Palmer, 1978). Focusing on the similarities and differences in the characteristics of the representations are important since they give insights about the affordances of each modality. For instance, as analog representations, gestures and graphical representations can convey the aspects of size and manner in a richer way compared to language modality.

1.3 Graph Perception and Graph Comprehension Theories

There is a large bulk of literature about visual graph perception and comprehension. Most studies address both issues since they are considerably intertwined. According to Cleveland (1985), several perceptual features (see Table 1-1) are involved in the perception of graphs (e.g. length, angle, area) and they differ in terms of their accuracy and reliability. The graph reader chooses the first property available from this set. Cleveland also proposed a hierarchy of these perceptual features based on their accuracy levels, which is shown below (Table 1-1). However, according to Hollands and Spence (2001), observers sample from the set of available perceptual features rather than choosing the most effective one and using it consistently. During the perceptual process, the graph reader may choose between the perceptual clues or features, however, the available perceptual features may vary with the graph designs. Aside from availability, the effective usage of the available features may also vary among graphs (Hollands and Spence, 2001).

Table 1-1 Hierarchy of perceptual features for discriminating proportions in graphs (Cleveland, 1985)

Accuracy	***Perceptual feature***
Most accurate	Position along a common scale
	Position along identical, non-aligned scales
	Length
⇕	Slope / Angle
	Area
	Volume
Less Accurate	Color hue –saturation - density

However, in order to be able choose most effective features, being able to ignore irrelevant cues plays an important role (e.g. Shah and Hoeffner, 2002). For instance, while the area is important, length information is irrelevant to read a pie chart. Additionally, to read a bar graph, depth information or closeness between two consecutive x-labels may be irrelevant. The graph reader picks up relevant information and ignores the irrelevant ones depending on the graph type they are exploring for the sake of appropriate comprehension.

Although the perceptual properties of graphs are crucial in the conceptualization, comprehension is still highly dependent on the conceptual properties (Zacks and Tversky, 1999), which may be hard to grasp via the haptic modality. Another factor that affects graph comprehension are working memory (short-term memory) and long-term memory constraints. Kosslyn (1989) stated that since the working memory has a limited capacity, this constraint affects our ability to integrate syntactic information and to keep the semantic information in mind during graph comprehension. Thus, the complexity of a graph will be a major factor in determining its comprehensibility. Through visual perception, the whole graph is grasped even before the individual parts are focused on consciously (Wagemans, Elder, Kubovy, Palmer, Peterson, Singh and von der Heydt, 2012). In contrast, for haptic perception, the local parts are perceived first in a sequential manner and then integrated during exploration process. Kalia and Sinha (2011) claimed that integrating local samples to global shape information for haptically explored shapes may be one of the most problematic issues because of the memory constraints. Moreover, long-term memory has also some major constraints, most importantly the person's domain knowledge. The way a graph is interpreted at both the semantic and the pragmatic level depends on which stored information is most closely associated with how the stimulus properties of a graph are categorized.

From many theories of graph comprehension in literature; here I focus on the three of them briefly since they are the most comprehensive regarding the coverage of graph types, performance predictions, and explanations for how people extract information from a graph (Trickett, Ratwani and Trafton, 2006). Trickett et al.'s study provides a thorough investigation in a well-organized format on the effects of the task to be performed (or the intention) on graph comprehension and whether the current general theories account for all tasks. These theories are Freedman and Shah's "*Construction Integration Theory*" (Freedman and Shah, 2002; Shah and Shellhammer, 1999; Shah, Freedman and Vekiri, 2000; Shah, 2002), Pinker's "*Propositional Model*" (Pinker, 1990), and Lohse's "*UCIE model*" (Lohse, 1993). Freedman and Shah base their interpretation of graph comprehension processes on the construction-integration model of text comprehension (Kintsch, 1988). During the construction phase, the graph reader notices the visual features of the graph, and alternately reads the graph and the legend. During the integration phase, the visual features from the construction phase are comprehended using prior knowledge about the graph and domain knowledge. Domain knowledge refers to any mental representation of the content of the graph (Freedman and Shah, 2002). Domain knowledge serves the purpose of making numerical relationships more apparent and facilitating inferential processes. Readers who lack domain and/or graph knowledge will be less accurate in their interpretations and will produce only surface level descriptions of the graph (Freedman and Shah, 2002). Freedman and Shah assume that working memory is limited, and as a result, graph readers alternate between the construction and integration phase until the information is comprehended. According to Pinker's *Propositional Model*, a graph reader first scans the graph (the scan patterns are addressed in this model), and constructs a visual array based on its perceptual properties. The graph reader then forms a propositional representation of the graph from this visual array. The appropriate graph schema, which allows the reader to create a conceptual question (the information the graph reader wants to extract from the graph), is triggered by the propositional representation. Four main processes operate on the propositional representation of the graph; (1) a matching process, which allows the reader to recognize a graph as being of a particular type, (2) a message assembly process that translates the visual information into conceptual information, (3) an interrogation process which is used when needed information is not present from the message assembly, and (4) inferential processes like being able to perform mathematical operations from the context (Pinker, 1990). Furthermore, Lohse's *UCIE* (Understanding Cognitive Information Engineering) theory aims at making quantitative predictions about how long it will take a graph reader to extract specific information from a graph. It contains the following operations: comparing two units in memory (digits, colors, words, shapes, forms), interpolating on a linear scale, making saccades, and making a perceptual judgment (Lohse, 1993). When interpreting a graph, Lohse suggests that (1) Early visual processes detect and encode visual features such as shape and color, (2) Short-term memory (STM) builds a visual description from the early visual processes (3) Information in the STM triggers an association to a memory trace in long term-memory (LTM) which instantiates the graph schema. The graph schema directs the interpretation of the graph.

Those graph perception and comprehension studies might be also beneficial in order to understand underlying mechanisms of haptic graph comprehension. Therefore, I will try to highlight a few issues starting from possible differences between vision and haptics regarding some features. For the line graph representation, the line connects data points and invites people to focus on trends. The value for each data point is *the first order information*, whereas the slope that represents a trend becomes *the second order information*. Then the turning angle in the junction of two line segments, which is dependent on the slope values of the two line segments can be considered *as third order information*. The abovementioned use of this terminology (*the first/second/third order changes*) exemplifies the distinctions on the different properties concerning the static graph shape from the perspective of the graph visualization. A kindred use of this terminology also occurs in the domain of event segmentation. According to Thibadeau's second order hypothesis (1986), second-order changes are important for detecting psychological event boundaries. This hypothesis suggests that the segmentation of an event is influenced by the changes in the first order changes (namely *second order changes*) rather than by just first order changes. From the event perception point of view, for example, the velocity of an object (in this case, it may correspond to velocity of the haptic explorer's action on the graph line) is *the first order change* whereas the acceleration in the action corresponds to *the second order change*. This issue is elaborated on in Chapter 6.2.2.

All these properties can be considered as amodal geometric properties regarding haptic and visual modalities, since they can be accessed from both. To sum up, slope and angle are the two main features in the comprehension of line graphs. Despite the fact that these two parameters are well investigated in the visual-graph domain, their effect in haptic graph comprehension still lacks research. Moreover, the investigation of the effect of irrelevant features on graph comprehension may also be useful in understanding the underlying mechanisms of haptic graph comprehension. It should be noted that length constitutes one of the problematic issues when it comes to line graphs. Length is also an amodal feature. However, due to perceptual differences, the interpretation of this particular feature may exhibit differences. Figure 1-3 aims to illustrate the possible misreadings of the length. The line segment's projection on the x-axis corresponds to a temporal interval. The length information does not interfere with reasoning about the comparison of the sub-events represented by the segments (l1) and (l3). Since they have the same slope value (*no slope*), we can use the length information of (l3) and (l1) to claim that the third sub-event took longer than the first one. But in other cases, using length to reason about time leads to misunderstandings. That is why graph readers need to ignore length information when comparing two sub-events like (l2) and (l3). Sub-event-2 lasted for one time unit, while sub-event-3 took 5-time units. However, this property may interfere seriously with temporal judgments of the events explored through sequential haptic movements. This is touched upon in more detail in upcoming sections. Since controlling the length parameter in line graph design is practically not reasonable, length information was left aside.

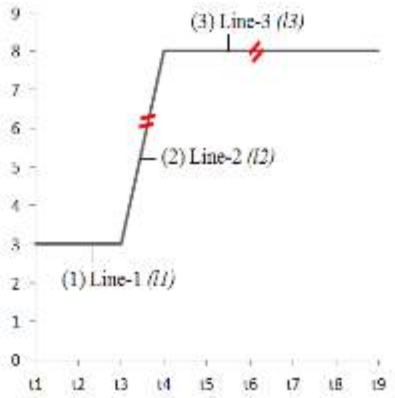


Figure 1-3 An illustration of length feature on a line graph

Another feature that needs to be ignored during haptic graph exploration is friction caused by the *force-feedback mechanism*. A study by Fiehler, Burke, Engel, Blen and Rösler (2007) showed that pressure and friction perceived via a force feedback stylus has an effect on the perception of the engraved line patterns, highlighting the difficulty of experimentally isolating kinesthesia.

Existing graph comprehension theories concentrate on visual representation of graph (as expected), and they mostly consider visual images and visual features. However, graph comprehension inherently requires reasoning. The latest studies (e.g. Knauff, 2013 for a comprehensive review) claim that reasoning is done via spatial representation rather than visual image as it was claimed before. Therefore, it may not be extraordinary to assume that if both haptic and visual graph comprehension is based on spatial representations then they may exhibit similarities when it comes to reasoning and communicating over graphs.

1.4 Sensory Substitution and Equivalence between Modalities

The advantages of graphs as a source of information are not directly accessible for blind and visually impaired people. A promising approach to provide blind people with access to external representations is sensory substitution. In text reading, both the haptic modality (Braille) and the auditory modality (using text-to-speech systems) are successfully used to substitute the visual modality. In contrast, pure haptic line graphs have emerged as inferior to visual graphs. While visual perception supports comprehension processes which switch between global and local aspects of a graphical representation, haptic perception has a more local and, in particular, a more sequential character. Thus, compared to visual graphs, one drawback of haptic graphs is the inherent restriction of the haptic sense regarding the possibility of simultaneous perception of information (Loomis and Klatzky, 1991).

Ferres et al. (2013) provided a categorization of existing assistance methods used to facilitate haptic graph comprehension for visually impaired people. A simplistic system may be consisting of haptic-only interfaces as presented in several studies (Wall and Brewster, 2006; Yu and Brewster, 2002; Yu, Ramloll and Brewster, 2001; Riedel, 2001). The deficits of haptic-only exploration have led to the development of multimodal graphs, which employ two sensory modalities simultaneously, namely, haptic and audio. A more advanced way is to load some information to the auditory modality by sonification (Alonso-Arevalo, Shelley, Hermes, Hollowood, Pettitt, Sharples and Kohlrausch, 2012; Cohen, Yu, Meacham, and Skaff, 2005; Brown and Brewster, 2003; Ben-Tal, Berger, Cook, Daniels, Scavone and Cook, 2002; Ramloll, Yu, Brewster, Riedel, Burton and Dimigen, 2000). In sonified graphs for example, statistical data is represented by the pitch of the sound, e.g. a tall bar produces a high-pitched sound, while a short bar produces a low-pitched sound (Yu and Brewster, 2003). Sonified graphs have demonstrated facilitating effects in the perception of haptic graphs of certain types. But given the complexity of cognitive processes involved in the comprehension of line graphs, a lot more assistance is necessary than sonified graphs can provide. Thus, haptic interfaces can be also accompanied with natural language interfaces (Ferres et al., 2013, 2010; Giudice, Palani, Brenner and Kramer, 2012; Demir et al., 2010; Abu-Doush et al. 2009; Kurze, 1995). The next chapter will address verbal assistance systems in more detail.

Haptics is commonly defined as a perceptual system that incorporates inputs from two subsystems; from *cutaneous receptors* in the skin and from *kinesthetic receptors* embedded in muscles, joints, and tendons (Loomis and Lederman, 1986). Cutaneous receptors are distributed all over the body surface and play a role in the perception of varying object properties such as temperature, softness, texture etc. On the other hand, kinesthetic input from mechanoreceptors, that involves active exploration, contributes to human perception of limb position and movement in space and is also involved in sensing, perceiving and thinking about objects, their properties and the object's environment (Lederman and Klatzky, 2009). Furthermore, perceiving by touch includes both passive (being touched passively) and active perception (active and voluntarily done exploration). A perceiver may directly touch the object or surface by using his/her one-finger, two-fingers, one hand, or two hands. However, it is also possible to perceive the object or surface via apparatus, in other words by using an intermediate tool (e.g., a pencil or a stylus) that allows exploring it. In this case, haptic perception is considered remote or indirect, and the information is delivered through the force-feedback mechanism of the device in the form of vibrations or friction (Lederman and Klatzky, 2009). Within the scope of this dissertation, like previously mentioned, I employed the Phantom Omni force-feedback device with a stylus to perceive haptic graphs, all of which are represented as an engraved line on a virtual plane, The graph readers were invited to explore the graph actively with the stylus. Therefore, although going into the underlying haptic sensory mechanism in detail would take us too far afield, I should note that haptic exploration for line graphs in this study requires remote haptic exploration guided by kinesthetic receptors and the shape perception involves the contribution of kinesthetic inputs.

Lederman and Klatzky (1987) introduced a set of six manual exploratory procedures regarding tactile/haptic object perception and established a systematic relationship between those exploratory procedures and object properties. In short, these procedures are (1) lateral motion (associated with texture), (2) unsupported holding (for weight), (3) pressure (for hardness), (4) static contact (for temperature), (5) enclosure (for global but coarse shape and volume) and finally (6) contour following (for global and exact shape). Using the Phantom Omni, the graph line is provided as an engraved concavity. Thus, the line and its shape are foregrounded. The explorer is then invited to trace it in a line-following manner. Among all these exploratory procedure, only the contour following procedure can be applied to the virtual haptic line graphs, therefore the haptic actions mentioned in the rest of this dissertation only refer to this procedure.

Loomis, Lederman and their colleagues (Giudice, Betty and Loomis, 2011; Loomis and Klatzky, 2008; Loomis, Klatzky, Avraamides, Lippa and Golledge, 2007; Loomis, Klatzky and Lederman, 1991; Lederman and Klatzky, 1987; Loomis and Lederman, 1986) investigated how similar abstract spatial representations resulting from different input modalities (vision, language, auditory, touch) are. Loomis and Klatzky (2008) touched upon the idea that the perceptual representations created by vision, hearing and touch may give rise to amodal spatial representations; this will be elaborated on more in the upcoming subsection 1.5. While showing many similarities, the study of Loomis and Klatzky (2008) indicated that vision, hearing and touch vary in terms of encoding, in terms of their precision of how they represent space. When the sensory bandwidth of vision was matched to that of touch by artificially limiting the spatial resolution and or field of view, the visual and tactile modalities were found to be functionally equivalent or nearly so with respect to pattern perception. This indicates that the two modalities are similar regarding their higher-levels of processing. For object recognition, it was shown that that when participants are forced to use only one object property (namely shape) for recognition, the recognition time for haptic objects is longer than that for visual objects (Loomis and Klatzky, 2008; Klatzky et al., 1993). A study conducted by Casey and Newell (2004) showed that equating the modalities resulted in a considerable difference in the encoding duration. Accordingly, 4 minutes of touch can be the functional equivalent of 1second of vision.

Visual representations can successfully carry precise metrical information such as distances and directions between objects. On the other hand, these relations are carried over categorical labels via language and therefore they usually have to be inferred (Loomis et al., 2007). However, due to common underlying representations, namely, amodal spatial representations, the abstract representations constructed from different modalities and stored in memory, function similarly. Taylor and Tversky (1992) showed that people can create very accurate spatial representations for environments described linguistically. When participants are induced to create spatial images from language, there is considerable functional similarity between the spatial images from visual perception and language (Loomis et al., 2007). These findings are promising for us to suggest that verbal assistance can be a very powerful method to close the functional gap between the visual and haptic modalities.

In light of these studies, haptic sensory substitution appears to be very appropriate for providing access to graphs since it allows active exploration of the line shape. However, a haptic-only system (providing information via a single modality) would suffer from perceptual disadvantages and might have problems in achieving functional equivalence. These limitations of single-modality exploration led to the development of multimodal graphs, which are able to simultaneously represent two sensory modalities (i.e., haptic and audio). Verbal assistance provided during active haptic exploration has proved to be useful in various types of diagrammatic representations, such as “you are here” maps (e.g. Lohmann, 2012).

1.5 Spatial Representations and Reasoning over Graphs

When reading a sentence or when exploring a graph, internal mental representations of the content presented by these stimuli are constructed first. The construction of this mental representation is dependent on the given task and on the salient features of graphs or sentences, which can make the extraction either easier or more difficult (Schnotz, 2002).

It has been suggested that the information coming from spatial senses arrive at a common region in the brain (at the striate cortex and the posterior parietal cortex, see Cohen and Andersen, 2004). This information gives rise to amodal spatial representations (Loomis and Klatzky, 2008) and is involved in spatial processing. There are numerous theories that suggest that spatial representations are crucial for communication about objects, graphs in this case, (“Representational Modularity

Hypothesis" Jackendoff, 2002) and for inference (e.g. Jonhson-Laird, 1983, 2001, 2006, 2010, 2013; Knauff, 2013; Lacey, Campbell and Sathian, 2007; Bryant, 1997). Briefly, Bryant (1997) proposed that humans possess a spatial representational system (SRS), which enables them to create modal representations of space. According to Bryant, linguistic and perceptual input is first analyzed by modality-specific systems before the information is directed to the SRS, which operates to represent it in a format that is neither perceptual nor linguistic - the resulting spatial representation is modality-independent. As proposed by Johnson-Laird and his colleagues, a mental model is an integrated representation of the information presented via representational modalities, namely a representation of objects and relations (Johnson-Laird, 2006, 2010; Goodwin and Johnson-Laird, 2005). According to Tversky's definition (1991), spatial mental models contain spatial properties presented in representational modality (or modalities), but they are neither visual nor propositional. Furthermore, spatial relations are categorical relations among objects; they allow perspective-taking and spatial reasoning (Tversky, 1993). According to "Space to Reason Theory" recently proposed by Knauf (2013), people construct non-metrical and qualitative spatial representations. Reasoning occurs over these amodal representations, which convey only relevant information for inference, rather than visual images, which can be precise about metrical information. He also claims that these representations are more abstract than visual images and more concrete than propositional representations (Knauff, 2013, p.16). They capture spatial relations from various sensory modalities such as language, visual, auditory or tactile perception and represent necessary and relevant information in a categorical, and amodal representational format. Therefore it has been suggested that they are not physically equivalent to what they represent. In other words, spatial relations in spatial representations are discrete and ordinal relations, which only contain relevant information for the reasoning task at hand. This continuous-discrete transformation (similar to the analog/symbolic differences in the representation format mentioned earlier in *1.2*), is one of the highlighted issues in Knauff's definition. Reason to spatial theory proposes that visual distractors effect only visual reasoning, while spatial distractors impair both visual and spatial reasoning. This leads to the conclusion that the role of spatial representations is more prominent in reasoning. In order to make reasoning easier and faster, spatial aspects should be available and foregrounded. While there are differences in the use of terminology, all these terms "mental model", "spatial mental model" and "spatial layout model" correspond to integrated amodal representations, and the distinction between them is not within the scope of this dissertation.

The distinction between visual and spatial representations has been in the focus of cognitive science for decades. According to the widely accepted working memory models (Baddeley, Grant, Wight, and Thomson, 1973; Logie, 1986; Baddeley, 2003), the visuospatial sketchpad (VSSP), one of the two slave systems of working memory, is responsible for integrating visual and spatial information into a unified representation. There also seems to be evidence that the integration of kinesthetic information may be also performed by this component (Baddeley, 2003; Kaas, Stoeckel, and Goebel, 2008). It has been claimed that the VSSP consists of two sub-components; one for processing visual information and another for processing spatial information. All spatial information seems to be processed in the latter component regardless of its modality. Therefore, it can be assumed that both visual and haptic graph comprehension are working based on the same underlying cognitive mechanism despite their perceptual differences. Similarly, Kosslyn and his colleagues emphasized the role of the two subsystems in regards to the processing of spatial information (Kosslyn, Koenig, Barrett, Cave, Tang, and Gabrieli, 1989; Kosslyn, 1994). One of them is claimed to be responsible for processing quantitative-metrical information with respect to a continuous coordinate system, whereas the other processes qualitative categorical information that conveys coarse spatial relations between objects and discrete spatial concepts.

The results of functional equivalence studies are also in favor of this idea (Loomis and Klatzky, 2008; Loomis, Klatzky, Avraamides, Lippa and Golledge, 2007). In more specific for haptic modality, Loomis and Klatzky's research (2008, p.172) proposed "modality-specific encoding gives rise to the visual, haptic or auditory representations or the linguistic meaning of the utterance. Through some additional conversion process from those representations, spatial images are created (one should note that the latter research do not aim going into details of the cognitive components involved).

Graphs consist of both modality-dependent and independent features. For example, two different properties can be represented by two different colors (i.e. a bar graph that represents two-variables), with different frictions or with different sonified tones. Acquiring these modality-dependent features (color, friction and tone) through another modality is not possible. On the other hand, statistical line graphs are spatio-temporal representations and they are based on spatial features. Hence, they employ a wide variety of amodal geometric properties, such as local and global shape consisting of curvature, length, orientation etc. and that modality-independent information can be accessed by vision and touch and well described by spatial language as well.

A computational approach from the perspective of information processing to these amodal representations was presented by Tschander, Schmidtke, Habel, Eschenbach and Kulik (2003) and Eschenbach, Tschander, Habel and Kulik (2000). In this framework, they are called as *integrated conceptual and spatial representations*. The conceptual implementation of this approach in the graph comprehension domain was discussed in Habel and Acartürk (2009), see also Acartürk (2010). In this implementation, the graph comprehension module is based on Pinker's graph comprehension theory (1990) introduced earlier. Integrated conceptual representations, which are accessible by both modalities, are constructed from the modality-specific processing of linguistic entities (by a language comprehension module) and graphical entities (by a graph comprehension module), both of which form a graph-text constellation.

Together, haptic graphs exhibit different features than visual graphs, since first they are presented in different representational modalities (visual versus haptic) and also the way of perceptions (simultaneous versus active and sequential perception) exhibit differences. Due to these differences, haptic graph comprehension might differ from visual graph comprehension. Detecting these differences is important to close the informational gap originating from these perceptual dissimilarities. On the other hand, as densely elaborated on in various sub-domains of spatial cognition, the spatial information captured through all sensory modalities is processed by the same component in the working memory. Graphs are spatio-temporal representations that convey spatial relations regardless of their modality, and the reasoning and communication over graphs rely on the same amodal spatial representations. Therefore similarities in regard to graph comprehension are also to be expected. The chapters, which are presented under Section-II, are dedicated to the investigation of these issues.

CHAPTER 2

Towards Verbally Assisted Graphs

There are various ways of providing assistance for haptic graphs; verbal assistance is one of the methods that have been gaining more attention. As already introduced in the previous chapter, functional equivalence between language, vision and haptics can be achieved and spatial representations can be successfully built based on verbal descriptions of these representations. Furthermore, language is already a commonly used accompanying modality for visual graphs since it can be effectively utilized to highlight the important aspects by using annotations or by summarizing the message of the graph. By this means, the gap between the visual modality and the haptic modality, due to the perceptional sequentiality inherent in the latter, may be successfully minimized. Even this multi-modal representation may outperform the vision-only cases with the contribution of verbal assistance, as in the case of verbally annotated graphs (Habel and Acartürk, 2011; Acartürk, Habel, Çağıltay and Alaçam, 2007. In addition, the contribution of language as an accompanying modality may also have a facilitating effect on the memory in that remembering the trends for later use is made easier through naming (Kaas, Stoeckel and Goebel, 2008; Millar, 1999, 1978).

This chapter first summarizes different approaches employed to provide visually impaired people with verbal assistance for access to graphical representations (2.1). Then I discuss crucial issues related to the production of verbal descriptions for graphs (2.2). Afterwards, I introduce a verbal assistance system for haptic graphs (2.3), and finally, the assistive system is elaborated on from the perspective of human- computer interaction (2.4).

2.1. Different Approaches for Verbally Assistance Systems

Verbal assistance can be provided in different ways. One approach is to give a full summarization of a graph at once without requiring any interaction with users. The summarization of each graph can be pre-determined or it can be derived from a graph automatically. Wu et al. (2010) proposed an automatic natural language generation method on inferring the intended message of a line graph based on a graph's visually salient features. First, they start by automatically segmenting the line graph into visually distinguishable trends. The output of this segmentation step is used as a basis for suggesting possible messages out of 10 predetermined messages by using a Bayesian Network and accordingly a full summarization is produced by combining the suggested descriptions for the segmented parts. On the other hand, providing assistance specifically tailored to the user's actions may fulfill the user's needs more effectively. One of the promising examples of an assistance system was introduced by Ferres and his colleagues (2013). The iGraph is an interactive advanced graph navigation system. Graph summarization is provided by text-to-speech, and users can have access to information at different levels of granularity depending on his/her need.

Furthermore, although the method of segmentation based on visually salient points is practically useful and successful in most of the cases, it should be noted that event segmentation is driven by both conceptual and perceptual effects (Chapter 6 provides comprehensive overview on this topic). The design of graphical representations is inherently intended to highlight conceptual saliency through the use of visual saliency, in line with the Gestalt principles in the first place. As such, a graph type that carries the right perceptual features are chosen in accordance to the type of conveyed message (bars for discrete comparisons, lines for trend descriptions). Therefore, it can be assumed that perceptually (visually) salient points would very likely correspond to conceptually salient points. However, here two issues need to be taken into careful consideration: First, haptic perception may hinder some features, which have higher weight for visual perception, and facilitate other irrelevant ones as already pointed out in the previous chapter. Second, although perceptual salience may point to conceptual salience due to the underlying design principle, we cannot hold the same for the opposite relation, that is, conceptually salient regions can be perceptually indistinct. Figure 2-1 (that is identical to Figure 0-2b) illustrates an example, where detecting the global minima among local minimums becomes a challenge. In order to detect the year with the lowest population value in a graph with three local minimums that are very close in value, relying on perceptual saliency may not help at all. That is why a system that considers conceptual ascriptions as well as considering sensory saliency when deciding on the content of verbal assistance may bring the user's knowledge up to higher level by successfully communicating the conceptual event that the graph depicts.

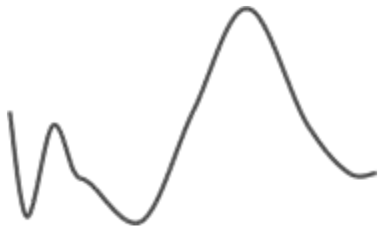


Figure 2-1 A graph sample where detecting the global minimum is challenging

The conceptual design of the proposed system varies from the previous works in several aspects. Figure 2-2 illustrates properties of the proposed system and also other possible design choices that were made for existing systems. First, the proposed system allows the user to access the graph via active haptic exploration in addition to provision of verbal descriptions of that graph. Second, since explorers can explore and perceive the graph actively, verbal assistance is given sentence by sentence just in time (namely, in an incremental way) instead of giving a full summarization at once. Third, this design takes into account the user's current and previous exploration patterns (their haptic actions) as well as the user's and system's previous utterances when deciding "*what to say*", "*how to say*" it and "*when to say*" it. When users interact with a haptic graph by means of a haptic device interface, their exploration movements can be observed by the system's assistance module similar to how a human observes a blind person's tracing movements when giving additional assistive verbal information during the course of the knowledge acquisition task. A similar verbal-assistance approach in the domain of tactile maps (Lohmann and Habel, 2012) and of floor plans (Yu, 2014) has been realized successfully.

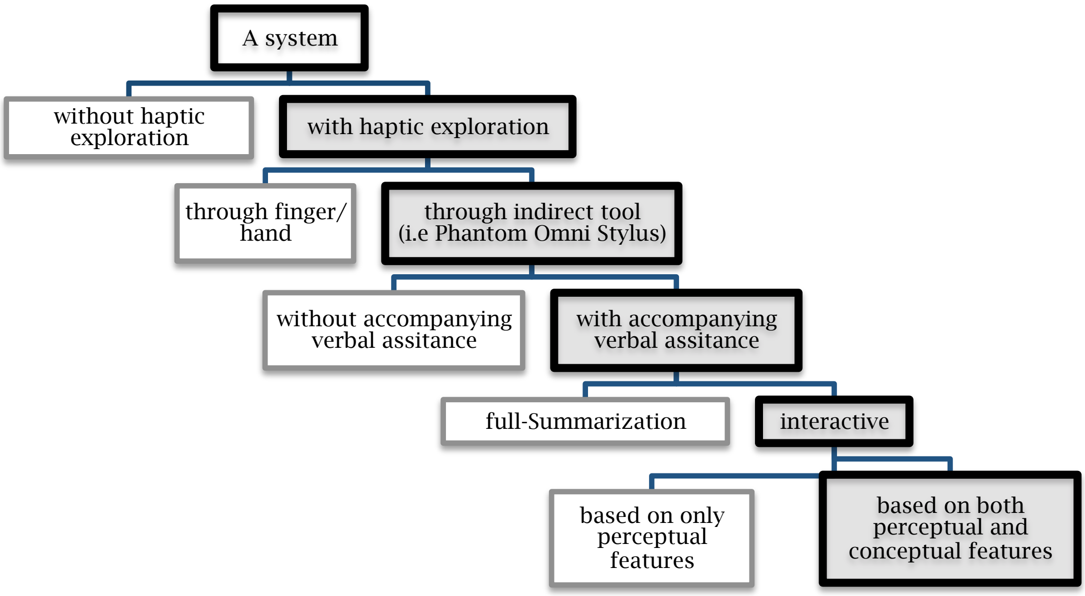


Figure 2-2 Properties of our conceptual verbal assistance system (highlighted using a darker background)

To that end, as well as for the purpose of investigating the underlying mechanism of haptic graph comprehension (introduced briefly in Chapter-1), how language interacts with graph comprehension and the dynamics of collaborative activity should also be considered for an effective and efficient system design.

2.2. Referential Links in Multi-Modal Graphical Representations

For the last decade, our research group has consistently employed statistical line graphs retrieved from a waterbird census report (PRBO) in the investigation of various research questions regarding graph comprehension for the sake of having a standardized stimuli set[4]. To comprehend language-graph constellations (ranging from annotation on graphs to verbal assistance during haptic graph exploration), the link between *reference* and *co-reference* has to be established at the outset. One of the topics covered in this report is the summarization of information about the changes in a population distribution provided in the *Bolinas Lagoon Population Trends* report. A sample excerpt is given in (1) to exemplify this. In this report, the summarization contains a line graph (see Figure 2-3) and an accompanying text for each observed bird species.

(1) *Bolinas Lagoon Population Trends*

> From a peak of about 60 wintering birds in 1976, numbers have declined to about 20 birds currently.

[4] see Alaçam, Habel and Acartürk, 2015, 2014, 2013ab; Alaçam, Acartürk and Habel, 2015, 2014; Acartürk, 2014, 2010; Acartürk and Alaçam, 2012; Habel and Acartürk, 2009.

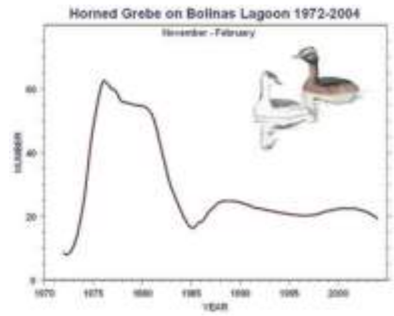


Figure 2-3 The population graph of wintering birds

Communication through line graphs is a specific case of communication about physical space since conceptual representations are built upon a set of abstract structures (Habel and Eschenbach, 1997; Eschenbach, Habel, Kulik and Leßmöllmann, 1998; Habel, 2005; and also see Jackendoff, 2002 for *the representational modularity hypothesis*). As already said in the previous chapter, graphs depict abstract events via physical entities, such as lines and bars. Referential expressions about the graph (represented in Figure 2-3), like 'peak of about 60', can be interpreted as referring to the domain of discourse (waterbirds at Bolinas Lagoon), namely abstract-conceptual event, or as referring to the domain of external representations, in particular, the graph line, which depicts data about the domain of discourse. Therefore, when communicating through graphs, whether in written or spoken format, two communicational modalities, language and graphics, contribute to a common conceptual representation. And *inter-* and *intra-representational coherence* by employing internal conceptual representations has to be established for a successful communication (Habel and Acartürk, 2007; Acartürk, 2010). Sentence (1) is a verbal characterization of the population trend of Bolinas Lagoon's Horned Grebes specifying it as a specific change, namely a *decrease_of_value* change, lexicalized by *decline,* with a *begin-value* (realized by a *from_PP*) and an *end-value* (realized by a *to_PP*). The graphical counterparts of these verbal items are the *global maximum* and the *right end-point* of the graph line. From the perspective of a seeing human, who produces a trend description based on what is depicted by a line graph, the salient parts of the graph line are primary candidates to refer to[5]. In other words, the referring expressions are evoked by visually salient graph entities. In general, i.e. both in static and in dynamic scenes, *color* and *motion* play major roles as visually salient attributes in identifying entities using referring expressions (see, e.g., Koolen, Goodbeek and Krahmer, 2013; Carmi and Itti, 2006). In the static domain of line graphs *shape* is the dominant saliency attribute.

For verbalizing line-graph descriptions (i.e. as trend descriptions) focusing on either a particular segment or a combination of line segments, the conceptual inventory has to fulfill requirements from perception and from language. In other words, the content of the verbal description is dependent on the segmentation of the event into meaningful sub-events (on a partonomic level) and then on the decision of appropriate referring expressions (in regards to taxonomic level) in a way that it facilitates both on-line comprehension and memory.

Qualitative approaches to shape representation usually focus on the shape (e.g. Hoffman and Richards, 1984; Eschenbach et al., 1998), and on the curvature landmarks (Cohen and Singh, 2007) of contours. These landmarks involve visually salient regions, such as positive maxima and negative minima, depending on the concepts of convexity and concavity of contours, and inflection points. Graph lines can be seen as a specific type of contours. A qualitative analysis of graph lines requires additional shape representations and shape cognition characteristics, which go beyond that of contours. In general shape segmentation, no specific frames of reference are presupposed. In contrast, line graphs

[5] see Habel and Acartürk (2007, 2009) and Acartürk (2010) on details of these semantic and conceptual analyses

mostly provide a distinguished frame of reference, explicitly via the axes or implicitly through the medium of presentation (such as a printed page or a computer screen) such that the interpretation of graphs lines requires graph-schema knowledge (Kosslyn, 1989; Pinker, 1990). That is, some additional concepts are needed in graph-shape segmentation (see Alaçam, Acartürk and Habel, 2014; Habel, Alaçam, and Acartürk, under revision). Here, I aim to introduce the fundamental issues that need to be taken into consideration in the design of a verbal assistance system. Details regarding each of those issues elaborated here are discussed in the upcoming chapters, together with the interpretation of empirical results.

Producing a description for a graph line can be considered as a sort of path description. As also investigated in Lakusta and Landau (2012), when producing basic path descriptions, the prominent role should be given to the 'goal', which is done by backgrounding the source/start. To illustrate the inventory of shape concepts for language production in a simplified way for the sentence (1) depicted in Figure 2-3, only prominent (i.e., highly salient) shape landmarks (*sl*) will be exemplified here.

sl_1 – left-endpoint

sl_2 – global-max & positive-curvature max /local max.

sl_3 – local-min & negative-curvature-min

sl_4 – right-endpoint

Choosing the more prominent *shape landmark* (sl_2) for the source role and neglecting sl_3 leads to the core structure of sentence (1). The production of appropriate referring expression depends on fine-grained conceptual representations of the selected prominent shape landmarks. Accordingly, granularity can be adjusted from fine-grained to coarse-grained to avoid repeating the same long descriptions constantly for the sake of communication. For producing such referential expressions, using the *incremental conceptualizer* INC seems to be appropriate (see Guhe, Habel and Tschander, 2003). But this dissertation focuses mainly on the conceptual and empirical investigation of verbal assistance system. The computational aspects are left to future studies, although the technical infrastructure for such a system has been developed already (Kerzel, 2015; Kerzel, Alaçam, Habel and Acartürk, 2014; Lohmann, 2012).

Following a method proposed by Acartürk (2010), in the geometric specification of a graph, the line (proper or specifier) in line graphs in time domain is a representation of the "Path" concept. To exemplify, a line segment *sp4-sp5* contains a *PATH* structure as well. In line graphs, spatial verbs such as "rise", "going up" and "increase" are representations of another subordinate concept, namely *Path of Motion*. As a symbolic representation, it can be represented as "Increase Of Value", where *temp* and *value* are path arguments (Habel and Acartürk, 2007). Temporal aspects of the process are carried by the *temp* argument, while the *value* argument stands for the amount of change. The geometric specification of the "Increase" concept inherently carries the necessary conditions such as "*Value*(Begin(Increase)) < *Value*(End(Increase))" (see Eschenbach et al., 2000; Acartürk, 2010; Acartürk (2012) for more details regarding Geometric Concept Specification).

According to Landau and Jackendoff (1993), the shape is the dominating criteria in labeling shape-based representations. Cohen and Singh (2007) showed that shape characteristics such as contour length, turning angle, curvature polarity and curvature magnitude (also orientation, location and size, see Cohen and Singh, 2007; Denisova, Singh and Kowler, 2006) have an influence on the segmentation and identification. Basic object segmentation starts with the automatic division of complex shapes into sub-units. There are consistent and predictable rules that govern this segmentation (such as using visually salient features as a basis). The resulting sub-units play a crucial role in the storage of visual information (e.g. Cohen and Singh, 2007; Tversky and Hamenway, 1994; Biederman, 1987). Incremental methods require a rich database to facilitate the generation of verbal descriptions on different partonomic and taxonomic levels for each shape landmark, each shape segment and also for the global shape. Such a system can also produce verbal descriptions on different levels of granularity (in regards to partonomic and taxonomic relations) regarding the same line segment and combinations of line segments[6]. To illustrate this, example (2) presents two verbal descriptions that can be produced for the complex line in Figure 2-4 consisting of *ep1-sp1* & *sp1-sp2*. Example (3) illustrates descriptions that have the same partonomic structure but exhibit differences in the taxonomic granularity. The choice of the granularity in the content may lead to a different conceptualization of an event caused by the differences in the highlighted information.

(2) Referring expressions on different partonomic levels;

- It decreases$_{[ep1\text{-}sp1]}$ and increases$_{[sp1\text{-}sp2]}$.
- It has a V-shape$_{[ep1\text{-}sp1\ \&\ sp1\text{-}sp2]}$.

(3) Referring expressions for sp1 on different taxonomic levels;

- Its value is 65.
- This is a local minimum.
- This is a local minimum of 65.
- This is the first local minimum.
- This is one of three local minima.
- This is the second global minimum[7].

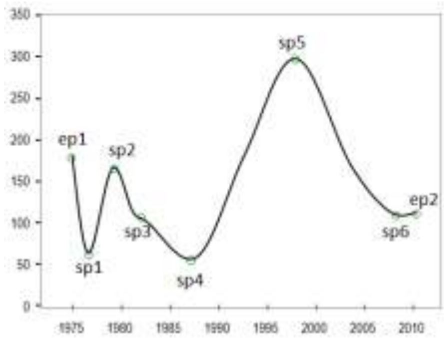


Figure 2-4 A sample graph with labeled shape landmarks

In the generation of referring expressions (Reiter and Dale, 2000), the identification of the referent properties, such as prominent shape landmarks, has been conceived as an important phase. The identification phase alone, however, is insufficient for providing a complete specification of the generation process (van Deemter, Gatt, van Gompel and Krahmer, 2012). Over the last few decades, research on referring expressions have focused on various communicative aspects that have gone beyond the identification phase, such as forming conceptual pacts during the course of interaction in a collaborative environment (Brennan and Clark, 1996), and on overspecification (Pechmann, 1989). In this field, the concept of overspecification is mostly associated with Grice's (1975) maxims, which target adequacy, efficiency and sensitivity. The studies that focus on the production of referring expression (e.g. Koolen, Gatt, Goudbeek and Krahmer, 2011, 2009; Pechmann 1989) proved that humans usually produce

[6] The effect of shape characteristics on event segmentation is discussed in Chapter 6.

[7] The descriptions should be selected carefully. For instance, this description may be misundestood since the sequential order of the peaks is more available information than the ordinal ranks in the values of the local points.

overspecified referring expression for a referent. On top of it, the reason for having an assistance system is to close the gap caused by the sequentiality inherent in haptic perception. That is why assistive interfaces that provide more information than required would be acceptable in many cases.

Another important aspect in daily-life communication is the saliency in the sensory modality through which the communication takes place. There are many different factors that could define saliency, including spatial features, perceptual properties and conceptual properties (Carlson, 2008). Previous research has shown that not only saliency in the domain of discourse in the linguistic context but also saliency in the visual context influences humans' choice of referring expressions (Fukumura, van Gompel and Pickering, 2010). Similarly, saliency in the haptic context is also expected to influence the choice of referring expressions (Section-II addresses this issue in more detail and also see Habel, Alaçam and Acartürk, 2013).

The investigation of multimodal interactions (namely interaction by means of language, graphs and also gestures) requires both systematic qualitative and quantitative analyses. A widely accepted method that addresses the generation of referring expressions was developed by Dale and Reiter (1995), and, is utilized to characterize the semantic properties of graphical segments and the referring expressions produced during collaborative activity. Here, I do not aim to go into implementation level in detail, instead I employed this method as a tool to systematically map the semantic properties of the graphical features to the referring expressions produced by the participants. According to Dale (1992), a system that generates referring expressions should at least satisfy Gricean-like conversational maxims for the sake of adequacy, efficiency and sensitivity. In more detail, a referring expression should contain enough information to allow the hearer to identify the referent, it should not contain unnecessary information and it should be sensitive to the needs and abilities of the hearer.

Most of the research on referring expressions in discourse, in particular in situated and co-operative dialogues, focus on the production of REs for objects or events in visual worlds that is, for entities perceived visually by the interlocutors (e.g. Bard, Foster and Arai, 2014; Gatt, Krahmer, van Deemter and van Gompel, 2014). The referring expressions produced by the haptic explorers and verbal assistants during collaborative activity give insight into how graph readers comprehend graphs, which elements are mentioned most, and how they are referred to.

In a line-graph comprehension setting, referring expressions are not produced for distinctively separate objects, but for line segments, i.e. the entities that belong to a continuous stream of data. Objects are space-bounded, and events are both space- and time-bounded. However, comprehensive analyses performed by Zacks and Tversky (2001) indicated that although there are differences between the two, *events* can be also treated and analyzed as *objects*. The events represented by a graph line (e.g., changes in population) unfold in time. That is why, despite the static nature of the visual stimulus for the graph reader, the spatio-temporal characteristics of the line segments (the entities to be referred to by shared or distinctive attributes) are inherently bounded among one another. Accordingly, the segmentation of the continuous data is a fundamental step in the production of the referring expressions, and involves the use of *gradable modifiers* inherently. In the graph domain, all data points (which are possible candidates to be selected as a landmark) are positioned on the graph line therefore they are temporally located due to the restrictions of a spatio-temporal representation. This means that they are all connected, and relatively small-size and specific set of spatial relations can be used in reference to them; such as "Earlier/later", "faster/slower", "higher/lower", and "upper/lower".

Asking for descriptions about and summarization of the spatial layout of the space is the most employed common experimental paradigm in the investigation of spatial cognition (e.g. Tversky and Lee, 1999; Landau and Jackendoff, 1993) and also in the investigation of graph comprehension (*see Footnote-4* for the references). Acartürk (2010) proposed a conceptual inventory of spatial terms and a guideline for a holistic description generation system. His study showed that the vocabulary for communicating through line graphs consists of shape nouns, spatial prepositions, adverbial modifiers, and verbs of change in space. Additionally, the findings also indicated that the name of the domain variable, time numerals (e.g. the unit of time (e.g., 'year') and the unit of domain value (e.g., 'population') were the entities of the conceptual event most frequently referred to. He concluded that these four components should be included in a typical verbal description template for an automatic verbal generation system. Therefore, the verbal descriptions of the graphical entities and the spatial concepts, which are expressed linguistically, should be systematically investigated.

Graphs are not just spatial representations; they contain both spatial and temporal information. In order to talk about a spatio-temporal representation (namely a graph line) that depicts what happened and when happened regarding abstract event, the event depicted by the event needs to be evaluated (production perspective) and constructed (generation perspective) by taking those aspects and their relation into account. The semantic annotation method that combines abovementioned <attribute, value> pair representation with a time/event sensitive annotation format has been proposed for a systematic analysis. The use of a method, which is sensitive to time/event, allows preserving these *what* and *when* dimensions without losing the relation in-between (Pustejovsky, Castano, Ingria, Sauri and Gaizauskas, 2003; Schilder and Habel, 2001). Annotation is a labeling process that discretizes the content to ease the analysis and to highlight the similarities and differences. As such, the annotation process may suffer from a loss of relations and semantic details might be leveled out. Schilder and Habel (2001) proposed a semantic tagging system that separates event- and time-denoting expressions and focuses on the extraction of temporal information from a text by preserving the semantic relation. Pustejovsky and his colleagues (Pustejovsky, Lee, Bunt and Romany, 2010; Schilder, Katz and Pustejovsky, 2007; Pustejovsky et al., 2003; Hobbs and Pustejovsky, 2003) proposed a standard annotation framework, namely ISO-Space, which relates events (Spatial-ML) to time (Time-ML). Both approaches are based on Allen-like interval relations (such as "one before the other", "simultaneous"). This interval-like tagging is particularly suitable for the analysis of communication over graphs as well. Graph readers explore the graph before segmenting them into meaningful sub-parts, and use these parts for online comprehension (by linking perceptual cues to conceptual information about the depicted event) or for memorization. Consequently, annotation schemes that maintain the order of the intervals, the size of the interval and the number of intervals (Pustejovsky et al., 2010) are useful in this domain.

2.3. Haptic Graph Comprehension as a Collaborative Activity

One of the first HCI examples of a collaborative task between both visually-impaired and sighted students was presented by Moll and Pysander (2013). They investigated the dynamics in a collaborative problem solving environment that enables both haptic access through the Phantom Omni device and visual access to the same learning material. They found that the verbal utterances produced by the sighted students contributed positively to the visually impaired students' comprehension of changes that are hard to perceive haptically due to sequential exploration, providing successful example for such collaborative instructional environment.

Verbally-assisted haptic graph exploration is a task-oriented collaborative activity between two partners (Alaçam et al, 2014; Habel et al., 2013), a (visually impaired) haptic explorer (*E*) and an observing assistant (*A*) providing verbal assistance in the process (Figure 2-5). Verbally assisted haptic exploration has a dialog-like character:

(2.a) *A* has to synchronize language production with *E*'s hand-movements in a turn-taking manner.

(2.b) The quality of the verbal assistance depends on how well appropriate referential and co-referential links are established.

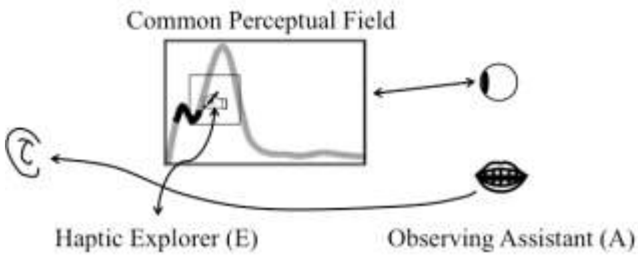


Figure 2-5 Assisted haptic graph exploration, a joint activity

A and *E* share a common field of perception, namely the haptic graph, but their perception and comprehension processes are highly different. For example, while *E* explores the segment of the haptic graph highlighted in black, *A* is able to perceive the global shape of the graph, in particular, *A* is aware of the shape landmarks and line segments. When E explores the first local maximum followed by a local minimum (see Figure 2-5), and does not have any information on the global maximum, unlike *A*. Therefore, the explorer and the assistant have different internal representations of the graph line, and *A*'s referring to the graph could augment *E*'s internal model substantially. Uttering (at the position depicted in Figure 2-5) "Now you have reached the heights of the last peak", for instance, would provide *E* with additional information. Another suitable comment would be "You are in the increase to the population maximum", or even "You are in the increase to the population maximum of about 90 that was reached in 1985". As mentioned in *2.2*, verbal assistance is a type of *instruction*[8], *thus* over-specified referring expressions are adequate for our domain (see Koolen et al., 2011, 2009).

The success of the joint activity of the explorer *E* and the observing assistant *A* in general, and also the success of *A*'s utterances in particular, depend on the alignment of the interlocutor's internal models, especially on building an *implicit common ground* (Garrod and Pickering, 2004). Since *E*'s internal model of the *activity space* (i.e. the haptic graph and *E*'s explorations) is perceived via haptic and motor sensation and *A*'s internal model of the same space is built up via visual perception, similarities and differences in their conceptualization play the central role in aligning on the situation-model level. To be assistive, *A* should verbally provide *E*—in particular—with content (i.e. conceptual representations) difficult to acquire haptically. This—haptically difficult to build up—content has to be combined with haptically-explored content in the same sentence (or phrase) to fulfill the given-new contract (Clark and Haviland, 1977).

In addition to communicating over this kind of descriptive information about abstract graph domain knowledge, "*instructions*[9]" such as navigational commands regarding the use of haptic stylus play a crucial role in the communication as well. In general, all the information exchanged between the explorer and the assistant can be categorized under two communicative goals; descriptive and instructive (see Chapter 4.5.2, Chapter 9 and also Alaçam, Habel and Acartürk, 2014). Taken together, the success of a collaborative activity is dependent on several factors; establishing common ground can be treated as a fundamental factor. As will be elaborated later in Chapter 9, this alignment may occur at different layers, from linguistic alignment to alignment of situation models. Based on the established common ground, the switch between different reference frames (spatio-temporal or action perspective/ego-centric) depending on the communicational goal can be achieved more fluently.

---

[8] Here, this term is used as an umbrella term that covers both descriptive and instructive assistance
[9] This use of "instruction" is a subset of abovementioned term

2.4. An Assistive System from the Perspective of Human-Computer Interaction

Most of the existing research in the HCI domain on developing an assistance system for visually impaired people is relying on literature on visual perception and visual graph comprehension as there is a lack of research on haptic graph comprehension. The empirical findings of the investigations presented in the chapters of Section II and III provide insights that can be used as guidelines for the design of verbal assistance during the course of communication in the implementation of user interaction for visually impaired people.

According to the ISO (The International Organization for Standardization), usability is defined as "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in a particular environment", (ISO/CD 9241-11, 2015). These three components were also defined by the standard as follows:

> ***Effectiveness.*** "accuracy and completeness with which specified users can achieve specified goals in a particular environment"
>
> ***Efficiency.*** "the resources expended in relation to the accuracy and completeness of the goals achieved"
>
> ***Satisfaction.*** "the comfort and acceptability of the work system to its users and other people affected by its use".

These are well-accepted definitions in the field of HCI, however they can be considered as umbrella notions that need to be evaluated with more refined concepts. The cognitive science-oriented HCI perspective introduces a more branched and user-oriented approach to the field by incorporating the role of mental models in the HCI system design. Mental models are considered as a key concept in the development of instructional technologies, tutorials, and other forms of user assistance. According to Preece (1994), the use of mental models in HCI design brings crucial components to the surface, such as simplicity, familiarity (building on prior knowledge), availability, flexibility, feedback, safety and affordances.

From the perspective of HCI, a mental model is a model of what a user knows about a system. According to the definition by Norman (1983, p.8), "in interacting with the environment, with others, and with the artifacts of technology, people form internal mental models of themselves and of things with which they are interacting. These models provide predictive and explanatory power for understanding the interaction". In classical HCI interface design that takes the user's mental model into consideration, the concept of mental models is more attributed to the user's understanding of how the system works (Gentner and Stevens, 1983). On the contrary, the design of an assistive system requires a more elaborated approach, namely that the system should be sensitive enough to adapt itself to the user's and to then help the user to form correct the mental representation of the task at hand.

There are various approaches that can be followed in order to design usable interfaces. Empirical user testing, heuristics (design guidelines) and user modeling tools are the most standard methods among these approaches (e.g. Kieras, 1997). In empirical user testing, users are observed interacting with a system and performance measures are collected through various methodologies in the process (i.e. verbal descriptions, think aloud procedures, eye movement analyses) and analyzed thoroughly afterwards. However, conducting empirical user testing is a highly costly and time-consuming process. Therefore, low-cost methods such as *heuristic evaluation* by following design guidelines (Nielsen, 1994) and *(cognitive) user modeling tools* (also called a "*cognitive walkthrough*" by Lewis, Polson, Wharton and Rieman, 1990) are alternatives to this method. Design guidelines or design heuristics are constructed by experts based on their expertise and proficiency in the field or they can be also based on empirical user testing. Following design guidelines is a highly cost effective and rapid process, therefore they constitute one of the most valuable tools employed in the domain of HCI design. However, considering rapid improvements and emergence of new technologies, finding guidelines that target a specific topic is a challenging issue. Haptic exploration interfaces in general, and haptic graph comprehension specifically, suffer from a lack of this kind of research. Another frequently used HCI evaluation method are cognitive modeling tools, such as *the GOMS-family* of HCI tools (*Goals, Operators, Methods and Selection*, see Card, Newell and Moran, 1983; Card, Moran and Newell, 1980) and *CogTool* (developed at Carnegie-Mellon University, see John, Prevas, Salvucci and Koedinger, 2004). These tools provide a descriptive model of the way a user performs tasks on a system, that is, it predicts user exploration preferences based on the interface and the task. The development of models using these tools also requires domain knowledge and can benefit from empirical user studies.

In the graph domain, successful communication means correct mental representation of the abstract event depicted in the graph. Acartürk (2010) proposed a cognitive model for statistical line graph comprehension based on the GOMS analysis model. The findings suggest that the users' eye movement patterns over a graph can be predictive in creating the referential links between the graph and the accompanying text. Furthermore, the findings also showed that the type of prepositional phrases (i.e. 'since', 'between', 'from') may result in different gaze pattern characteristics. The results would be particularly relevant for designing a cognitive model of haptic graph comprehension for line graphs. It should be also noted that this study provides a solid basis for this purpose since the graphs employed in both studies were taken from same database (PRBO, the Bird Population database). On the other hand, cognitive modeling of haptic graph comprehension goes beyond the scope of this dissertation. Although the empirical findings obtained from the experiments provide valuable information in that sense, haptic graph comprehension involve many components that need to be dealt with, and this study can be considered as a first step towards achieving this goal. Therefore, follow-up analyses and studies are planned to reach to a comprehensive understanding of haptic graph comprehension, as will be briefly mentioned in Chapter 11.

In this dissertation, empirical studies were employed in gaining a better understanding of haptic graph comprehension and of collaborative activity dynamics towards the design of a verbally assisted system for visually impaired people. The results of the empirical studies contribute to the formulation of design guidelines (see Chapter 10). To my knowledge, this is one of the first studies that provide design guidelines for verbal assistance systems in the domain of haptic graph comprehension. It should be noted that the conceptual design of the task-oriented collaborative system presents multi-modal information in an incremental way. In short, it is based on the interaction between two agents; a human explorer who perceives the graph haptically through active exploration and a verbal assistance agent that provides helps the explorer with the requested help automatically. Therefore it contains different and intertwined layers. As can be seen in the schema illustrated in Figure 2-6, the conceptual design of the proposed system combines different dimensions of human-computer interaction systems. Although the guidelines coming from each of this field are of great use, they are not directly applicable to such multi-modal interactive systems, which allow haptic exploration with accompanying verbal assistance tailored to users' needs. Therefore, the design guidelines, which target specific dimensions, need to be taken into consideration and be converged for this purpose.

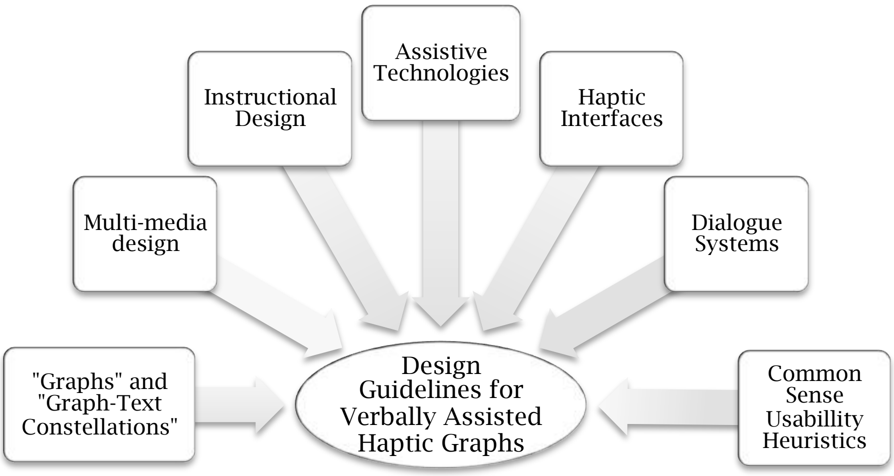


Figure 2-6 Different aspects of the proposed assistive system[10]

Within the scope of this dissertation, these existing design guidelines had contributions in two layers. First the collaborative experimental setting, which is introduced in 2.3, was designed by taking the advices mentioned in those guidelines. The empirical studies (which are presented in Section II and III) were conducted by employing this experimental setting. In the second layer, these guidelines were combined with the results of the empirical studies and this integrated knowledge was used as an input to construct the design guidelines for a verbally assisted system for haptic graph comprehension. In the following, I address the first layer, leaving the discussion of the second layer to Chapter 10.

First item in this scheme (Figure 2-6) addresses the design guidelines regarding graphs and graph-text constellations (Shah and Hoeffner, 2002; Acartürk, 2010). The statistical graphs with written or spoken accompanying verbal elements are forms of a multi-modal design, therefore design guidelines specific to graph domain, as it was proposed in the abovementioned study, provide a solid basis. Furthermore, although graphs have wide variety of usage areas, such as newspaper, scientific documents, reports etc., one of the main uses of them is in the domain of education and instruction. The instructional system design has a specific goal of conveying an instruction to facilitate learning. This field is known as *instructional system design* (also called *instructional design* or *instructional development*) (Gustafson and Branch, 1997; Schrock, 1995; Seels and Richie, 1994; Reigeluth, 1983). Therefore, the design guidelines concerning the instructional design should be also taken into consideration (Burgstahler, 2007; Silver, Bourke and Strehorn, 1998). Furthermore, the design of assistive systems as a form of instructional system can be considered as a domain-specific sub-category. The proposed system targets visually impaired people who may have varying degree of impairments. Therefore the guidelines regarding the assistive technologies are also highly relevant for my purposes. The concept of the "Universal access design", which usually co-occurs with the assistive technologies, is one of the growing fields in the HCI area (Lidwell., Holden, and Butler, 2010; Emiliani, and Stephanidis, 2005; Stephanidis, 2001; Shneiderman, (2000). The term *universial access, universial degin* or *design for all* in HCI has two kindred but distinct meanings. One use of this term addresses the needs of all potential users in a more politically correct perspective (Schneiderman, 2000). Second use has more specific and HCI-oriented meaning that addresses people with special needs and disabilities and elderly people as well.

[10] For each item, one design guideline was selected based on the coverage and relevance. However it should be noted that the items in this scheme and also the chosen guidelines could be diversified.

Moreover, focusing on the usability of haptic devices and haptic space (Sjöström, 2002) and also on the usability of the dialogue systems is getting popular in the domain of HCI. Although the usability of the haptic interfaces and hardware design (e.g. O'Malley and Gupta, 2008) plays important role in the use of such systems, the technical aspects concerning haptic graph design is outside the scope of this dissertation. Yet, the decisions and some of the crucial points, which were taken into consideration in the design process, are touched upon later on.

The design of the conceptual experimental setting is based on several decisions, which were supported by the abovementioned existing design guidelines. Mayer's multimedia principles (2001, 2009) address 12 design principles and these guidelines provide a solid and comprehensive body of knowledge that affords valuable information for a successful multimedia design. The list of the principles (retrieved from Mayer, 2009, p. 267–268) consists of "*Coherence*", "*Signaling*", "*Redundancy*", "*Spatial contiguity*", "*Temporal contiguity*", "Segmenting", "*Pre-training*", "*Modality*", "*Multi-media*", "*Personalization*", "*Voice*" and "*Image*". More instruction-oriented design framework and guidelines were introduced by Silver et al. (1998) in higher education with the aim of highlighting the importance of accessibility issues as an integral component of instructional planning. Inclusion of this framework in instructional design has been gaining increased attention since then. The guidelines from the *Universal Design of Instruction* (*UDI*) (e.g. Burgstahler, 2007) covers 7 main issues, these are "*Equitable use*", "*Flexibility in use*", "*Simple and intuitive use*", "*Perceptible information*", "*Tolerance for error*", "*Low physical effort*" and "*Size and space for approach and use*". Also ISO 9241-920 (2009) that provides guidance on tactile and haptic interactions provided very useful insights during this process. It should be noted common sense usability heuristics (Nielson, 1994; Preece, 1994; IBM Guidelines, 1992; Norman, 1983) apply to this domain as well.

In the light of those guides, the collaborative experimental setting of a haptic explorer and a verbal assistant during exploring haptic graphs

- utilizes sensory substitution to close functional and informational gap between visual and haptic modalities
- allows graph readers to explore graph actively and haptically
- assigns asymmetric activity roles to an explorer and an assistant
- aims to provide relevant information

The details of each of these design decisions are discussed in the following.

Using Sensory Substitution and Allowing Active Haptic Exploration

As discussed throughout this dissertation, haptic perception is local and sequential whereas visual perception allows access to both local and global information simultaneously. Due to this critical perceptual difference, functional and informational inequivalence of haptic graphs to their visual counterparts need to be taken care of for successful communication over haptic graphs. ISO 9241-920 (Guidelines 3.1.2 and 4.1.4) recommends sensory substitution for haptic interfaces for providing accessible information and easing the encoding of the haptic properties. As stated in another multi-modal design guideline that targets visuo-haptic interfaces (*SH-8: haptic spatial structures and TH-1: temporal haptic metaphors* in Nesbitt, 2005), the haptic sense can carry both temporal and spatial dimensions successfully, so do the statistical line graphs. Therefore instead of providing only verbal description of the event (i.e. the full summary of the graph), allowing the graph reader to explore the graph haptically, letting them explore the spatio-temporal features through haptic modality would lead to a successful communication. This decision is supported by Mayer's "Multi-media design" principle stating that "people learn better from words and pictures than from words alone". Although the Mayer's guideline specifically addresses the interaction between the visual and verbal modalities, it also points out the superiority of the multi-media settings, which allow explorer to extract information from different kinds of modalities, which are good at carrying specific knowledge, (i.e. the visual modality is very good at carrying metrical information). Acartürk's study (2010) on the graph-text constellations proposes a similar principle, which is more specific to the graph domain. He states "users recall better from text-graph constellations rather than from text alone". The haptic exploration (as analog representational modality, see 1.2 for a reminder) via one-point interaction foregrounds the shape, which is an integral part of graph comprehension. On the contrary, the language modality can represent this information in discrete/symbolic categories (i.e. words and phrases). From the instructional point of view, providing information in a modality which can be equipped well to carry that particular information may lead to simple and intuitive comprehension of the intended information. This issue is also addressed by the "*Simple and Intuitive use*" principle of the UDI Guidelines (Burgstahler, 2007; Silver et al., 1998).

Additionally, ISO 9241-920 guidelines also suggest that active exploration is superior to passive exploration since it involves kinesthetic perception that allows the extraction of many material and structure related haptic properties. This guideline is also in line with the existing literature on haptic object perception and comprehension (see Chapter 1.1 and 5.1). By means of active exploration, the explorer can focus on the graphical entities with respect to their own needs and the task (summarization, extrapolation etc.) at hand.

Based on those principles together with the insights elaborated on the literature reported beforehand, allowing visually impaired graphs readers to explore the graphs haptically instead of providing just verbal summaries of the graphs was the first design decision for the proposed system.

Asymmetric Activity Roles Leading to Communication of Only Relevant Content

In this collaborative design, the activity roles are assigned in an asymmetric way between haptic explorers and verbal assistants, namely the haptic explorers initiates the dialogue by requesting information and the assistants provide help based on the explorer's need. The haptic explorers actively explore the graph and acquire information by themselves, therefore initiating a dialogue when needed would decrease communicative overload in a dialogue. This design decision guides the system to provide only relevant information. Still many decisions need to be taken for achieving successful verbal assistance. The evidence-based design heuristics constructed based on the empirical investigations regarding the task of providing relevant information is discussed in detail in Chapter 10.

Furthermore, in the experimental setting, the communication between the explorer and the assistant was restricted to only conversational utterances. To be more precise, people may think aloud when engaged in an activity, especially if it is cognitively demanding. Self-talks can be silent and may show a lack of grammatical structure. Thus the recognition and making sense from them would be highly challenging for speech recognition module of the to-be-implemented HCI system. Therefore, in order to create comparable environment, in the collaborative setting employed in the empirical investigations, the explorers are asked to turn the microphone on and open the communicational channel for asking help.

Design of Haptic Graphs

According to "Low Physical Effort" principle of the "Universal Design of Instruction" (Silver et al., 1998), "the design should be used efficiently, comfortably, and with a minimum of fatigue". This point is also one of the issues addressed in ISO 9241-920 (Guidelines 3.1.6). This could be achieved by helping participants staying inside of the engraved haptic line. As a part of the experimental procedure, warm-up sessions were conducted to familiarize the participants with the device and also with the haptic graphs. These sessions helped the participants in increasing their proficiency of the device use, still some haptic design parameters (i.e. depth or friction), may interfere due to their saliency. When the depth of the engraved graph line is relatively shallow, graph readers may have problems in staying inside of the graph line, therefore the depth parameter was adjusted so that this problem has been minimized. Although the guidelines concerning the design of the haptic graphs are not intended to be addressed here, the successful communication over graphs is highly dependent on successful design of the haptic lines, thus minimizing user's falling out of the graph serves to the purpose of providing uninterrupted exploration causing less anxiety. However, restricting falling out from the graph line may also cause problems in exploring graph space. When the explorers are only allowed to stay in the graph line, establishing relationship between the reference frame and the graph line, which is particularly helpful in locating a landmark on the reference frame, would be very difficult. Therefore, allowing the participant to explore all graphical entities is important, however while doing that, the separation between surface, namely the graph space, and the graph line should be apparent as suggested in ISO 9241-920 (Guideline 6).

Another issue that was taken into consideration is the use of appropriate spatial workspace, which is addressed in the ISO 9241-920 (with the title of "Using appropriate spatial addressability and resolution", Guideline 4.1.5). The haptic workspace was adjusted as not too small so the details of the graph shape can be represented, and as not too large as well in order to prevent the misrepresentation of a linear line at the edges due to arc-shaped movement of the device's stylus (discussed in Chapter 5.1.1).

The presence of haptic data labels can be also discussed under this topic. According to Mayer's signaling principle, "people learn better when cues that highlight the organization of the essential material are added". In visual graphs, data labels are more than just cues, actually they are crucial informative parts of graphical representations, However providing *haptic* data labels would cause distraction in exploration and encoding as well. In haptic modality, relating data labels represented in the axis and also data point requires effortful integration processes due to sequential and local haptic perception. Another design principle of Mayer, which is called as "spatial contiguity principle" states "people learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen". Thus it can be stated that the usability guidelines also supports the decision of not providing haptic data labels.

The consistency between the modalities of the multi-modal system should be also taken into consideration. The consistency concerning the interaction of the visual and haptic modalities was addressed by Nesbitt (2005: guideline SH.4 and SH.5), while Acartürk (2010) highlights the importance of providing consistency in the scalar entities provided in the graph-text constellations. Furthermore, for visually impaired people with low impairment, who can have partial visual access to graphs (or who acquired graph-domain knowledge through visual experience), coherence and consistency between visual and haptic modalities would be critical and inconsistent information would impair judgments. Additionally, this issue is also important for the sake of systematic analysis across visual and haptic modalities in the empirical investigations. Therefore, the haptic graphs are transformed directly from the visual graphs by using 3D virtual environment design programs, and they are identical to each other in representing amodal geometric properties, see Figure 4.3 in Chapter 4 for graph samples in each modality.

A systematic task and user analysis is another important aspect for the acceptance of the system by the target group. Not only the users' performance but also their willingness to use the system is one pivotal usability criteria and is considered as vital to the system's success (Davis, 2006; Venkatesh, 2000; Dillon and Morris, 1996). User acceptance is defined as "demonstrable willingness within a user group to employ information technology for the tasks it is designed to support" (Dillon and Morris, 1996, p.8). Human-like expressions (van Deemter, Gatt et al., 2012; Gatt and Belz, 2010), voice (Lee, 2010) or gestures (Bergmann, Kopp, and Eyssel, 2010; Kopp and Bergmann, 2012) have been shown to be of positive influence on the users' reaction to a system. In providing helpful instructional verbal assistance system for visually impaired users, the content of referring expressions is the most prioritized issue. As stated in Goudbeek and Krahmer (2012), human likeness is an important criterion in the generation of referring expressions and also in their evaluation. Nevertheless, humans can make errors and they can produce over- or under-estimated expressions. As such, exactly modeling how humans produce referring expressions may not be suitable and acceptable for assistive system. Therefore the results obtained through empirical research that contain human production of referring expressions should be evaluated by taking those into account with the aim of providing optimal solution by preserving naturalness and being assistive.

The design guidelines can be incorporated in different stages of the development process of a HCI system. From a software engineering perspective, there are many software development models (also referred to as life cycle models, such as the Waterfall model, the V model, the incremental model etc.). One of the design methodologies is the rapid prototyping, which can be considered as the fastest one with coarse level analysis (Tripp and Bichelmeyer, 1990). In this methodology, the analyses and the development are conducted in parallel. On the other side of the scale, more structured models are located such as ADDIE (analyze, design, develop, implement, evaluate), which is one of the standard instructional system design framework developed by Branson, Rayner, Cox, Furman, and King (1975). According to Reigeluth (1983), instructional design is basically about understanding, improving and applying methods of instruction. Applying such techniques optimizes the effectiveness and efficiency of a product since it emphasizes each phase from analysis to evaluation separately (see Allen, 2006; Akıllı, 2004)).

Figure 2-7 depicts the realization of the 5 phases of the ADDIE model towards the design of verbally assisted haptic graph comprehension system. The experiments presented in this dissertation were conducted incrementally. First, single-user studies and then a dual-user study with different research questions were performed. The lessons learned in those experiments were applied in an experiment conducted with visually impaired people. Within the scope of this dissertation, I narrowed my research down to conceptual and empirical research on haptic graph comprehension and human-human communication (Step 1A). The technical implementation of providing rule-based (incremental) canned text for the current location of the stylus on the Phantom Omni device (Step 1B) has been done already by other researchers in our research (Kerzel, 2015; Lohmann, 2012). The future studies will address the implementation and evaluation (Step 2) of the proposed system.

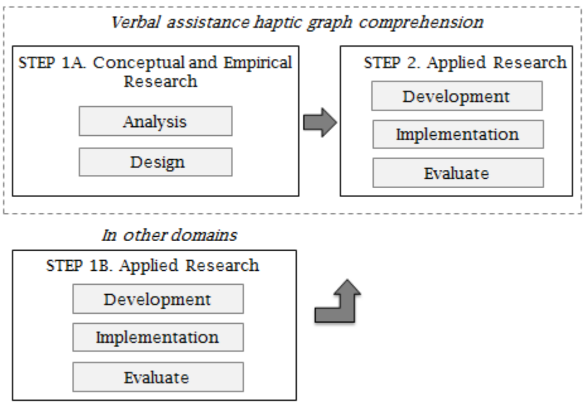


Figure 2-7 Design steps for the proposed verbal assistance system (following the steps of the ADDIE model)

The topics investigated in this dissertation are approached from several interdisciplinary research fields; HCI, cognitive science and computational (psycho-) linguistics. Therefore, I intend to use the most prominent and widespread terminologies regarding all these fields. However, because of the multitude of approaches, differences regarding the terminology are inevitable. In a nutshell, one of the most prominent differences between cognitive scientists and HCI practitioners are the evaluation methods. Cognitive science demands statistically significant and systematically controlled studies, whereas HCI evaluations are less strict in terms of scientific approach. Besides, the statistical power is not one of the required specification due to tendency to achieve optimal cost-benefit balance, although it requires thorough task and data analyses, but in a more qualitative manner (Rogers, Rutherford and Bibby, 1992). Within the scope of this dissertation, both quantitative and qualitative data analysis methods were applied.

CHAPTER 3

Gesture as a Research Paradigm for Verbally Assisted Haptic Graphs

The analyses of verbal descriptions and drawings are frequently used empirical methods in the investigation of (graph) comprehension. However, in addition to the rich data provided by verbal descriptions, gestures point out the hard-to-encode information and conceptually salient points as well as perceptually salient regions and entities of the graph. In this chapter, first a definition of gestures is given (3.1), after which well-known gesture frameworks (3.2) are introduced. The role of gestures in conceptualization is discussed in 3.3. Afterwards, the current state of gesture use in HCI domain and the use of gesture analysis as a research paradigm in the field of human computer interaction is discussed in 3.4.

3.1. The Definition of Gesture

In everyday language, the movement of one's hands and arms, adjustment of one's posture, the touching of oneself during a communication, nervous tics, and other movements that people use while talking can be regarded as a gesture (see e.g. McNeill, 2005; Roth, 2001). However, for scientific purposes, the scope of what is regarded as a gesture in cognition is usually defined by the "hand and arm movements that are interpreted by others as part of what a person says" (Roth, 2001). According to the well-defined classification proposed by McNeill (2005), there are three main categories of gestural movements: First, simple up-and-down movements without carrying semantic information are called *beat gestures*. The second category is comprised of *interactive gestures*, which are movements used to manage turn taking and other aspects of an interaction between multiple speakers. The third and final main gestural movement category consists of *representational gestures* and these can be organized in three different sub-categories, namely *deictic*, *iconic* and *metaphoric gestures*. An explicit pointing action, usually involving a finger or a forearm purposefully directed toward a display or an item in the environment is referred to as deictic gesture. They are very crucial because they let both listener and speaker know what the object of attention is. Iconic gestures are considered to be movements that depict a referent with the motion or shape of the hands. The so-called metaphoric gestures depict a concrete referent or indicate a spatial location for an abstract idea. Here, only representational gestures (focusing more on the iconicity) were considered. Gestures are idiosyncratic, mostly spontaneous movements that accompany speech and they are shaped directly by the meaning McNeill (2008). More specifically, iconic gestures convey information about shape, size motion and action characteristics of the events described in the speech (e.g. Tversky, 2011; Özyürek, 2010; Özyürek, Kita, Allen and Brown, 2005). Therefore, they reveal the speakers' mental representations (McNeill, 2008).

3.2. Gesture Frameworks

It is almost impossible to address the gesture without mentioning the embodiment approach (e.g. Gibbs, 2005; Shapiro, 2007). The embodied theories of cognition emphasized the link between perception and action. Perception and action are intricately linked, each influencing and determining the other in a broader system of coordination. No event can be thought of as being purely perceptual or purely motor. This tight coupling of motor and perceptual processes that is so important for physical interaction with the world, may also be important for the mental representation of the world (Barsalou, 1999).

The perception/action cycle is also considered as the origin of the gestural movements. The studies on gesture-language interaction are mainly based on the assumption that concepts are sensorimotor, emphasizing that they are grounded in the physical world and based on perceptual experiences (Barsalou, 1999; Garbarini and Adenzato, 2004). If the concepts are concrete and easy to visualize, speaker gestures more, besides, even abstract concepts are claimed to be grounded in physical terms in our understanding of the physical world (Gibbs, 2005 Johnson, 1987; Lakoff and Johnson, 1980) and people also produce gestures while talking about those abstract concepts (e.g. Hostetter and Alibali, 2008)

There are several frameworks that investigate gestures from various perspectives. These frameworks (that will be reviewed shortly) mainly focus on different aspects of these key issues; firstly, some of them focus on explaining the nature of the representations that underlie gestures, whereas others investigate whether gestures are communicative or not. How gestures facilitate speech production and how gesture and speech are integrated are also further topics that are extensively examined by gesture researchers. However, although there are significant differences on these issues, all frameworks converge on that gestures derive from spatial representations and they embody entities that are conceptual and abstract (Hostetter and Alibali, 2008).

Gesture studies can be divided into two main categories, the studies that focus on gesture comprehension and the studies that focus on their production. In the former branch, researchers are mostly interested in understanding the communicative role of gestures and their effect on the interlocutor's comprehension. The latter branch deals with the underlying principles of gesture production by focusing on the gesture producer's perspective. In this dissertation, I employed the analysis of speech-accompanying gestures as a tool to understand how a graph reader conceptualizes the events represented by a graph. Therefore, the review on frameworks mostly covers the second line of research.

Some views (e.g. Krauss, Chen and Gottesman, 2000; de Ruiter, 2000) claim that speech and gestures are processed independently from each other in a parallel fashion. According to these views, gestures are generated and processed directly and solely from the spatial and motoric representations whereas speech is generated from propositional representations and without interactions between the two during the production process. For example, according to Krauss, gestures are generated from spatial representations, "pre-linguistically" and independently from how certain information is linguistically (semantically or grammatically) formulated. On the other hand, an increasing amount of evidence supports the second view that claims that language and gesture production are two parts of the same cognitive system. As such, they interact with each other, but the level of this interaction and the stage in which it is available shows variance in different frameworks (e.g. Bergmann, Kahl and Kopp, 2013; Kopp, Bergmann and Kahl, 2013; Hostteter and Alibali, 2008; Mcneill, 2005, 1992; Kita and Özyürek, 2003; Goldin-Meadow, 2003).

One of the main theories for gesture production, called McNeill's Growth Point theory, claims that the origin of gestures is growth points (McNeill, 2008). In a growth point (GP), synchronized gesture imagery and linguistic categorical content together form an idea (in other words, they are mental packages that combine both linguistic categorical and imagistic components. According to the GSA framework (‘*Gesture-as-simulated-action’*, in Hostetter and Alibali, 2008), which is one of the frameworks focusing on how gestures are produced, gestures are a byproduct of speech. In particular, linguistic planning involves the simulation of visuospatial events; this activation during articulation is considered to be a source of speech accompanying gestures. To put it another way, when speakers activate a concept in order to express a meaning during production, they also activate perceptual and motor information related to that concept, just as activation of a meaning from language input during comprehension. Another framework, that is closely aligned with the GSA framework and that focuses on how gesture and language production are integrated, is the “Interface Hypothesis” (Kita and Özyürek, 2003). The preparation for the production of language requires the organization of rich and comprehensive information into small packages/processing units that contain an appropriate amount of informational complexity. According to the “Interface Hypothesis”, this processing unit may correspond to a clause in speech production. The content of a representational gesture is affected by the organization of these information-processing units, which are prepared for speech production. A recent cognitive modeling approach has been proposed by Bergmann, Kahl and Kopp (2013). According to underlying assumption of this approach, propositional and visio-spatial representations are activated and linked dynamically during pre-planning for language production. This simultaneous process on working memory facilitates memory organization (also see Kopp, Bergmann and Kahl, 2013) and give rise to semantic coordination of speech and gesture. All these suggest that gesture and speech complementary components of one communicational system bound together by semantic relations and temporal congruity to form a meaning (e.g. Wagner et al. 2014 see for a comprehensive overview).

3.3. The Role of Gestures

As mentioned in the very comprehensive review of the role of gesture by Goldin-Meadow (2010), gestures reflect the knowledge of the speaker. Speakers who produce gestures that convey information not contained in the speech (exhibiting gesture-speech mismatches) are more likely to benefit from instruction on that problem than speakers who produce redundant gestures. Moreover, gestures bring out implicit knowledge and have a role in reducing cognitive load. The positive effect of gesturing on learning is another important aspect. The results of a study conducted by Cook, Mitchell and Goldin-Meadow (2008) provided evidence that the children’s own hand movements help to cement what they had learned, suggesting that gesture can play a role by making learning last. Hostetter and Alibali (2008) found that the gesture has an enormous role in speech production due to its effect on translation of mental images into verbal output. The lexical access hypothesis (Krauss, Chen and Gottesman, 2000) indicates that gestures facilitate the retrieval of lexical items for spatial and motor ideas. In addition to its benefits for lexical access, the highlighting of features of images through gestures may also have benefits for the grammatical organization. When translating an image into a verbal stream, speakers must choose which aspects of the image to convey and in what order. The information packaging hypothesis (Kita, 2000) suggests that gestures help to break images down into smaller chunks that are suitable for verbalization. Support for this hypothesis has come from studies that systematically increase the difficulty of conceptualization for speaking without increasing the difficulty of lexical access (Alibali, Kita and Young, 2000; Hostetter, Alibali and Kita, 2007; Melinger and Kita, 2007). For example, Hostetter et al. (2007) found that participants gestured more when describing dot patterns that had to be parsed into separate geometric shapes than when describing the same dot patterns that had already been broken down into discrete shapes. There is also strong evidence that points out speakers gesture more when a verbal description is more challenging to plan or produce (e.g., Hostetter et al., 2007; Melinger and Kita, 2007; Alibali et al., 2000). Finally, highlighting mental images through gestures may also benefit speaking because gestures can bypass the need for verbally encoding imagistic ideas. Although most gestures occur with speech that conveys related semantic content, gestures sometimes stand alone or provide richer information than the co-occurring speech (e.g., Alibali, Bassok, Solomon, Syc and Goldin-Meadow, 1999; Church and Goldin-Meadow, 1986). For example, Melinger and Levelt (2004) instructed speakers to convey particular information about objects (e.g., size, shape) and found that speakers sometimes conveyed the target information in gestures and not in speech.

Furthermore, it has been suggested by Goldin-Meadow (2000) that "the information conveyed in gesture may not be conscious knowledge, but it is knowledge that speakers have at some level". This extraordinary feature of gestures makes them a very valuable starting point in the investigation of internal cognitive processes.

Gesture and language are two different representational formats, and each modality has its own advantages and disadvantages concerning the transfer of specific information (McNeill, 1992). The concepts might have various semantic features, which are represented mentally (Cassell, 1998). The different aspects of the same referent can be represented by different modalities, which highlight different perspectives. Gestures are analog representations, whereas language is a discrete representation, thus gestures can express thought that could be hard to convey in terms of language's categorical system (Goldin-Meadow, 2000; McNeill, 1992); they can, for example, express shape (analog representations) better than speech. As stated by Özyürek (2014), they can depict the event as a whole, describing the shape, manner and direction at the same time. On the other hand, language is sequential and the overall message to be conveyed is distributed in smaller meaningful phrases.

Not just relation between speech and gesture, but also their relation with the referent is also another crucial issue that needs to be touched upon. McNeill (2005) claimed that gestures refer to a referent (stimulus) and that they should be interpreted in relation to that referent instead of in relation to the verbal description of it. In his terms, "only in this way, we can observe co-expressiveness beyond redundancy. Speech content enables us to identify the event to which the gesture is related, but it is not the content against which the gesture is interpreted". The global and synthetic properties of gestures are similar to the global and synthetic properties of images (both physical and mental). Images, like gestures, convey meaning globally, such that the entire image's meaning influences the interpretation of each part. The isomorphism between images and gestures is manifested in their frequent co-occurrence. Gestures often occur when speakers express information in a way that evokes images. As stated in Özyürek et al. (2005), gestures can be claimed to share certain visual characteristics with pictures, they, however, do not have their exact semantic specificity because unlike pictures the full interpretation of gestures depends on the semantic content of the accompanying speech.

3.4. The Analysis of Gestures for HCI Research

This close relationship between gestures and language, which was elaborated in 3.2 and 3.3, makes gestures an effective tool in the assessment of the reader's conceptualization of events simultaneously being described verbally (Goldin-Meadow and Beilock, 2010). Although the interaction between language and gesture has been investigated for the past several decades in a variety of domains (Hostetter and Sullivan, 2011; Hostetter and Alibali, 2008; Goldin-Meadow, 2003; McNeill, 2005, 1992) specific investigations of graph comprehension in interaction with language and gesture have been one of the scarcely covered topics in the field of HCI and other relevant disciplines. Besides, it should be noted that so far little attention has been given to the effect of the sensory modality on gesture production.

As stated by Alibali (2005), spatial and motor information can be successfully carried over gestures, since they are relying on the same analog underlying mechanisms unlike discrete and propositional verbal codes. Mental imagery is claimed to be an embodied process that relies on simulation of perceptions and actions. As stated by Hostetter and Alibali (2008, p.500), “motor images are offline representations that utilize perceptual and motor systems” and they involve simulated action”; visual images are also offline representations but they involve simulated perception. According to the GSA framework, both language production and mental imagery employ the sensorimotor systems, and gestures are produced due to the activation in sensorimotor areas. Therefore, speech-accompanying gestures during post-exploration verbal descriptions of graphs may exhibit differences in regards to sensory modality of the exploration. More specifically, gestures produced for haptically perceived graphs are affected by motor-image based simulated actions, whereas gestures produced for visually perceived graphs are influenced by visual-image based simulated perception.

For communication through/via graphs, a graph reader first perceives the graph through a sensory modality (visual or haptic), and then communicates the information extracted from that graph based on how s/he conceptualized it. To that end, the graph reader refers to the conceptual event utilizing speech and/or gestures. So basically, graphs embody abstract representations of conceptual events and graph readers describe these abstract concepts over already embodied graphical entities. Therefore mental representations are neither solely originated from abstract representations nor from concrete representations. The sensory modality, language and gestures exhibit multiple and intertwined interactions among each other. Both gestures and graphical communication are visuo-spatial modalities sharing similar perceptual visuo-spatial features that convey meanings such as quantity, direction and relations (Tversky, 2011). When describing some visualization with an accompanying gesture, the places (or punctual events in the domain of our interest) become “fleeting positions” whereas marks and forms on the visualization become “fleeting actions” (Tversky et. al., 2009). Following this idea originating from the resemblance between the two modalities, the vocabularies of gestures, speech and diagrams can be considered as parallel (Tversky, 2011). For instance, in the context of communication over graphs, a fluctuating increase in a line graph may be verbally described by the term “increase” and it may be simultaneously accompanied by a gesture that represents the fluctuation in that increase. One of the studies focused on communication through line graphs (Acartürk and Alaçam, 2012) showed that the perceptual features of the graphical cues that highlight certain aspects of the visualization (e.g., a graphical cue such as an arrow) influence the conceptualization of the presented information. This effect was observable in the gestures produced by the graph readers. The results of that study indicated that in order to emphasize process concepts (e.g., increase, decrease) humans produced more vertical and diagonal gestures, whereas more pointing gestures were produced in order to emphasize punctual state concepts (e.g., a peak). Similar findings were obtained from a comprehension perspective, in the sense that vertical and diagonal gestures were efficient in conveying information about processes.

In summary, the role of gestures is to provide additional information that enhance comprehension and resolve ambiguities during the course of communication. They constitute a convenient tool for carrying spatio-temporal information. Besides, they highlight or complement the information presented by another. Because gestures provide additional information aligned with the visuo-spatial aspects of communication through graphs, their analysis can be considered to be a practical tool for assessing how a graph reader interprets a graph and conceptualizes the processes, events and states represented by it. In addition, gesture analysis helps to detect the hard-to-encode information and the disambiguities, which are generally highlighted with the presence of accompanying gestures. All these characteristics and relations make gesture a unique tool in understanding the conceptualization of spatial / shape-based representations. The contributions of gesture analysis as a HCI methodology in that respect are discussed in detail in upcoming chapters with empirical support from the experimental investigations.

One of my claims in this dissertation addresses why the investigation of gestures might be considered as a very effective research paradigm not only for cognitive scientists or psycholinguists but also for HCI specialists. Here, the distinction between the use of gesture as a methodology and incorporating them in human computer interfaces should be made clear. Therefore, I would like to briefly touch upon the current state of gesture use in HCI domain. The gesture research is one of the areas that catch exponential attention over the last two decades by HCI specialists, more specifically in Human-(virtual) Robot Interaction. Human communication involves many modalities, such as language, gestures, eye gaze and facial expressions. For generation human-like agents, understanding those behaviors and generating them in a dialogue is a crucial aspect that need to be dealt with thoroughly. The agents that can perform such behaviors are called as embodied conversational agents (Pelachaud, 2005; Kopp, Jung, Leßmann, and Wachsmuth, 2003; Cassell, 2000; Cassell, Bickmore, Campbell, Vilhjalmsson and Yan, 2000), and they can engage in face-to-face conversations with the user, perform joint actions and demonstrate those communicational modalities as humans exhibit in a natural conversations. In addition to their contribution in sustaining naturalness in conversation, the inclusion of those behaviors has other merits as well. Limiting communication to a single mode requires completeness of that mode and spontaneous spoken words typically lack the completeness since we use the language in corporation with other modalities to communicate what we want to convey (Tversky, 2009, 2011). Therefore it is also important for the robotics research to create systems first that can understand its human partner's gesture, and the relation between the speech and gesture. Secondly, the robots should be able to produce gestures to sustain efficient communication as well. The human computer interaction research usually incorporates gesture generation and gesture recognition from those perspectives.

Gesture provides referential links between the space and the speech, and the use of deictic references such as pointing gestures are important elements of human-human communication, especially in an assistive and educational domains, unfortunately these referential links carried through visual modality are not accessible for visually impaired people. A system developed by Oliveira, Quek, Cowan, and Fang (2012) provides a sensory substitution for those crucial deictic gestures in an instructional classroom environment. In this system, blind students wear a glove that can provide vibrotactile feedback. Both the instructor's pointing behavior, and the student's the position of the reading hand are tracked by cameras. When the instructor point a diagrammatic representation on the board, the distance between the location of pointing and the location of the user's hand and the direction (north, northeast, etc.) are calculated and a vibrotactile feedback is given to the user's hand to direct him/her to where the instructor is pointing. This system allows the instructor act more naturally and loads some of the information (such as direction, location) into the gesture modality that can successfully carry those.

Another example of the use of gesture for reference generation and understanding is presented by Giuliani, Foster, Isard, Matheson, Oberlander, and Knoll (2010). In their setting, a user and a robot work together to assemble wooden construction toys on a common workspace. In order to perform the required actions successfully, they coordinate their actions through speech, gestures, and facial expressions. In this case, the robot has a role of assistant and it knows the order of the actions that need to be done to build the toy. In this case, the comprehension of the gestures produced by the human user (the technical term is *gesture recognition*) is a key to detect unexpected actions (such as errors) of the user and to correct them accordingly.

It is also possible to incorporate gestures for virtual agents that provide human users with assistance. Max (the "multimodal assembly expert") developed by Kopp and his colleagues (2003) is a multimodal assistant, which is able to produce speech, gestures (both deictic and iconic), eye gaze and emotional facial expressions in virtual reality environment. The main task of this agent is to assist the user in virtual construction tasks. In this system, information is distributed among modalities; to illustrate, instead of providing only verbal instruction for the required action to complete a step, the virtual agent Max demonstrates visually what needs to be done by performing iconic gestures.

Within the scope of this dissertation, I employ different approach from those above-mentioned aspects concerning the gesture analyses. Here, rather than focusing on gesture generation and recognition, I prefer to use the gestures produced by the user as a *research methodology* to assess their conceptualization of the event. Using gestures for this purpose is not a common method in the domain of HCI. However, especially considering the underlying common mechanism of both gesture production and haptic exploration (the activation in motor cortex, and the construction of motor (mental) images), I strongly consider that they might help us to explain the differences and similarities between the two modalities.

Although my approach in this dissertation varies from the abovementioned HCI research fields that focus on gesture generation and recognition, the insights resulting from my research has possibility to converge with them at a point. As a part of the experimental paradigm employed, the visual or haptic graph readers produced single sentence verbal descriptions of the abstract events depicted in the statistical graphs and also demonstrated spontaneous speech accompanying gestures. This richly annotated – multimodal (incorporating gesture-speech and also referents' qualitative and quantitative characteristics) data may also provide valuable information that can be utilized in a design of such assistive systems that employ human- (virtual) robot setting. The experiment conducted here is limited with the statistical line graphs, however the graphs have variety of shape properties resulting different use of gradable modifiers regarding size, shape, orientation etc., and temporal/event denoting expressions (see Chapter 6 for more detailed information). Therefore, the results may also provide solid basis for an HCI system that incorporates communication over other types of shape based (diagrammatic) representations.

CHAPTER 4

Experimental Paradigms and Methodology

This chapter aims to provide a full overview of the experimental setup and experiments employed within the scope of this dissertation. The research questions regarding haptic graph comprehension and the dynamics of collaborative environment for designing a verbal assistance system were investigated by conducting four experiments and by employing two different experimental paradigms. In part 4.1, I present the overview of the experimental paradigms. The equipment and the stimuli utilized in those experiments are presented in 4.1 and 4.2 respectively. Next, the experimental procedures are explained in 4.4. Final part (4.5) summarizes the annotation schemes and the procedure of data coding.

4.1 Experimental Paradigms

In this dissertation, two experimental paradigms were conducted. The single-user paradigm, where only one user explore the graph visually or haptically without having an assistant during exploration session, was employed in three experiments (the experiments I, II and III). In dual-user paradigm, one haptic explorer and one verbal assistant were employed for each session, and two participants with different activity roles perform a joint activity for graph exploration. This paradigm was employed in one experiment (the experiment IV). The details of the experimental procedures are presented in the following. The experimental procedures for both paradigms can be divided into three phases; the procedures, which are followed (i) before the exploration, (ii) during the exploration and (iii) after the exploration, are presented in tables below. Table 4-1 and Table 4-2 illustrate the procedures employed in *the single-user paradigm* and in *the dual-user paradigm* respectively. The outputs that are obtained through these procedures are also given in lower part of the table (titled as *Output*). The details of each step are elaborated in the upcoming "*Empirical Studies*" part. For the rest of the dissertation, *the icons (mnemonics)* are presented next to each experimental method in order to give quick organizational cues for a reader.

Table 4-1 Overview of the experimental paradigm-I (a single-user study)

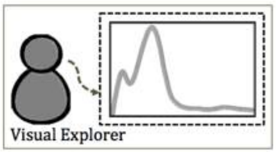
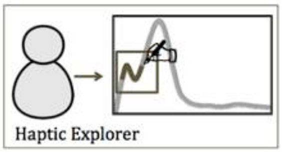
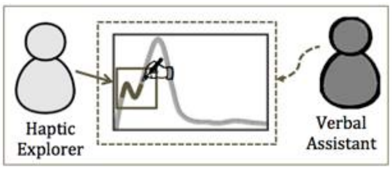
Before the exploration	**During the Exploration**	**After the Exploration**
The participants were asked		
— to fill out the consent and demographic data forms — to complete a warm-up session — to read the instructions	— to explore the graph ▪ visually  or ▪ haptically 	— to describe the graph to a hypothetical audience — to draw the graph on a paper
OUTPUT		
Consent and demographic data form	Haptic exploration actions of the explorers Eye Movement data of the visual readers	Verbal descriptions (all) Speech-accompanying gestures (all) Sketches (all) Survey (Exp-II)

Table 4-2 The overview of the experimental paradigm-II (a dual-user study)
(E: explorer, A: verbal assistant)

Before the exploration	**During the Exploration**	**After the Exploration**
The participants were asked		
— to fill out the consent and demographic data forms (both *E and A*) — to complete a warm-up session (both *E and A*) — to read the instructions (Exp.: all) (both *E and A*)	— to explore the graph jointly (*E* was instructed to explore the graph and ask for verbal assistance when needed, whereas *A* was instructed to provide verbal assistant shortly and plainly, when asked by *E*.) 	— to describe the graph to a hypothetical audience — to draw the graph on a paper
OUTPUT		
Consent and demographic data form	Dialogue acts by both E and *A* Haptic exploration actions of *E* Eye Movement data of *A*	Verbal descriptions (All) Speech-accompanying gestures Sketches

4.2 Equipment

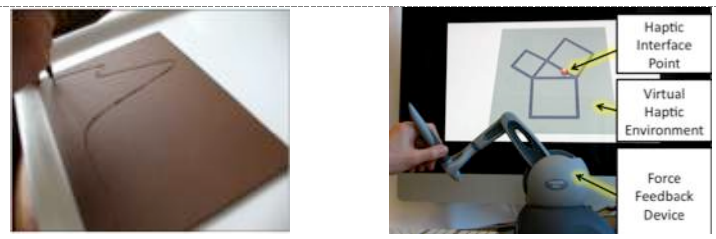
4.2.1 Phantom Omni Haptic Device

Haptic graph exploration is performed by moving the handle of the Phantom Omni haptic device (recently renamed as GeoMagic Touch). The pen-like handle of the device is attached to a moveable arm, which can be considered as a reverse robotic arm. The arm can be moved in all three spatial dimensions (with six degree-of-freedom). Figure 4-1 illustrates the moving parts and the directions of the movements for each part. The device contains three parts; the "body", "shoulder", and "elbow". The body part (or can be also referred as turret, see Phantom Omni User Manual for more information) is responsible for performing right and left motions. The shoulder part is responsible for up and down movements of the pen, whereas the macro elbow motion is associated with in/out movements. The shoulder and elbow parts can also perform micro scale movements (see Figure 4-1b), but those movements are irrelevant for the exploration of a haptic graph. The exploration of a graph on a 2D virtual surface (since the depth is irrelevant and kept constant as much as possible) is mainly performed by the actions on the body and on the shoulder indirectly through the handling of the pen. However due to cyclical movement of the body, the stylus follows an arc-shaped trajectory for the very right and very left sides (this issue is elaborated under Chapter 5.1.1). This may lead to a misreading of line based representations therefore the haptic graph workspace (384 x 228 pixels in 2D) was limited to a region (190 x 90 pixels) where linear exploration can be performed without having this side-effect.

(a) (b)

Figure 4-1 (a) Macro movements and (b) micro movements of the Phantom Omni arm and body (retrieved from the Official User Manual)

The Phantom Omni haptic device allows for one-point interaction. In haptic graph representation, the graph (the line of the line graph) was represented by engraved concavities on a horizontal plane; therefore the graph readers perceived the line as deeper with respect to other area on the surface. This resembles to exploration of a physical haptic map depicted in Figure 4-2a.



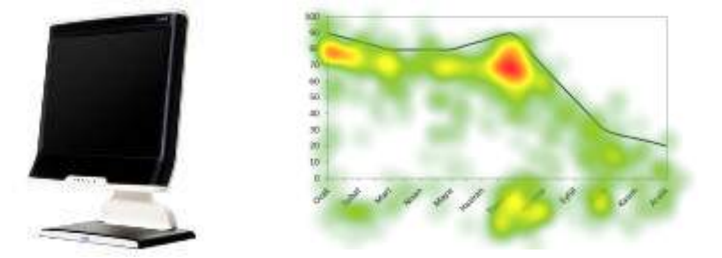
(a) (b)

Figure 4-2 (a) Exploration of a physical haptic map and (b) visualization in a geometry domain (see, Kerzel and Habel, 2013, Fig.1)

Exploration based system that allows one point interaction has important advantageous in terms of graph exploration. This exploration process particularly foregrounds the shape of the graph line and the changes on the graph line unlike other solutions such as raised-dot tactile displays that allow exploration with two hands or vibro-tactile gloves (e.g., Zeng and Weber, 2010 or Oliveira et al.2012). Furthermore, the Phantom Omni device is commercially available and comparably low priced. It also allows the system to know the location of the stylus, therefore it enables the system to react with respect to users' current state. Besides, the exploration patterns of the users (such as increase in the back-and-forth movements, or decrease in the velocity of the movement) can be also tracked. Although it was not intended and addressed within the scope of this dissertation, the haptic graphs would be generated automatically by employing software that converts them into haptic graphs based on numerical data. This kind of solution may lead to a fast and instantaneous generation of virtual haptic graphs. In this dissertation, instead of generation of haptic graphs, my empirical research focuses on the issues regarding the design of automatic verbal assistance. Namely, I focus on the content and timing of the verbal assistance than can be generated automatically based on user's exploration movements, and the perceptual/conceptual significance of the graphical entities (such as shape landmarks and shape segments). All qualitative ascriptions which constitute the base of the verbal assistance can be derived automatically from the quantitative data, namely data point values.

4.2.2 Tobii T120 Eye tracker

Regarding the experiments III and IV (see the upcoming parts in this chapter for the details of the experiments), the visual graphs were displayed on a Tobii non-intrusive 120 Hz eye tracker (T120, see Figure 4-3), integrated into a 17" TFT monitor with a resolution of 1024x768 pixels. The spatial resolution and the accuracy of the eye tracker were 0.25˚ and 0.50˚ respectively. Analyzing eye-tracking data is one of the frequent methods employed in the investigation of graph comprehension. The eye tracking data were obtained and analyzed by using Tobii Studio Software. In the experiments, in which I collected eye movement data, I aimed to investigate the effect of data labels on graph comprehension by looking at the gaze patterns. However, since this dissertation has emphasis on the haptic graph comprehension, the results obtained from this methodology was left out for further publications, and this methodology was just reported here as a part of the experimental setting.



(a) (b)

Figure 4-3 (a) Tobii T120 Eye Tracker (b) Heatmap visualization

4.3 Stimuli-Sets

Three types of graphs were presented to the participants in a between subject design. In the first condition (see Figure 4-4 (left)), the participants explored line graphs haptically (without data labels). In the current state of art of haptic line graph design, providing haptic information on the graph axes, such as information regarding the numerical labels and the axis title, is hard to implement and they would be rather distracting. Therefore the haptic graphs, utilized in this experiment did not contain the data labels. In the second condition (see Figure 4-4 (middle)), the graphs with data labels were presented on a computer screen, thus the participants had visual access to the graphs. In the third condition (see Figure 4-4 (right)), the participants inspected the visual graphs without data labels. This condition served as a control condition for both former conditions. The effect of the sensory modality in the investigation of various graph comprehension related research questions was explored by comparing the condition-1 and the condition-3 and the data belongs to the visual and the haptic graphs without data labels are called *Modality-Group*. The effect of data labels was investigated by comparing the condition-2 and the condition-3 and the data that belongs to these two groups were named as *Label-Group.*

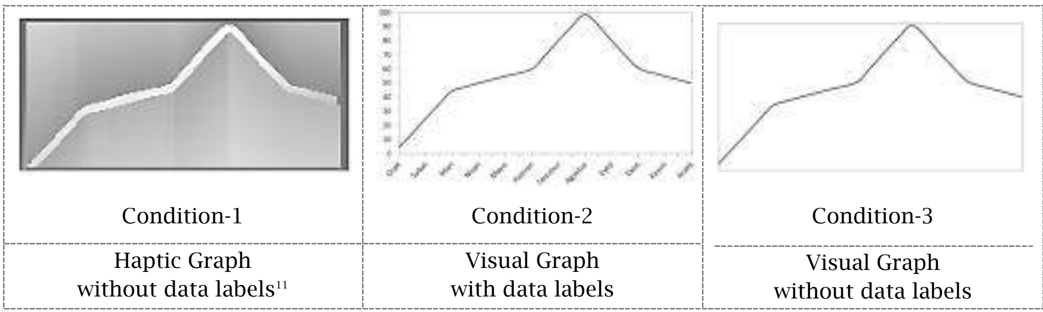


Figure 4-4 Sample Graphs for each condition

In addition to providing different conditions for controlling the sensory modality and the presence of data labels, the contribution of amodal geometric properties were also investigated by employing two different sets of line graphs.

4.3.1 Graph Set I: Bird-Population Domain

The first graph set contains line graphs with smooth edges taken from a publicly available consensus report (PRBO, 2012) about bird-population distributions and they were redesigned for the purposes of the study (see Figure 4-5 and Figure 4-6 for all graphs in this set). While two graphs were employed for the familiarization part, five graphs were presented in the experiment session.

Each graph had a different pattern in terms of the number and polarity of curvature landmarks, length and direction of line segments, as shown in Figure 4-6. The parameter of polarity covers the concepts of minima and maxima. In more detail, the salient curvature landmarks of *positive maxima* and *negative minima* of curvature (Cohen and Singh, 2007) depend on the concepts of *convexity* and *concavity* on contours in closed curves. However, the line of the statistical line graphs is a directed open curve. Based on this, the convexity and concavity were determined with respect to horizontal axis and elaborated under the umbrella notion of polarity in this dissertation. This graph-set was employed to investigate the modality-dependent differences as well as the effect of the presence of the data labels. However, the shape of the graph and the shape properties were not controlled intentionally in this set.

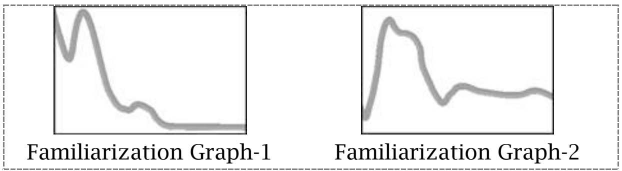


Figure 4-5 Familiarization Graphs

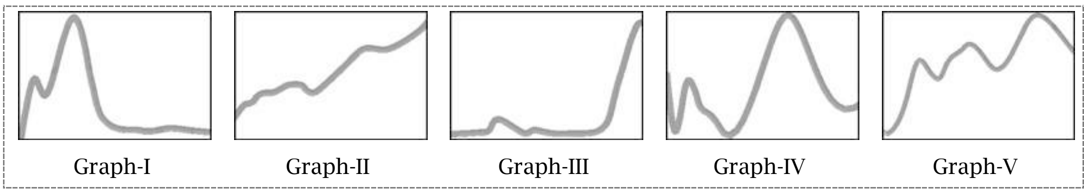


Figure 4-6 Five Stimulus-graphs in Graph Set-I

[11] The screenshot of the 3D haptic graph

4.3.2 Graph Set II: Tourist-Visit Count

The line graphs in this set depict averaged monthly tourist visits for different cities in Turkey. Unlike the previous graph-set, the steepness of the line segments is one of the systematically controlled parameters in this graph set. Steepness of the lines (a within-subject variable) w.r.t x-axis were set as 0˚ (no change), 15˚ (a slight change), 45˚(a moderate change) or 75˚(a steep change) in both direction (namely an increase or decrease). In addition to the segment's properties, the turning angle at the intersection of two line segments in varying steepness values was also controlled. This graph set was used in the Experiment-III in order to investigate the effect of amodal geometric properties on event segmentation and description. More detailed discussion about the graphical features is presented under the corresponding chapter (Chapter 6.4.2).

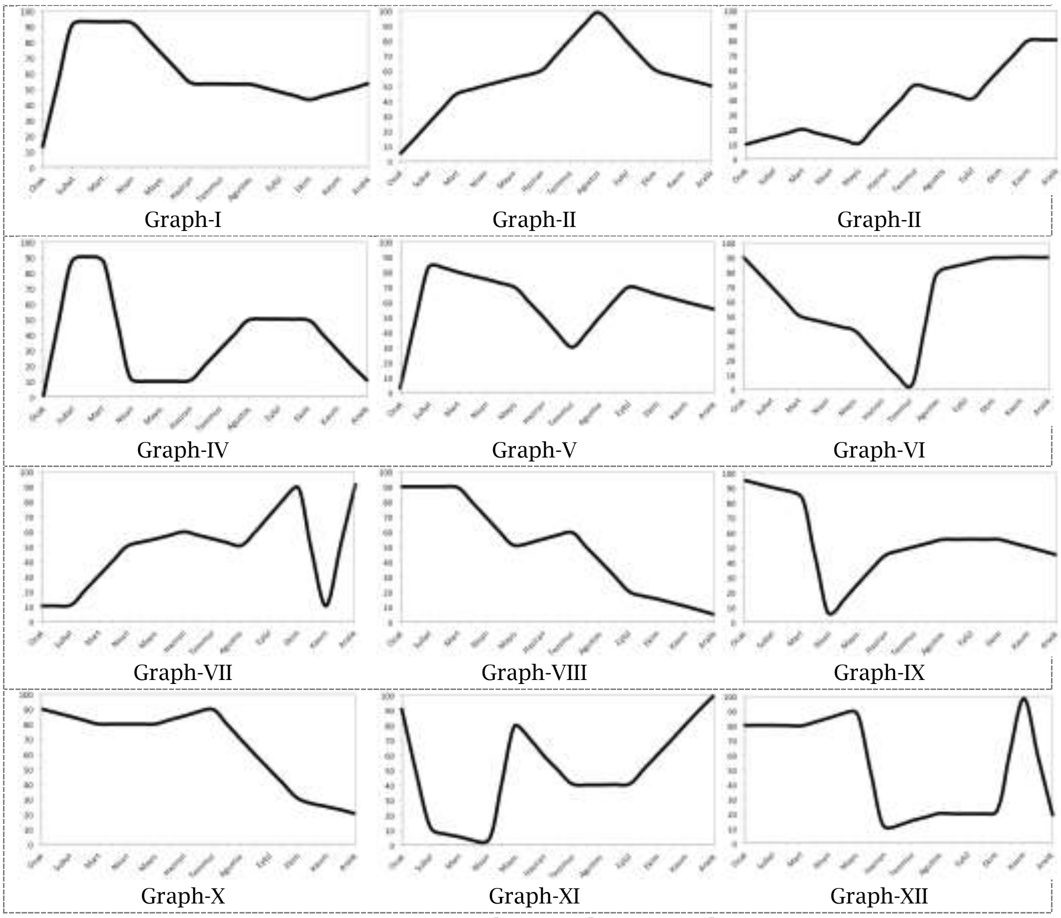


Figure 4-7 12 Stimulus-graphs in Graph-Set II

4.3.3 Smoothing Method

In this experiment, smoothed line graphs were employed as stimuli. First reason for applying a smoothing is that the smooth line graphs are one of the frequent types of data visualization due to their efficiency in highlighting trend information. The PRBO database created by domain experts also chose this representation to illustrate their results about bird population distributions. Another reason is that they are more suitable for haptic graph representation. Haptic lines are consisting of voxels, which are 3-dimensional units. As also stated in FreeFrom Software Manual (a software for developing 3D surfaces for Haptic Omni device), it is not possible to achieve infinitely sharp edges, as 2D representations are able to achieve. There always occurs a natural smoothing due to how the haptic rendering procedure works. For this reason, it is not feasible to target non-smooth edges in haptic line graphs due to technicality. Still, in the design phase of the stimuli set, both haptic graphs created based on non-smooth and smooth 2D line representations were compared and they did not exhibit considerable difference in terms of haptic exploration. Due to this restriction in haptic modality, the smoothed visual counterparts were employed in visual modality as well for the sake of sustaining systematic analysis.

The smoothing method also seems very crucial for successful and accurate data visualization. Smoothing is a distortion process, where a designer (of the graph) makes a conscious sacrifice from the precision of the data value in order to highlight the extraction of the second order information such as relations and trends. However, the amount of distortion and also the dimension in which the distortion is applied are important. Figure 4-8(a) illustrates two lines graphs, both of which were created automatically by MS Excel from data tables. The dark line corresponds to non-smooth line, whereas the lighter line is the smoothed one (by Excel smoothing algorithm). As can be seen from the non-overlapped line segments, the distortion affect the y-value of February and makes it higher, but also more importantly it misrepresent the data point for February as a global maximum point. In the original data, both February and March have same y-values, and this kind of distortion on conceptual domain can severely affect the accurate comprehension of the abstract event depicted in the graph. Therefore, to avoid this kind of misrepresentations, semi-manual smoothing method was applied. According to this method, the average value between each consecutive data points were calculated, and added as a data point. This step reduces the amount of distortion. In the second step (Figure 4-8 (b)), the distortion in y-value is partially spread over to x-axis (by manual manipulation of the handle) resulting more similar graphs compared to the original (Figure 4-8(c)). Figure 4-8 depicts the overlapped version of the two graphs; the non-smooth and the semi-smoothened graphs.

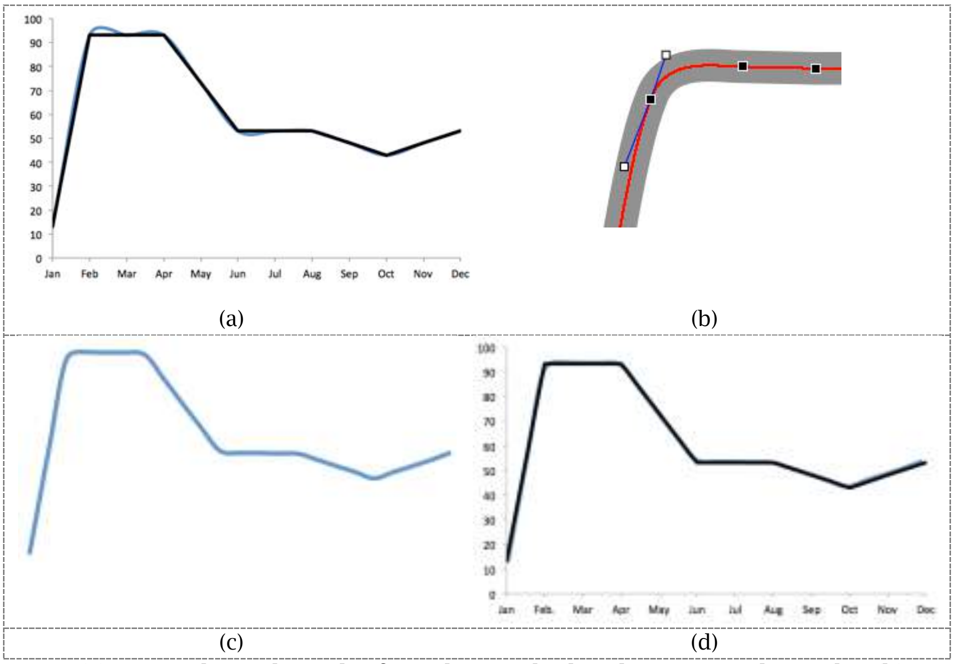


Figure 4-8 (a) an overlapped graph of excel smoothed and non-smooth graphs, (b) manual manipulation method for smoothing, (c) only manually smoothed graphs, (d) overlapped graphs of manual smoothed and non-smooth graphs,

4.4 The Empirical Studies

In this part, I aim to provide the overview of the four experiments presented within the scope of this dissertation. In order to maintain coherence and provide ease of reading, the information provided here was also summarized in "*the experimental setting*" parts where I provide the comprehensive analyses and results regarding each investigated domain in the upcoming sections. Therefore readers are encouraged to skip or look through this part (*4.4*) and come back if more information than the provided in the corresponding section is needed.

4.4.1 Experiment I: Differences and similarities between visual and haptic modalities

For designing haptic graphs augmented by audio assistance (sonification or speech), it is necessary to determine which information depicted by the graph or by the segments of the graph are appreciated as important. In this experiment, the visual and haptic graph readers explored the graphs in a single-user experimental setup. The aim of this experiment is to investigate the effect of sensory modalities and the effect of data labels on graph comprehension.

Participants. In the first experiment, 31 undergraduate and graduate students ($Mage$=25.42, SD=5.05, 15 female) from University of Hamburg participated for course credit or participation fee. The experiment was conducted in German, the native language of all participants. Haptic explorers were blindfolded sighted people.

Procedure and Design. The experiment was conducted in three conditions in a between-subject design (see Figure 4-4) by employing *Experimental Paradigm-I* presented in detail in Table 4-1. The *graph set-I* was presented to the participants. In the first condition, *"haptic graphs without data labels"*, (9 participants, 4 female), the participants explored line graphs haptically. All haptic explorers were right-handed and they used the haptic device with their dominant hand. In the second condition, *"visual graphs with data labels",* (11 participants, 3 female), were presented on a computer screen, thus the participants had visual access to the graphs. In the third condition, *"visual graphs without data labels"*, (11 participants, 8 female), which served as a control condition, the participants inspected the visual graphs without data labels.

Each session (including warm-up & instruction sessions, exploration processes and post-exploration tasks) took approximately 1 hour. The sessions were audio and video recorded. Before starting warm-up session, all participants were asked to fill out a demographic data form, which contains basic questions regarding age, gender, educational background and status, native language and handedness, see *Appendix A* for a sample of a demographic form in German and in Turkish. Accordingly, each participant signed informed consent form (see *Appendix B*).

— ***The Warm-up session***

Before starting haptic graph exploration, all participants completed a warm-up session that aimed to familiarize the participants with Phantom Omni® Haptic Device. In this warm-up session, the participants explored the character set consisting of two letters (L and G) and two digits (2 and 3). Those characters were chosen because of their resemblance to graphs with their edges and curves. Additionally, in order to assess the participants' comprehension, the characters were presented in different rotations (normal, left rotated, right rotated and upside down), and then they were asked to tell which character it is and how it is oriented. The experiment session started only if the participants succeeded in at least 9 trials.

— ***The instruction session***

After the warm-up, the participants were informed that the graphs represented populations of bird species in a lagoon and also about post-exploration tasks. The instructions were given in paper. Gesturing was not mentioned in the instructions.

— ***The exploration session***

In the experiment session, each participant was presented five line graphs (two additional graphs were employed for the familiarization part, Graph Set-I, see Figure 4-6). The graphs were presented in random order. The participants in haptic condition did not have time limitation, while visual explorers presented the graph for 10 sec.

— ***The post-exploration tasks***

After the exploration process, the participants were asked to imagine themselves in an online meeting, in which their task was to present single-sentence summaries of annotated graphs to the audience. According to the scenario, the audience was able to see the participant (i.e., the presenter) but not the graphs. In studying communication, using imaginary audience is a standard and frequent experimental procedure (Goudbeek and Krahmer, 2012). As shown by several studies, the differences between employing real and imagery audiences are small (Van der Wege, 2009; Ferreira, Slevc and Rogers, 2005). After they complete their descriptions, the participants were also asked to draw a sketch of the graph with paper and pencil (see *Appendix C*).

4.4.2 Experiment II: A Pilot Study with visually-impaired users

Many factors take part in the graph comprehension, ranging from bottom-up perceptual influences to high-level top-down effects. One of the most important top-down effects is the role of prior knowledge about how to read a specific graph type, namely facilitation of appropriate graph schemata. In the previous study, I employed blind-folded sighted participants in the investigation of haptic graph comprehension in order to keep this factor constant as much as possible. In this experiment, the haptic line graphs were explored by visually impaired participants.

Participants. Eight visually impaired participants (three partially sighted) participated in the experimental study. The participants were teachers from an elementary school for visually impaired, in Turkey. The category of blindness is identified by various parameters, such as onset-time of blindness (congenital-early or late) and level of deprivation (total, severe blind or partial sighted), (Vecchi and Cattaneo, 2011). Late-blind participants are also of interest since they have domain knowledge, and mathematical skills acquired via visual modality. However, in this experiment, only congenitally blind participants with or without partial sight were employed. The experiment was conducted in Turkish, the native language of all participants.

Procedure and Material. Each session (including warm-up & instruction sessions, exploration processes and post-exploration tasks) took approximately 1 hour. The sessions were audio and video recorded.

— ***The Warm-up session***

Same warm-up session that was employed in the previous section was used.

— ***The instruction session***

After the warm-up, the participants were informed that the graphs represented populations of bird species in a lagoon and also about post-exploration tasks. The instructions were read to the users by the experimenter.

— ***The exploration session***

In the experiment session, each participant was presented five haptic line graphs with smooth edges (two additional graphs were employed for the haptic graph familiarization), see Graph Set-I. The graphs were presented in a random order. The numerical labels were not represented, therefore the participants were asked to comment on the graph based on the line shape only. The participants did not have time limitation. Only basic assistance or alerts that aimed to help the participants to locate him/herself on the reference frame (e.g., “you are at the start point”, “you are out of the line”) was provided by the experimenter during the course of the haptic exploration.

— ***The post-exploration tasks***

After the experiment session, the participants were asked to present single-sentence verbal descriptions of the graphs to the experimenter. They were also asked to fill out a spatial term survey, which is also read by the experimenter (see *Appendix D*). The survey contained the list of words or phrases uttered by the participants in other experiments (the Experiment I and IV). The participants used a Likert scale to rate the meaningfulness of the terms, using 1 (less meaningful) to 5 (most meaningful). They were also asked to verbally explain the meaning of the terms in the questionnaire.

4.4.3 Experiment III: Investigation of Event Segmentation and Description

In this experiment, I focus on two main questions; (i) how the graph readers parse continuous stream of data into meaningful discrete events by using perceptual salience or conceptual significance, namely "event segmentation" (ii) in which circumstances they prefer to use or not to use gradable expressions (i.e. adverbial and adjectival modification), namely "event description". Furthermore, the amodal graphical properties (such as slope and angle) were systematically controlled to explore their effect on event segmentation and description.

Participants. Forty-eight participants (*Mage*=24.69, *SD*=5.85, 23 female) from Middle East Technical University participated to this experiment. 3 of the haptic explorers were left-handed and all subjects used the device with their dominant hand. The experiment was conducted in three conditions in a between-subject design. The experiment was conducted in Turkish, the native language of all participants. Each participant signed informed consent form. Haptic explorers were blind-folded sighted people.

Procedure and Material. In the first condition, the participants explored line graphs haptically (without data labels). In the second condition, graphs with data labels were presented on a computer screen, thus the participants had visual access to the graphs. In the third condition, the participants inspected the visual graphs without data labels.

Each session (including warm-up & instruction sessions, exploration processes and post-exploration tasks) took approximately 1 hour. The sessions were audio and video recorded.

— ***The Warm-up session***

Same warm-up session that was employed in the previous section was used.

— ***The instruction session***

After the warm-up, the participants were informed that the graphs represented populations of bird species in a lagoon and also about post-exploration tasks. The instructions were given in paper. Gesturing was not mentioned in the instructions.

— ***The exploration session***

In the experiment session, each participant was presented 12 line graphs (two additional graphs were employed for the familiarization part). The line graphs with smooth edges presented averaged monthly tourist visits for different locations (the Graph-Set II, see Figure 4-6). The graphs were presented in a random order. The participants in the haptic condition did not have time limitation, while visual graph readers had 10 sec for visual inspection.

— ***The post-exploration tasks***

After the exploration session, both participants were asked to present single-sentence verbal descriptions of the graphs to a hypothetical audience. They were also asked to draw a sketch of the graph with paper and pencil.

4.4.4 Experiment IV: A Verbally Assisted Graph Comprehension as a Joint Activity

This experiment that employs a dual-user experimental setup was conducted to explore the dynamics of the joint activity, such as the effect of verbal assistant on graph comprehension, and alignment in the interlocutors' (a haptic explorer (*E*) and a verbal assistant (*A*)) reference frames.

Participants. In the experiment, thirteen sighted-participant pairs (an explorer and an assistant, Mage=25.3, SD=3.27, 16 Female) from Middle East Technical University collaborated in exploring haptic line graphs. Only 1 haptic explorer was left-handed and the decision of which hand is used for the exploration was left to the participant, and he preferred to use his right-hand to handle the device. The experiment was conducted in Turkish, the native language of all participants. Haptic explorers were blind-folded sighted people.

Procedure and Design. Experimental Paradigm-II was employed; **e**ach session (including warm-up, instruction sessions, exploration processes and post-exploration tasks) took approximately 1 hour. The sessions were audio and video recorded.

— ***The Warm-up session***

Same warm-up session that was employed in the previous section was used.

— ***The instruction session***

After the warm-up, the participants were informed that the graphs represented populations of bird species in a lagoon (Graph Set-I) and also informed about the post-exploration tasks. The instructions were given in written format on paper.

— ***The exploration session***

In the experiment session, each participant was presented five haptic line graphs (two additional graphs were employed for the familiarization part). The numerical labels were not represented. The line graphs represented populations of bird species. The graphs were presented in random order to each pair. Each participant in the pairs was located in separate rooms so that they communicate through speakers without visual contact. *E* explored the graph haptically. During the course of exploration, *A* was able to display the graph and the current location of *E*'s exploration (as a moving point on the visual graph). *E* and *A* did not have an identical task space. Only haptic explorer´s current location of exploration was available to the verbal assistant's screen. Therefore, pointing was possible only for the haptic explorer. Both participants in each pair explored informationally equivalent graphs, except for the difference in the modality of presentation (haptic and visual). Another difference between the two participants was that *E* was instructed to explore the graph and ask for verbal assistance when needed, whereas *A* was instructed to provide verbal assistant shortly and plainly, when asked by *E*.

— ***The post-exploration tasks***

After the experiment, both participants (*E* and *A*) were asked independently to present single-sentence verbal descriptions of the graphs to a hypothetical audience. They were also asked to draw the graphs on a paper.

4.5 Data Analysis and Coding

As discussed before, the data were gathered through several experimental methodologies. Before going into details of each of these methodologies concerning the procedure of coding and analysis, first I briefly present the segmentation procedure applied on the graph shape to extract qualitative ascriptions.

4.5.1 Qualitative Ascriptions of Line Graphs

In order to provide verbal assistance regarding graph shape, first the perceptual and conceptual features of the entities (the shape landmarks and shape segments) that construct a graph shape should be extracted from the raw numerical data values. The geometric shape concepts for describing graph lines are exemplified with a graph used in the experimental studies (see Figure 4-9).

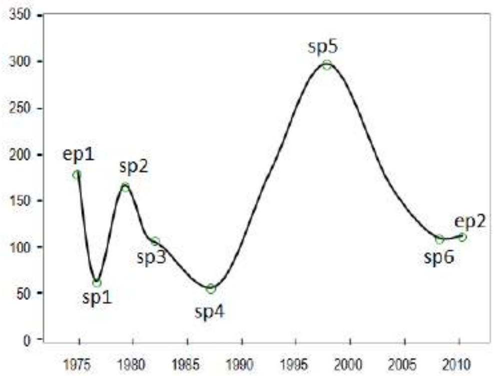


Figure 4-9 Shape Landmarks of Graph-IV (in the graph-set I)

The line graphs consist of data points (see Table 4-3a) and these information is usually enough to draw conclusions about the perceptual and conceptual saliencies (Table 4-3b). Besides, the relevant qualitative ascriptions can be also derived from this raw data with simple calculations. More detailed information is given in Chapter 5.2.

Table 4-3 Quantitative values used for the derivation of qualitative ascriptions

(a)

	Y-Value	X-Value
ep1	175	1975
sp1	65	1977
sp2	160	1980
sp3	110	1982
sp4	55	1988
sp5	300	1998
sp6	110	2008
ep2	120	2010

(b)

	Value	Shape Landmark
Max. Y-label Unit	350	NA
Min Y-Label Unit	50	NA
ep1-ep2	-55	NA
Gmin*	55	sp4
Cmin*:	65	sp1
Gmax.	300	sp5
Cmax*.	175	ep1
Cmin-Gmin	10	NA
Gmax. - Cmax.	125	NA

*Gmin: Global Minimum, Cmin: Closest value to Gmin, NA: not applicable

Graph lines require some additional shape representations and shape cognition characteristics beyond the characteristics of contours. In particular, graph lines are conventionally oriented corresponding to reading and writing direction and they are comprehended with respect to an orthogonal system of two axes. The haptic graphs we use in the experiments are realized in a rectangular frame that induces an orthogonal system of axes. Table 4-4 gives a tabular summary of qualitative representations for selected shape landmarks and induced line segments. The functional character of statistical line graphs leads to the prominence of value extrema (in contrast to curvature extrema of contours). Since we use in the experiments presented here smoothed graphs, these extrema are called smooth points (sp)).

Table 4-4 Qualitatively described shape landmarks and shape segments

Shape landmarks		
Landmarks	**Landmark characteristics**	**Global properties**
ep1	left end pt., local max.	Higher than sp1, sp2, sp3, sp4, sp5, ep2
sp1	smooth pt., local min.	Higher than sp4
sp2	smooth pt., local max.	Higher than sp1, sp3, sp4, sp5, ep2
sp3	smooth pt.	Higher than sp1, sp4, same height as sp6
sp4	smooth pt., local min.	Global min.
sp5	smooth pt., local max.	Global max.
sp6	smooth pt., local min.	Higher than sp1, sp4, same height as sp3
ep2	right pt., local max.	Higher than sp1, sp3, sp4, sp5

Shape segments		
Segments	**Shape characteristics**	**Vertical orientation**
ep1–sp1	curved	steeply downward
sp1–sp2	curved	steeply upward
sp2–sp3	curved	steeply downward
sp3–sp4	curved	diagonally downward
sp4–sp5	curved	steeply upward
sp5–sp6	curved	steeply downward
sp6–ep2	curved / slightly straight	slightly upward

4.5.2 Analyses of Verbal Descriptions & Dialogues

An Utterance. The term "utterance" refers to speech parts produced coherently and individually, in other words, syntactically, semantically or prosodically complete speech units, by each participant. All post-exploration verbal descriptions gathered in all experiments and also the conversations observed in the Experiment IV, where I investigate the dynamics of a joint activity were first split into utterances.

A Phrase or Segment. The data collected through Experiment III was used to investigate the event segmentation and event description; therefore the verbal descriptions (the utterances) are split into more refined units, namely phrases. A phrase (or a segment) was defined as any unit containing a predicate (i.e. a verb) that expresses a single sub-event or state by following Berman and Slobin (1994). Based on this definition, all verbal descriptions collected in the Experiment-III were spitted into phrases (or sub-events) and annotated w.r.t annotation scheme, which is introduced in the upcoming part.

Dialogue acts. The utterances produced during joint action as a conversation acts were categorized intro three; (i) Request-Response Pairs, (ii) Alerts initiated by *A* (but do not require response from *E*) and (iii) think-aloud sentences.

Communicative goals. The utterances produced by both explorers and assistants were also classified w.r.t *their communicative goal* (1) instructive, (i.e. navigational, such as 'go downward from there'), or (2) descriptive utterances. Moreover, assistants' descriptive responses were also classified as follows; (2a) confirmative assistance (confirming the information which haptic explorer has already access), and (2b) additional assistance (introducing new property or updating the value of already stated property)

The referring expressions produced by either haptic explorers or verbal assistants give insight about how graph readers comprehend graphs, which elements are mentioned most, and how they are referred to. The investigation of multimodal interactions (namely interaction by means of language, gesture and graph) requires systematic qualitative analysis, as well as quantitative analysis. I followed a widely accepted method[12] developed by Dale and Reiter (1995), which addresses the generation of referring expressions, to characterize the semantic properties of graphical segments and the referring expressions produced during post-exploration verbal descriptions or during collaborative activity. According to Dale (1992), a system should start to generate referring expressions by choosing less computationally complex properties and then it should continue with more complex properties when still necessary; these properties are type properties, absolute properties, relative properties and relations respectively. By following this method, ⟨attribute, value⟩ pair representations to characterize the qualitative representations of graph shapes and landmarks were constructed (Alaçam, Habel and Acartürk, 2014; Habel, Alaçam, and Acartürk, under revision). To illustrate this, the attribute set, which is available for the "ep1-sp1" shape segment (see Figure 4-9 and Table 4-4) possesses the following properties: ⟨type, curved⟩, ⟨manner, steep⟩, and ⟨direction, up⟩. For the

[12] Reminder: Within the scope of this dissertation, it is not aimed going into implementation level in detail, instead I used the method as a tool to make systematic mapping between semantic properties of graphical features and participants' referring expressions, which is introduced in Chapter 2.2

systematic data analyses, the verbal data produced both in the post-exploration descriptions and in the dialogues of a joint activity were also annotated by using this method since it successfully foregrounds the common properties of multimodal data, see Table 4-5 for semantic attribute scheme for the verbal data.

Table 4-5 Semantic attribute scheme

Event-Denoting Expressions
Type Properties: Terms – ⟨term, peak⟩, ⟨term, something⟩ Location – Frame of Reference Terms (“start point”) – Haptic Ostensive Expressions
Absolute Properties: – ⟨value, 0⟩ for “it is 0” – ⟨count, 3 peaks⟩
Relative Properties: – ⟨size, small⟩ , ⟨manner, slowly⟩ – ⟨direction, up⟩
Relations: – ⟨temporal relations[13], after the fall⟩ – ⟨spatial relations, higher⟩
Others: – Interjections (hmm, ah…) – Affirmations/Negations

The experiment-III addresses the issues regarding event segmentation and description in graph comprehension. Therefore more refined annotation scheme, particularly concerning temporal information temporal information referred by the participants, is needed. In addition to those event-denoting attributes, the temporal information referred by the participants were also annotated with respect to scheme presented in Table 4-6.

Table 4-6 The annotation scheme for time-denoting expressions

Time-Denoting Expressions		
Reference to seasons		**Reference to general trend**
• Summer		**Reference to Spatial Location**
• Spring		**Reference End points**
• Winter		• Start point
• Fall		• End Point
Reference to months (explicit)		
• Jan.		**Temporal Prepositions**
• Feb.		• around
• Mar.		• for some time
• Apr.		• then
• May		• between, during, within
• Jun.		• towards, till
• Jul.		• from to
• Aug.		• …
• Sep.		
• Oct.		
• Nov.		
• Dec.		

In addition to the attributes stated by Dale and Reiter (1995), the haptic ostensive expressions (HOEs) were also identified. The haptic explorers produced HOEs that referred to the pointed locations, which are also accompanied by assistance request from the verbal assistant. Foster and colleagues (2008) define the HOE as a reference, which involves deictic reference to the referred object by manipulating it haptically. Since haptic

[13] Only this item in this annotation scheme refers to temporal information

explorer location is visible to verbal assistant during joint activity, haptic actions are useful to provide joint attention between *E* and *A*. The questions presented below exemplify the use of HOEs in a dialogue.

- Is this a start point?
- Which one is higher, this one or this one?

4.5.3 Gesture Annotation

The gesture coding scheme was based on both McNeill's (2005) semantic (*function-oriented*) and syntactic (*form-oriented*) features. Although the categorization proposed by McNeill (2008) is well defined and widely accepted, coding and identifying gestures for experimental purposes is always considered as problematic, since the production of gesture are highly context dependent, and gestures may have different forms and functions and exhibit variety from person to person. The annotation methods for gesture analysis can be classified under two categories; function-oriented and form-oriented annotation. In the function-oriented annotation method, gestures are classified w.r.t their semantic attributions, i.e. classification of gestures as beat, representational or interactive is an example for function-oriented method. Additionally, gestures can be also classified in another semantic level such as with respect to being supplementary, redundant and complementary. These are most common function-oriented categories. On the other hand in form-oriented annotations, labeling is performed with respect to syntactic properties, such as hand shape or direction.

As well as the classification, segmenting gestures is another issue that needs to be handled very thoroughly. According to McNeill, Levy and Pedelty (1990), a gesture phrase consists of three phases; preparation, stroke and retraction. On the other hand, it is also acceptable to perform gesture analysis in more surface level. Kita, Van Gijn and Van der Hulst (1998) name this broader gesture phase as a maximum unit is defined as "the movement that starts at the moment of the hand's departure from its resting position and end at the moment it return to resting position".

Gestures were annotated on the basis of frame-by-frame analysis by using the video annotation software ANVIL. The annotated features are summarized in Table 4-7. In the first classification, the gestures were categorized according to their semantic classifications, such as beat gestures and representational gestures. Then each representational gesture was classified as *static* or *dynamic*. According to this classification, the hand movements conducted in small space without having any directed trajectory were categorized as *static*, whereas the hand movements with aimed trajectory on the air were classified as *dynamic* gestures. Then they were categorized in terms of its directionality: *non-directional* and *directional* (vertical/diagonal/horizontal). Directional gestures were also classified into two categories; (i) *single direction*, and (ii) *multiple directions*. The gestures that contained movement in only one direction (such as upward) were classified under the "*single direction*" category, while category of "*multiple directions*" covers the gestures formed with the combinations of one-directional gestures. They were also identified in terms of hand shape (ASL), direction of the movement, start and end location of the movement on gesture space, the movement coverage (finger, hand, arm) etc.

While describing the graph, participants may produce more than one verbal phrase and also more than one speech accompanying gestures. Therefore, the gestures were also segmented into a sequence of discrete units (Kita et al., 1997). An abrupt change of direction in the trajectory of the hand movement that co-occurs with a temporal pause or a change in the speed of the movement was considered as a point for gesture segmentation. The interrater reliability analysis between the gesture coders are presented in the subsection of the corresponding experiments.

Table 4-7 The annotation scheme for gesture coding

Gesture Annotation Scheme	
Link to verbal data • accompanied speech part • its <attribute, pair> representation	Start Location • Bottom • Center • Top
Mapping between gesture and speech part (Phrase), see Chapter 6.4.6.2(B) • One-to-one • One-to-many • Many-to-one	End Location (bottom, center, top) • Bottom • Center • Top
Static or Dynamic	ASL Hand Shape
Directionality • non-directional • 1-directional • multi-directional	Movement coverage • Arm • Hand • Finger
Directions • Vertical • Horizontal • Diagonal • Their combination for multi-directional gestures	

4.5.4 Sketch Analyses

Sketches represent what retains in participants' memory (Tversky and Lee, 1999). Furthermore, as stated by Tversky (1999), "drawings reveal people's conceptions of things, not their perceptions of things". In particular, sketches provide complementary data for the analysis, because they can reveal details that the graph reader skips in verbal description for various reasons (e.g., the concept may be hard to express verbally or it may be considered as redundant by the reader). Furthermore, Kalia and Sinha (2011, p.14) also suggest that drawings are less noisy than the observers' mental image of the shape, which also allows for easier recognition. Their argumentation for this is that in order to draw a contour, which was recently explored, motor images from motoric representation are constructed. And, they might be more accurate than visual images, since same underlying mechanism is involved both for exploring and drawing. For those reasons, sketch analysis of explored graphs is considered to be an appropriate methodology to evaluate the conceptualization of the events represented by the graph

All sketches were assessed by two raters w.r.t their similarity to stimulus-graphs by employing five point Likert Scale (1: least similar, 5: most similar). For the analyses of the sketches, inter-rater reliability between two raters was tested using two-way mixed consistency average-measures ICC (Intra-class correlation). The detailed test results are presented in the subsection of the corresponding experiments.

4.5.5 Haptic Exploration Movements

The raw haptic exploration data set consisted of the recorded locations of the stylus of the haptic device on the 2D horizontal plane (horizontal and vertical coordinates). As the first step, the raw data points (the coordinates of the stylus) were segmented and labeled with respect to the landmarks for each graph (w.r.t. the qualitative ascriptions, see Figure 4-9 and Table 4-4). After the labeling of each data point, the Euclidean distance between each consecutive data point was calculated. This data was used to calculate several parameters such as time spent for each landmark and each segments, or the average speed. Besides, the use of haptic exploration movements as an input to decide on whether the user needs assistance requires more complex feature extraction and data cleaning, a part of Chapter 9.3.5 is dedicated to this issue.

4.5.6 Eye Movement Analysis

An appropriate methodology for investigating the construction of referential links between language and graphs is the eye-tracking methodology. As previously mentioned, providing haptic data labels is challenging issue, however, their role in graph comprehension is crucial, since they convey temporal and value related information that needs to be taken into consideration for fine-grain data extraction. The data labels are inseparable parts of the visual graphical representations, therefore the studies that investigate visual graph comprehension usually incorporates this component. Therefore, there is lack of research in understanding what happens in the lack of data labels especially regarding graph (event segmentation). Within the scope of this dissertation, I focus on haptic graph comprehension and employed the visual modality as a control group, thus I evaluate the effect of the data labels with the methods that can be applied to the analysis of the data collected from the haptic modality condition. For this reason, the analysis of the eye movements during visual graph reading (with and without data labels) were left for the future studies, however I would like to briefly explain why the incorporation of this methodology is important and included in the experimental settings of this research.

Eye tracking has been one of the techniques that provide comprehensive information about online cognitive processes of a graph reader since it lets to trace the allocation of attention. In addition to the fact that eye movements are central to the visual system extremely fast, and metabolically cheap, they have a lower threshold for being triggered as compared to other motor movements. This makes eye tracking a very powerful and accurate tool to investigate cognition (Richardson et al., 2007).

Eye movement features during graph comprehension such as fixation duration, gaze time, number of fixations (instances when the eye remains relatively still within a particular location), the occurrence of regressions (transitions between the areas of interests), and a number of variations on these measures can be used to investigate moment-by-moment cognitive processing of a graph by the reader in order to assess comprehension strategies and the effectiveness of the graph (Just & Carpenter, 1980; Rayner et al., 1989; Renshaw, 2004).

SECTION - II

Haptic Graph Comprehension

CHAPTER 5

Exploration and Comprehension Of Haptic Graphs

Graphs are one of the efficient ways of visual communication to convey the highlights of data, however visual perception differs from haptic perception; therefore a piece of information that can be clearly perceived as highlighted in the visual modality can appear to be hidden when it is converted to the haptic modality. The saliency of the properties of an object (of a graph line in our case) may differ for visual and haptic exploration, even for the nearly identical visual and haptic representations (Klatzky, Lederman and Reed, 1987). Both the visual and haptic modalities exhibit different characteristics about perceptual information intake, thus leading to potential differences in how the situations represented by the graphical entities are conceptualized. In particular, compared to visual exploration, haptic exploration is subject to limited perceivability: The visual perceptual system is capable of accurately computing many attributes simultaneously with considerable speed, while haptic object perception works in a more sequential way as it involves the performance of (manual) exploratory actions on an object. Hence, in the course of the development of fully automatic verbal assistance for haptic graphs, investigation of differences in two sensory modalities (visual and haptic) is necessary to detect and close the gap caused by these differences. Additionally, the quality of the verbal assistance as an enabler of efficient interaction highly depends on whether appropriate referential and co-referential links are established. For designing haptic graphs augmented by audio assistance, be it sonification or speech, it is necessary to determine, which information depicted by the graph or by segments thereof, are important. To find appropriate content for the verbal assistance, I focus on studying representations at a conceptual level in what follows by tapping into the data obtained by means of various experimental paradigms, including the analysis of verbal protocols and referring expressions, of gestures, and of sketches and haptic exploration patterns.

5.1 Exploring Haptic Graphs

As previously mentioned in Section-I, a Phantom Omni force-feedback device (recently renamed as Geomagic® Touch™) was used by the graph readers to actively explore the haptic graphs represented as an engraved line on a virtual plane with a stylus. I will not go into detail concerning the underlying haptic sensory mechanism; nevertheless I should note that haptic exploration for line graphs in this study requires remote haptic exploration. According to J.J. Gibson (1962), active haptic exploration guides the explorer's attention to the external properties of the explored object. An experiment conducted by Klatzky et al. (1985) on the identification of familiar objects using the tactile sensory modality demonstrated that active haptic object recognition can be rapid and accurate since the haptic perceptual system can encode many different object properties. Existing research on haptic object recognition is mostly based on the classification of exploratory actions proposed by Lederman and Klatzky (1987, 2009). The properties for haptically perceived objects are commonly divided into two main categories, namely *substance-related* and *structure-related* properties (Lederman and Klatzky, 1987). In their later publication, they switched the terminology to *material* and *geometric* properties respectively. Here, I will continue with the later one (Lederman and Klatzky, 2009). Their classification consists of a set of six so-called *exploratory procedures*, each of which is associated with certain object properties: In brief, these procedures are (1) lateral motion (associated with texture), (2) unsupported holding (for weight), (3) pressure (for hardness), (4) static contact (temperature), (5) enclosure (global but coarse-grained shape and volume information) and finally (6) contour following (global and exact shape). The first five exploratory actions require direct touch or involve the exploration of some features (such as weight) which haptic graphs do not possess. As for haptic graph comprehension, only the contour following procedure associated with both coarse and fine-grained (geometric) shape properties is applicable.

According to Lederman and Klatzky (2009), haptic exploratory procedures can have an influence on object processing *bottom-up and top-down*: Exploratory movements may effect in a bottom up manner when they induce explorers to encode particular object attributes like the above-mentioned associations. In contrast, top-down effects may take a part when instructions direct explorers to learn about a particular object property. In our case, one of the top-down factors is the graph schemata (introduced in Chapter-1) that guide explorer about which feature is important for the comprehension of haptic graphs that depict abstract event in a spatio-temporal frame. Besides, the choice of exploratory actions is also influenced by the task that the graph explorer needs to perform; for example, the task (or the intention) of looking for the relation between two points in the graph may be performed differently than the task of acquiring the general trend of the graph. Although both tasks have to be completed by the same exploratory action, namely contour following, the specific action patterns performed in the process could be different is aspects like speed, or the number of back and forth actions.

5.1.1 Amodal Geometric Properties

A series of experiments conducted by Klatzky et al. (1987) demonstrated that the haptic system varies from the visual system in regards to encoding. Moreover, an attribute, which is salient in one modality, may be indistinct in another modality. Hence, whether an attribute will be used for encoding is dependent on its saliency in a particular modality. The mental representation of an object perceived through different modalities may differ substantially as a result.

In the comprehension of line shape, a crucial role is played by amodal geometric properties of global shape or of local shapes (such as curvature, orientation). As previously stated, material properties (such as texture, temperature, hardness etc.) are irrelevant for graph comprehension and will therefore be excluded in my further considerations. The amodal geometric properties extractable through contour following are the following:

Curvature. Previous work has shown that the perception of curvatures depends on whether they are convex or concave (van der Horst and Kappers, 2008). In this study, the participants were asked to compare the curvatures of convex and concave shapes by touch. Their findings indicated that people underestimate the curvature of convex shapes compared to that of concave shapes. Additionally, according to the results of a study by Kappers (2011), first-order stimulus information[14] (i.e. the difference in attitude or slope) is the dominant factor determining the threshold for curvature detection or discrimination.

[14]Within the scope of this dissertation, the slope is categorized as a second-order property. Kappers (2011) categorizes it as a first-order property; however she starts to name categories from zeroth-order (such as height, which is referred as first-order in this dissertation). Therefore although the terms are different, they correspond to same level concerning the underlying hierarchy.

Shape. Shape plays a primary role in visual object recognition: The visual sensory system is very good at processing information simultaneously and performing fine spatial analyses of patterns over a broad range of distances (Biederman, 1987). For the haptic modality, however, shape is particularly difficult (Klatzky et al, 1987). In order to acquire fine-grained shape information haptically, it is necessary to use contour following. But this exploratory procedure is relatively slow and complex especially if it is performed with two hands. The Phantom Omni device simplifies this procedure by reducing the information acquisition channel into a single point with the stylus. Another reason for why shape is considered as problematical is that the comprehension of this feature requires the involvement of a memory component due to the sequential nature of haptic perception (Lederman and Klatzky, 2009).

Size. Similar to how vision is superior to haptics when it comes to shape, the size of an object, or of a line segment in a graph in this case, also appears to be more readily encoded by vision (Klatzky, Lederman and Reed, 1987). The extraction of size information for line segments also involves a contour following process. Because size is a relative measure used for discrimination and comparison, its acquisition of also necessitates the involvement of memory.

Orientation and Spatial Anisotropy. Orientation is another critical property as far as haptic graph perception is concerned. Of particular concern for the comprehension of this property is spatial anisotropy. It is crucial with respect to two issues regarding haptic graph design. First, the stylus of the Phantom Omni has a limit in left to right rotation. Its body part (see Phantom Omni User Manual) is responsible for performing right and left motions. However due to cyclical movement of the body, the stylus may follow an arc-shaped trajectory for the very right and very left sides. In order to eliminate this effect in the empirical investigation, the haptic graph workspace was limited to a region where linear exploration can be performed.

Figure 5-1 Physical Space of Phantom Omni Device (Body: left/right, Arm: up/down, elbow: in/out) (retrieved from the Official User Manual)

Another issue is that even for a raised linear line explored with the hands or the fingertips, the perception may differ from the center of the haptic space to its edges. Isotropy means displaying of a uniform distance metric regardless of the direction and magnitude of the distance across space (Burtt, 1917; Lederman and Klatzky, 2009). Numerous studies have reported that haptic space is anisotropic in several respects: First of all, according to *radial/tangential illusion*, "linear extents felt along a radius toward and away from the body are perceived as longer than the same extents felt along a tangent to that radius" (Sanders and Kappers, 2008; Marchetti and Lederman, 1983). Furthermore, another one is called as the *horizontal–vertical illusion*. For a T or L-shaped stimuli on the horizontal plane, vertical lines are perceived as longer compared to horizontal lines that have same length (e.g., Burtt, 1917). Moreover, the third one is called the *oblique effect*. Observed also for visual perception, the haptic perception of oblique lines has been suggested to be problematic relative to that of vertical and horizontal lines (Lederman and Klatzky, 2009; Lechelt, Eliuk, and Tanne, 1976; Lechelt and Verenka, 1980). According to Gentaz and Hatwell (1995), the haptic oblique effect occurs because of the kinesthetic gravitational cues produced during manual exploration by the hand–shoulder system, and also because of the memory constraints induced by sequential exploration. Additionally their study also suggested that this illusion occurs at a relatively late stage of the orientation processing, the stage in which the perceptual information is converted into a more abstract representation of orientation, mostly resulting from the use of a reference frame consisting of vertical and horizontal axes. According to their interpretation, in order to encode vertical or horizontal line, taking reference to one of those axes would be sufficient. On the other hand, oblique line encoding requires taking the distance to both axes into account, therefore their encoding is cognitively more demanding (Gentaz, Baud-Bovy and Luyat, 2008; for a more elaborated review of spatial anisotropy, see Kappers and Schakel, 2011; Lederman and Klatzky, 2009; Cuijpers, Kappers and Koenderink, 2003; Faineteau, Gentaz and Viviani, 2003). In the statistical line graphs, each label on the x-axis can only have one data point on the graph line (i.e. a line graph that represents monthly bird population distribution, each month has only one population value), thus a graph cannot have a vertical line. However, this perceptual anisotropy in favor to vertical lines may lead to wrong interpretations. Aside from that, oblique lines are frequent parts of line graphs; hence, this should be also taken into consideration in the design of haptic graphs. Additionally, shape features such as the path length, or the turning angle (Vogels, Kappers and Koenderink, 1999, Sanders and Kappers, 2008), and the direction of the movement (Davidson, 1972; Hunter, 1954) are also known to affect the haptic perception of linear shape.

5.1.2 Remembering and naming haptic shapes

The matters of haptic shape recognition and naming are particularly important both for online haptic graph comprehension and also for recalling them for later use, since the perceptual graphical entities are linked to conceptual graph-domain information through these processes.

Similar to vision, two system take place in the functioning of haptic perceptual system; these are the so-called *"what pathway"* and *"where pathway"*. The "what pathway", also known as the ventral system, deals with the perceptual processes involved in object identification and recognition. The "where pathway", also known as the dorsal system", is based on spatio-temporal discontinuities, and deals with the perceptual guidance of action and provides a layout of objects on haptic space (Kalia and Sinha, 2012). However, as argued by Klatzky and Lederman (2007), haptic object recognition is different from visual object recognition for the reasons aforementioned therefore traditional models of visual object recognition (e.g., Biederman, 1987; Marr, 1982) are not suitable to explain the recognition of haptically perceived objects. Their claim is that spatially aligned edges are the fundamental features of those models, however haptic system has lower spatial acuity in terms of the extraction of these features (Weinstein, 1968).

The local and sequential character of haptic perception demands more perceptual and cognitive resources for the integration of local information to global shape (Loomis, Klatzky and Lederman, 1991). Furthermore, in order to make the relational reasoning over geometric properties through contour-following, memory processes have to be involved (Lederman and Klatzky, 2009). The recognition of tactile images is generally better when they are scanned actively rather than passively (with a stationary hand) (Gibson, 1962; Heller and Myers, 1983), since kinesthetic information obtained from the hand movements may aid the process of integrating the collected information about the shape (Magee and Kennedy, 1980). Besides, both sighted and blind observers perform well on tactile match-to-sample tasks with geometric shapes (Heller, 1989), suggesting that they can accurately obtain simplistic shape information by touch. Yet, it is still unknown just how well observers perceive the shape of more complex tactile images of objects, such as haptic statistical graphs in which small perceptual changes (for haptics) may correspond to crucial conceptual information.

The results of a study conducted by Kalia and Sinha (2012) showed that recognition of simplistic tactile pictures is largely inhibited by low-level shape processing rather than high-level object recognition mechanisms, since the latter one, recognizing or memorizing a shape, inherently involves the necessity of processing and determining the shape of the stimulus. They claim that shape acquisition is a bottom-up process requiring the integration of tactile information over space and time. As observers explore the object (in their study, through hands and fingertips), they must comprehend and remember local pieces of information. Additionally, Lederman and Klatzky's study (2009) also indicated that the duration of manual exploration has an effect on the preference for the processing granularity (i.e. focusing on local or global properties). For the exploration of two similar geometric objects with respect to their global shape, it appears that the attention is drawn to local features first. When provided with more exploration time, the global features are attended. Kalia and Sinha (2012) summarize the possible problematic issues regarding memory for shapes. First of all, this construction process (from local to global) is assumed to be challenging. With ongoing exploration, the explorer has to integrate local pieces of shape to each other consecutively to form a global shape, hence local features must be stored in memory and recalled during integration; therefore it is more cognitively demanding. They suggested that most of the errors occur during this integration process. Noisy motor control and poor spatial localization of the hand might be other reasons for the errors. They concluded that the facilitation of the integration process may help reduce the possible errors and, thus, is one of the important aspects to improve haptic object recognition. In addition to the problems regarding shape acquisition, associating it with a particular object, which is recognizing or naming it, would be also difficult. A study conducted by Heller, Calcaterra, Burson and Tyler (1996) suggested the reasons for poor recognition are largely due to the problems regarding semantic labeling of the patterns rather than poor shape acquisition. Their results showed that when categorical information about tactile images is provided to participants, they perform dramatically better. This finding is particularly crucial, since successful haptic graph comprehension also requires facilitation of conceptual information about abstract events, which are depicted in graph, rather than just the acquisition of shape properties. Instead of associating shape with a particular object, a graph reader needs to associate the shape of the graphs with abstract conceptual representations of events that are represented in them (either local graph shapes or global shape). This involves higher-level mechanisms for retrieving object representations that match the perceived shape of the image.

More insights about the encoding of shape properties and its effect on memory was provided by Lacey and Cambell (2006). In this study they presented an empirical research that focuses on the investigation of mental representation of familiar and unfamiliar objects in visual/haptic cross-modal memory. Their experiment that foregrounds the familiarity of objects was based on the dual-code theory (proposed by Johnson, Paivio, and Clark, 1989). Briefly, for familiar objects, visual representations can be formed easily. Moreover, they can also be represented verbally since the naming of a familiar object is also easy. It has been suggested that visual and haptic perceptual input may evoke a visual representation of the object that can trigger the name of the object (Johnson et al., 1989). However, the role of verbal processing appears to be more complicated for unfamiliar objects. First of all, naming is a challenging task for unfamiliar objects, but verbal descriptions can be produced any time during exploration. Therefore instead of having verbal processing at the later stage as naming, covert verbal descriptions that occur during exploration contribute to construction of mental representation of an object. The Lacey and Cambell's findings (2006) lead to the conclusion that, regardless of the sensory modality, verbal descriptions may aid both the encoding and the recognition of unfamiliar objects. For familiar objects, naming is relatively easy process, therefore it does not have much influence on encoding but it is considered as crucial for recognition.

These studies are relevant for the design of verbal assistance system with the purpose of easing haptic graph comprehension. First, the graphs are generated through data points.

As there is no pre-determined shape for the line that connects these data points, it may not be easy to name. Still, referring to the shape of the graph is quite common. Some graphs may contain familiar shapes that can be easily referred to and named instantaneously (as it is for example with “staircase”), and some of them can be hard to name and, consequently, hard to encode and to remember. If the global and local shape is familiar, providing its name would be enough for successful communication leaving easy to encode information to the user’s responsibility and focusing on hard to encode information for the sake of efficiency. If it is unfamiliar, more detailed verbal descriptions should be provided.

Another top-down factor which is highly relevant for haptic shape comprehension is the facilitation of proper schemata. The studies (Hunter, 1954; Davidson and Whitson, 1974) that compares the performance of blind observers and sighted observers on the same task showed that blind participants performed better and their judgments were more accurate when compared to the sighted subjects. In a follow-up study, the sighted participants were instructed about the exploratory strategies spontaneously employed by the blind participants. The result of this was that the accuracy of their judgments improved by a lot. This finding is not trivial from graph comprehension perspective, since it puts emphasis on the fact that the gap is not just due to perceptual differences but also due to triggering more appropriate graph schemata regarding the content and the task at hand. In that experiment, the schemata used by the blind participants were more appropriate for the task than the ones used by the sighted participants. Also, when the schemata of the sighted participants were updated with useful strategies, they performed better. In the haptic graph comprehension domain, the situation is reversed, since the graphs are widely used elements of multimodal communication that sighted people come across very frequently. Besides, these representations are based on the visual Gestalt principles making the facilitation of appropriate graph schemata easier (Shah, Mayer and Hegarty, 1999; Pinker, 1990; and for an overview of the visual Gestalt principles, see Wagemans et al., 2012).

5.2 Shape Geometry for Graph Lines

As introduced earlier, spatial representations are essential for recognizing objects and events from different sensory channels. Beyond perception and motor action, higher cognitive capacities such as memory, problem solving, and planning are also based on spatial representations. Furthermore, communication which can be seen as the production and comprehension of external representations (Habel, 2003) makes systematic use of spatial representations, for example in the communicational modalities of language, gestures, graphs, and maps. In language, spatial expressions are used not only for communicating about space, but also in other domains, such as time, (in descriptions like ‘before Christmas’ or ‘short time’) and numbers/quantities (in expressions ‘large numbers’ or ‘high value’). The pivotal role of spatial representations for connecting communication, cognition, and perception is based in their modal non-specific nature, described by Jackendoff as ‘amodal or multimodal’ (2012), and characterized by Loomis and Klatzky as amodal (2007). Here, geometric specifications of amodal abstract shape primitives were used to describe lines, in particular graph lines (on abstract spatial structures, see Habel and Eschenbach (1997). The geometric approach for line shape representations (see Habel, Alaçam and Acartürk, under revision; Alaçam, Acartürk and Habel, 2014) used in this section does not require concepts of differential geometry like differentiability, continuity or real numbers.

Looking at line and contour shape from the perspective of perception leads to different requirements and to, further, geometrical concepts. For example, the *curvature landmarks* investigated in Cohen and Singh’s seminal paper on shape segmentation are geometrically specified using the methods of differential geometry (Cohen and Singh, 2007). The shape-landmark points can be seen as qualitative counterparts to Cohen and Singh’s curvature landmarks.

Most qualitative approaches to shape representation focus on the shape of contours (see, e.g., Hoffman and Richards, 1984; Richards and Hoffman, 1985; Eschenbach, Habel, Kulik, and Leßmöllmann, 1998; Meathrel and Galton, 2001). However, lines and in particular graph lines have specific characteristics to be considered with respect to shape representation and cognition. First and foremost, the prominent curvature landmarks of positive maxima and negative minima of curvature (Cohen and Singh, 2007) depend on the concepts of convexity and concavity and, thus, on contours in closed curves. Additionally, graph lines (of statistical graphs) are based on functions. Therefore, a graph line is neither closed, nor can it cross with itself, nor it can branch.

Furthermore, graph lines are conventionally oriented in the prevalent reading and writing direction and they are comprehended in the context of an orthogonal system with two axes. This holds also if the graph axes are not explicitly given, for example in sketch drawings of graphs. The haptic graphs that were used in the experiments were realized in a rectangular frame that implied such an orthogonal system of axes. The use of the geometric shape concepts for describing graph lines is exemplified with one graph that was actually used in the experimental studies in Figure 5-2. Table 5-1 gives a tabular summary of qualitative representations for selected shape landmarks and segments.

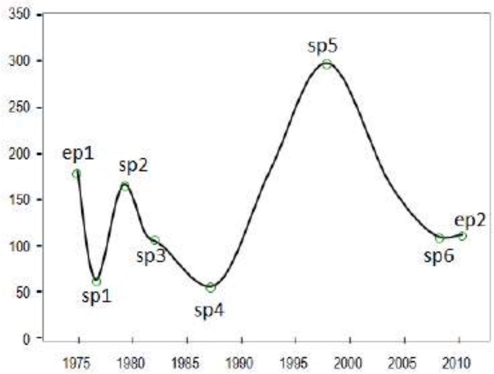


Figure 5-2 Qualitative shape landmark ascription for one graph in the stimuli-set

Table 5-1 Qualitatively described shape landmarks and line segments

Shape landmarks		
Landmarks	**Landmark characteristics**	**Global properties**
ep1	left end pt., local max.	Higher than sp1, sp2, sp3, sp4, sp5, ep2
sp1	smooth pt., local min.	Higher than sp4
sp2	smooth pt., local max.	Higher than sp1, sp3, sp4, sp5, ep2
sp3	smooth pt.	Higher than sp1, sp4, same height as sp6
sp4	smooth pt., local min.	Global min.
sp5	smooth pt., local max.	Global max.
sp6	smooth pt., local min.	Higher than sp1, sp4, same height as sp3
ep2	right pt., local max.	Higher than sp1, sp3, sp4, sp5
Shape segments		
Segments	**Shape characteristics**	**Vertical orientation**
ep1–sp1	curved	steeply downward
sp1–sp2	curved	steeply upward
sp2–sp3	curved	steeply downward
sp3–sp4	curved	diagonally downward
sp4–sp5	curved	steeply upward
sp5–sp6	curved	steeply downward
sp6–ep2	curved / slightly straight	slightly upward

These qualitative ascriptions can be automatically derived from quantitative measures with a few calculations. First of all, each data point contains two labels (time, value); one for the x- and one for the y-axis. Based on the data values illustrated in Table 5-2(a), each shape landmark is marked with its landmark type; *global points* (min. and max.), *local* *points* (min. and max.) and *other smooth points* (i.e. *sp3* in Figure 5-2). Then each landmark is also classified with respect to three types of relations; *higher*, *lower* and *same as*. Data values are also sufficient to make ascriptions for shape segments. The direction can be easily calculated by looking at the sign of the difference between values belonging to two consecutive x-axis units (i.e. *sp1*(65) - *sp2*(160) = -95 indicates a downward trend). The information of maximum and minimum y-axis units (see Table 5-2(b)) is used to infer the steepness of the line segment. For this sample, they are 350 and 50 respectively, indicating that the y-axis is consisting of 7-intervals. For each segment, the number of intervals that the line segment covers in direction of the x-and y-axes is calculated, and then the rest can be considered as a discretization process among steepness categories. Additional ascriptions can be also obtained about the general trend, i.e. looking at the difference between start and end points of the graph. Besides, the saliency of the global maximum can be inferred from its difference to the closest (in value) local maximum. The same calculation can also be done to infer the saliency of the minimum point. For this sample, the difference between the global maximum and the closest local maximum is 125, while the difference regarding the global minimum is 10 (given that the 1-interval difference on the y-axis is 50). From these ascriptions, it can be easily concluded that for this sample, the global maximum is salient; in contrast, the global minimum is indistinct. Moreover, the distance of the two points being compared is also important, since they are explored sequentially. Due to short-term memory constraints, comparing entities that are too far away from one another on the haptic graph is a challenging issue. In this case, one can infer that the local minimum and the global minima that have close y-values are also in 2-landmark distance to each other (a local or global max. has to exist between two minimum points by definition), so this comparison might be easier than comparing two far points (i.e. 4-landmark or 6-landmark distant).

Table 5-2 Quantitative values used for the derivation of qualitative ascriptions

(a)

	Y-Value	X-Value
ep1	175	1975
sp1	65	1977
sp2	160	1980
sp3	110	1982
sp4	55	1988
sp5	300	1998
sp6	110	2008
ep2	120	2010

(b)

	Value	Shape Landmark
Max. Y-label Unit	350	NA
Min Y-Label Unit	50	NA
ep1-ep2	-55	NA
Gmin*	55	sp4
Cmin*:	65	sp1
Gmax.	300	sp5
Cmax*.	175	ep1
Cmin-Gmin	10	NA
Gmax. - Cmax.	125	NA

*Gmin: Global minima, Cmin: Closest value to Gmin, NA: Not applicable

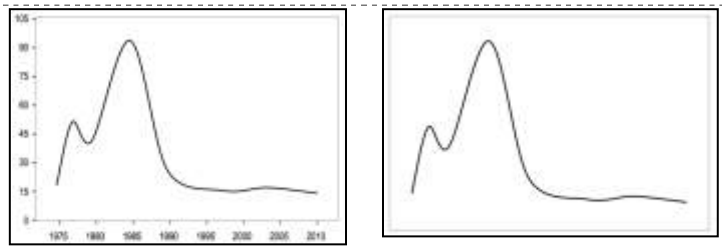
5.3 Empirical Study- I: Shape Concepts in Graph-line Comprehension

For designing haptic graphs augmented by audio assistance it is necessary to determine, which information depicted by the graph or by segments thereof, are appreciated as important (Habel and Acartürk, 2012). The analysis of verbal descriptions, speech-accompanying gestures, sketches and haptic exploration patterns could be helpful in assessing the graph reader's comprehension and in obtaining the important aspects considered as worth mentioning in verbal descriptions. The analyses in multiple experimental modalities provide complementary data for identifying appropriate line segments and other graph parts for verbal assistance, for designing a lexicon of verbal assistance, as well as for evaluating the formal approach presented in the previous section as an analytical tool for generalizing the findings. The empirical work that was done on haptic graph comprehension employing single-user protocol and without verbal assistance is discussed in the following. The results presented here were partially reported in the several publications (Alaçam, Habel and Acartürk, 2013; Habel, Alaçam, Acartürk, 2013; Acartürk, Alaçam and Habel, 2014).

5.3.1 Experimental Setup

Participants. 31 university students (*Mean age*=25.42, *SD*=5.05, 15 female) from the University of Hamburg participated to the study. The experiment was conducted the native language of all participants, which is German. The haptic explorers were blindfolded sighted people. All of them used the Phantom Omni device with their dominant hand. An informed consent form was signed by each participant before the experiment started.

Procedure and Materials: The experiment was conducted in three conditions in a between-subject design by employing Experimental Paradigm-I, which was presented in detail in Chapter-4. In the first condition, "*haptic graphs without data labels*", (9 participants, 4 female, *Mean age* = 25, SD = 6.3), the participants explored the line graphs haptically. No haptic labels (such as information for numerical labels and axis title) were provided in the haptic modality, because it is hard to implement and they would be rather distracting. In the second condition, "*visual graphs with data labels*", (11 participants, 3 female, Mean age = 23,4, SD = 1.7), graphs with data labels were presented on a computer screen, thus the participants had visual access to the graphs (see Figure 5-3(a) for a sample graph). In the third condition, "*visual graphs without data labels*", (11 participants, 8 female, Mean age=26.8, SD=5.5), which served as a control condition for the first two conditions, the participants inspected the visual graphs without data labels (see Figure 5-3(b)).

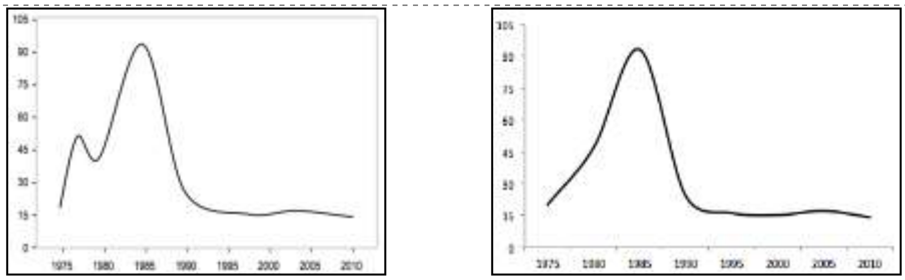


(a) (b)

Figure 5-3 Sample visual graph (a) with and (b) without data labels

Visual graphs without labels were used as the control group to understand whether the observed differences are modality-related or due to other factors such as the existence of data labels or amodal properties of the object (i.e. global shape, saliency of the segments, or conceptual properties). The experimental design of this study was intentionally devised to test the effect of the modality and the existence of data labels in a systematical way. The effect of data labels on event segmentation was investigated by comparing condition-2 and condition-3. The effect of the sensory modality on event segmentation was explored by comparing condition-1 and condition-3. Since original data graphs were employed in this experiment, a systematic control of amodal properties of graph shapes was left aside. This issue is comprehensively investigated in Chapter 6.

The graph-set used in this experiment was taken from the "Waterbird Census at Bolinas Lagoon, Marin County, CA" by the Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science and was redrawn based on the original. The reason for choosing data that represents a real population is to observe the modality-dependent differences and also the effect of data labels on original graphs, which are created by domain experts with a purpose/message to convey. The graph given in Figure 5-4(a) is the original graph that contains data labels on the x-axis for every 5 years. But the data samples for each population in this database were collected every year, and they were presented in 5-years intervals on the x-axis for the sake of visibility and readability Figure 5-4(b) illustrates how the graph would look like if the sampling rate was once in every five years, first the small details would not be present. Besides, the line between two data points would serve as a tool to highlight the trend without carrying information about the exact values of the intermediate years (if there are). Although this distinction is beyond the scope of this dissertation, it should be noted that the differences in the visualization and decisions made by a graph designer (about "real sampling rate", "presented sampling rate") and also the content carried by the x-axis data labels (i.e. yearly or monthly data on average can be considered both continuous and discrete in some sense) is important for graph visualization. Previous research (mostly on visual graph domain) already showed that they may have an effect on graph comprehension (e.g. Shah and Freedman, 2011).



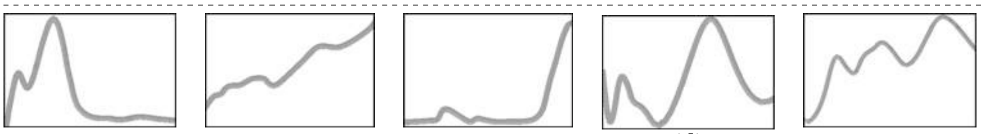
Sampling Rate: Once for every year (a)

Sampling Rate: Once for every five years (b)

Figure 5-4 Qualitatively described shape landmarks and shape segments

In the haptic graph condition, all participants completed a warm-up session that aimed to familiarize the participants with Phantom Omni® Haptic Device before starting the exploration of the actual haptic graphs.

In the experiment session, each participant was presented with five smooth-edged line graphs. In the instruction session, the participants were then informed that the graphs represented populations of bird species in a lagoon, see Figure 5-5 (two additional graphs were employed for the familiarization part). Each graph was shown for 10 seconds on a computer screen for the visual conditions, while haptic exploration was conducted without time limitation. Upon completing the exploration phase, the participants were asked to imagine themselves in an online meeting, in which their task is to present single-sentence summaries of the annotated graphs to the audience. According to the scenario, the audience is able to see the participant (i.e., the presenter) but not the graphs. Basically, the presenter first investigates the graph visually or haptically, then s/he turns towards the audience (an audience picture displayed on another computer screen), and presents a single-sentence summary of the graph.



(a) (b) (c) (d) (e)

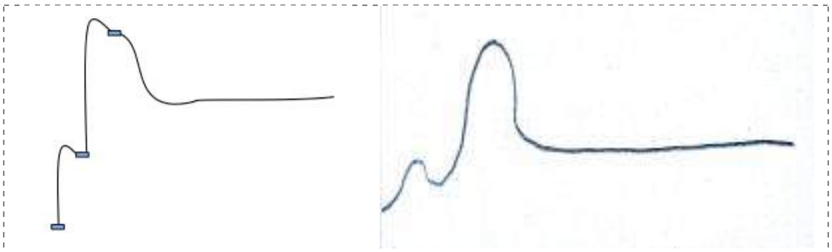
Figure 5-5 Five different haptic graphs

Gesturing was not mentioned in the instructions. Nevertheless, many post-exploration verbal descriptions produced by the participants were accompanied by spontaneous gestures. The participants' verbal descriptions, so the speech-accompanying gestures were video-recorded for five population-graphs of bird species. Additionally, the participants were also asked to draw a sketch of the graph using pen and paper.

The following is a translation of a German description given for the graph depicted in Figure 5-5 (a):

(1) "It is similar to the one before with the difference that it was a bit deeper. Then it had a small peak, than a large peak and somewhere in the midfield it levels off."

This description (1) was accompanied by a gesture sequence, which is depicted in Figure 5-6a. Figure 5-6b illustrates the post-exploration sketches produced by the same participant.



(a) (b)

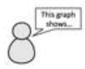
Figure 5-6 (a) Trajectories of the gestures accompanying the verbal description (the rectangles indicate pauses on the gestural movement) (b) post-exploration drawing

Table 5-3 presents the list of all dependent variables utilized in the analyses. The analysis results of the different types of data gathered during the experiments are reported in what follows.

Table 5-3 The list of dependent variables

Verbal descriptions
▪ Reference rate for shape segments and shape landmarks ▪ The use of modifiers
Gesture
▪ The number of representational gestures ▪ Static versus Dynamic Gestures ▪ Directionality in gestures — Non-Directional, 1-Directonal or Multi-Directional
Sketches
▪ Similarity scores for post-exploration
Haptic Exploration Movements
▪ Time Spent for landmarks and segments ▪ Average Speed for landmarks and segments

5.3.1.1 Verbal Descriptions



The shape of the graph line is the property that identifies the referents by distinguishing it from its distractors. Therefore verbal protocols provide valuable data for identifying the salient points of interest, which usually point to haptically difficult regions in graphs, identified in the course of exploration. In particular, the referring expressions that are produced during verbal descriptions[15] single out modality-specific similarities, as well as differences between graph lines of different shapes.

Verbal Coding. Graph shapes were evaluated under two main categories; (i) *shape landmarks* that emphasize a salient change in the pattern of the graph line and (ii) *line (shape) segments.* Since *end points* (both left and right ends) carry semantically different information as compared to other shape landmarks on the graph-line, *the category "landmarks"* was also split into two subtypes; (i) *end points* and (ii) *intermediate shape landmarks* (see Figure 5-2). Furthermore, verbal expressions that refer to *"landmarks"* on the graph were also classified into three groups according to the content of the referred information. If the verbal expression contained words such as peak, minimum etc., they were evaluated under the "*term*" (*type*) category. Expressions that referred to the quantity of the bird population (referring to y-axis labels) were assigned to the "*value*" category, as well as phrases like high value, and low value. The third category, "*time*" covered time related expressions (referring to x-axis labels).

[15] This analysis were presented in Habel, Alaçam and Acartürk (2013).

A. Results

A.1 The Reference Rate

A two-way ANOVA was conducted to investigate the differences in reference rates for different types of graph shapes (namely shape landmarks and line segments) across modality[16]. Average mention rates (*reference rates*) concerning shape landmarks and line segments per description were calculated and then compared. The results showed that the participants referred to line segments more than shape landmarks for all three conditions: visual graphs with labels (χ^2= 17.4, p< .05), visual graphs without labels (χ^2= 64.9, p< .05) and haptic graphs (χ^2= 22.0, p< .05).

Moreover, Pearson chi square tests revealed that the reference rate for shape landmarks in visual graphs without labels were significantly lower than that in visual graphs with labels (χ^2= 16.7, p< .05) and that for haptic graphs (χ^2= 9.9, p< .05), without a difference between the latter two conditions, χ^2= .7, p> .05 (Figure 5-7). This may indicate that the absence of data labels makes the detection of shape landmarks harder. On the other hand, haptic perception seems to recover some of the information mainly about these landmarks, which are easily accessible in visual modality due to presence of data labels.

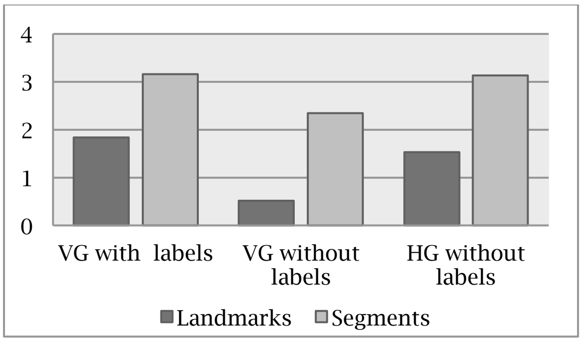


Figure 5-7 Reference rate for shape landmarks versus line segments per description

As mentioned in the coding section, the category "landmarks" has two sub-categories: end points and intermediate shape landmarks. A Pearson chi-square test revealed that while there was a significant difference on the average mention rate for end points and intermediate shape landmarks in the visual graphs with labels (χ^2= 19.3, p> .05), the differences in both the visual graphs without labels condition (χ^2= 0.0, p< .05), and the haptic condition (χ^2= .5, p> .05), were not significant, Table 5-4. Furthermore, intermediate landmarks were mentioned more frequently in the descriptions of the visual graphs with labels, compared to both the haptic condition (χ^2= 5.2, p< .05) and the visual graphs without labels (χ^2= 5.6, p< .05). However, the mention rate between the haptic condition and the visual graphs without labels were not significant (χ^2= .2, p> .05).

Table 5-4 Average number of referred shape landmarks

		Visual graphs with labels	Visual graphs without labels	Haptic graphs without labels
Intermediate	* [	6.40	1.50	3.60
End points		2.30	1.50	3.00

[16] In the haptic condition, behavioral data for only 8 of 9 participants was evaluated, because one of the participants had a misconception about the time domain, interpreting the x-axis in terms of months instead of years. For the visual condition without data labels, behavioral data for 10 of 11 participants was evaluated since she produce very long descriptions instead of short summaries.

The type of information mentioned in the referring expressions for those landmarks was also analyzed. It turned out that in all conditions, for the descriptions of end points, the participants preferred using the "value" category. Moreover, intermediate shape landmarks were mostly mentioned in the expressions under the "term" category for both haptic graphs and visual graphs without labels. On the other hand, as well as mentioning of the expressions in the "terms" category, the participants in the visual graph with labels condition also preferred using expressions regarding time and values represented by x- and y-axes respectively. As can be seen in Table 5-5, the absence of data labels in the visual modality resulted in a significant decrease in the production of expressions under the "term" category (χ^2= 9.1, p< .05). However, the results indicated that haptic modality, although no data labels are provided, recovers that information and enhance the conceptualization of terms, such as *peak*, *minimum* etc. compared to visual graphs without labels (χ^2= 8.4, p< .05). This analysis also suggested that expressions relating "time" and "value" were closely related to the presence data labels because they give direct access to that information.

Table 5-5 Average mention rate of intermediate shape landmarks

	Visual graphs with labels	**Visual graphs without labels**	**Haptic graphs without labels**
Term	3.20	1.20	3.13
Value	1.30	0.10	0.25
Date	1.90	0.20	0.25

The analysis of referring expressions points out the graph segments that are haptically and conceptually salient. One of the major differences between the referring expressions under the three conditions was observed for curvature minimums, as exemplified in Figure 5-8(a). The curvature minima landmarks were mostly highlighted by the participants in the visual condition with labels. Meanwhile, these landmarks were rarely mentioned in the other two conditions. Moreover, small line segments that corresponded to variation against the general pattern i.e. increasing trend (the section colored by red in Figure 5-8(b)) were easily detected and frequently mentioned in the verbal descriptions for the haptic condition. However, these regions were usually ignored in the visual conditions. This can be due to saliency originating from friction provided by the force feedback mechanism of the haptic device. Small changes congruent with the general pattern can be traced very easily, but the landmarks that contain significant changes, also require more physical effort during tracing since the haptic representation of this kind of points can be relatively difficult.

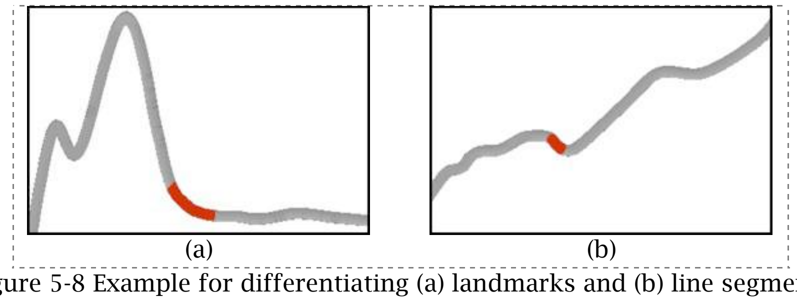


Figure 5-8 Example for differentiating (a) landmarks and (b) line segments

A.2 Use of Modifiers

In addition to the analysis on the mention rate of shape landmarks and line segments, the effect of the sensory modality in the production of modifiers (adjectives and adverbs) that occurred a lot in these expressions was also investigated. The results of a Mann Whitney U test showed that the participants in the haptic condition tended to use more modifiers in their descriptions (M= 3.65, SD= 1.33) than the visual graphs with labels (M= 2.62, SD= 1.22), U= 584, p< .05 and the visual graphs without labels (M= 2.33, SD= 1.59), *U*= 543, $p < .05$. As for the difference between the visual graphs with and without labels, it was not significant (*U*= 1113, $p > .05$).

A Kruskal-Wallis test indicated a significant main effect of the modality on the production of adjectives, $\chi^2 = 13.8$, $p < .05$. However, modality did not have an effect on the production of adverbs, $\chi^2 = 4.39$, $p < .05$. A more detailed analysis was conducted on the participant's preference of modifiers (adjectives or adverbs), since adverbs modify actions while adjectives modify shapes. This distinction is important since it may point to differences related to the sensory modality and also to possible help content for verbal assistance system. A Wilcoxon Signed Rank test indicated that while there was no significant difference between the number of adjectives (*M*= 1.5) and adverbs (*M*= 1.2) mentioned during the verbal description of visual graphs with labels (*Z*= -1.05, $p > .05$), participants produced more adjectives (*M*= 1.5) than adverbs (*M*= .9) in the visual condition without labels (*Z*= -2.53, $p < .05$). Similar to visual graphs without labels, the production of adjectives (*M*= 2.3) was significantly higher than adverbs also for haptic graphs (*M*= 1.3), *Z*= -2.62, $p < .05$. This indicates that the graphs without labels resulted in the production of more adjectives, the majority of which are shape-based. Especially, the expressions observed in the haptic condition (e.g., *like staircase shape*, *exponential increase*, *wave shaped*, *valley phase*, *damped cosine*, etc.) were highly distinguishing from the adjectives produced in the visual conditions in which the trends were mostly described in terms of the direction of the line segment (such as increasing/decreasing/fluctuating trend etc.).

B. Interim Discussion for verbal descriptions

Comparative analyses of the verbal descriptions produced by the haptic and visual explorers exhibited similarities as well as differences between the haptic and the visual graphs. Certain aspects of the graph segments turned out to be more difficult to acquire in the haptic modality than the visual one, largely due to the sequentiality and locality of the perception with a narrow bandwidth of information in the haptic modality. Due to those differences between the modalities the referring expressions that were produced by the participants exhibited diversity for the same type of graph entities under different sensory modalities, as reflected by the comparative analyses between the three groups of participants in the experiment.

This indicates that in communication through line graphs, line segments are a more prominent source of information compared to shape landmarks.

Another finding that was obtained in the reported experiment was that the presence of data labels (in the visual modality) resulted in an increase in reference to the intermediate landmarks, whereas reference to the end points was influenced neither by the modality nor the presence of data labels. A further analysis of the reference type (i.e., "value", "time" and "term") revealed that expressions used to refer to quantity and time (i.e., "value" and "time" categories) were more frequent in the presence of data labels. As expected, quantitative relations were difficult to acquire in the haptic condition. On the other hand, the production of referring expressions belonging to the "term" sub-category revealed a different picture: The haptic graphs and the labeled visual graphs resulted in a similar number of "term" referring expressions, both being more frequent than the visual graphs without labels. The similarity in the production of "term" referring expressions between non-labeled haptic graphs and labeled visual graphs may indicate that the participants who inspected the haptic graphs were able to conceptualize the represented information in a similar way to how the other participants did with the visual graphs with labels (rather than the participants did in the visual graphs without labels).

Finally, the participants, who verbalized the haptic graphs produced more modifiers (mainly, adjectives) compared to the visual graphs. Through the haptic modality, the participants produced shape adjectives that possibly made the memorization of the represented information easier for the perceiver.

5.3.1.2 Speech-Accompanying Gestures

Gestures comprise a highly relevant visuospatial modality of communication because they share similar perceptual and visuospatial features to convey the concepts such as quantity and relations (Tversky, 2011) and they express spatial information in communication through graphs (Acartürk and Alaçam, 2012). Therefore the analysis of speech-accompanying gestures[17] has the potential for providing complementary data to the analysis of verbal descriptions and referring expressions. Furthermore, haptic exploration and gesture production share underlying mechanism for performing hand movements. Due to this common mechanism, it is hypothesized that the effect of sensory modality has an effect on gesture production.

Gesture Annotation. As stated before, the experiment session consisted of five haptic line graphs for 31 participants (in three conditions), leading to 155 verbal descriptions, and 252 representational gestures.

Gestures were annotated w.r.t the annotation scheme provided in detail in Chapter 4.5.3. Each representational gesture was classified as *static* or *dynamic*. According to this classification, the hand movements conducted in a small space without having any directed trajectory were categorized as *static (non-directional)*[18], whereas the hand movements with the aimed trajectory on the air were classified as *dynamic (directional gestures).* To illustrate, the hand shape in a form of reverse L-letter was one of the frequently used *static* type of gesture to refer to peak value. On the other hand, the diagonal upward movement or drawing wave on the air is two examples of a *dynamic* gesture. The directional gestures (dynamic) were also classified into two categories; (i) *the single direction (*vertical/diagonal/horizontal), and (ii) *the multiple directions.* The gestures that contained movement in only one direction (such as upward) were classified under the "*single direction*" category, while category of the "*multiple directions*" covers the gestures formed with the combinations of the one-directional gestures in different directions. Two coders analyzed and classified the data. Interrater reliability was calculated by Cohen's kappa. The results revealed a value of .70 which indicates a substantial interrater agreement.

A. Results

A.1 The number of representational gestures

Haptic graphs (without data labels). All participants produced gestures for at least one stimulus during their session. For 75% of the protocols, speech-accompanying gestures (*N* =88) were observed. The results of a Chi-square test showed that more directional (dynamic) gestures (*N* = 76) were produced than static gestures (*N* = 12), $\chi^2(1) = 46.5$, $p<.05$. A more detailed analysis on directional gestures revealed that the participants produced the same amount of gestures consisting of hand movements with multiple directions (*N* =38) and with single direction (*N* =38).

Visual graphs with data labels. In each session, five visual line graphs were given to one of the eleven participants, leading to 55 verbal descriptions, and 102 representational gestures in total. Six participants produced gestures that accompany to their verbal descriptions of the seen graphs. For %44 of the protocols, speech-accompanying gestures (*N* =102) were observed. The results of a Chi-square test showed that more directional gestures (*N* = 67) were observed than static gestures (*N* = 35), $\chi^2(1) = 10.1$, $p<.05$. Unlike

[17] The results of this experiment regarding speech-accompanying gestures were partially published in Alaçam, Habel and Acartürk (2013a).

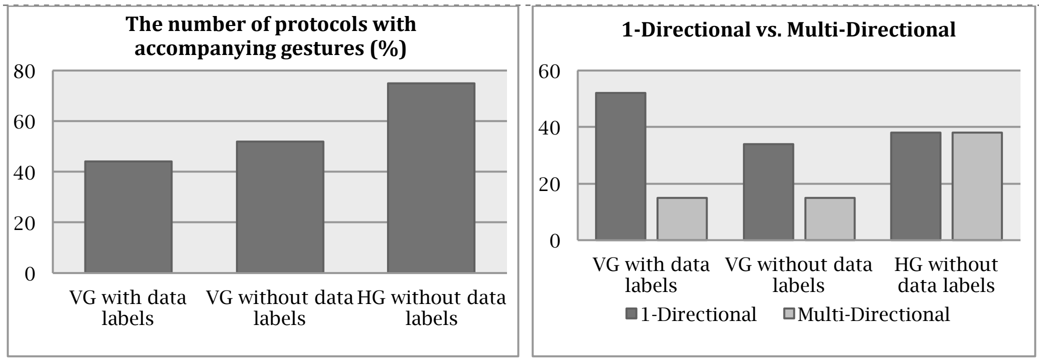
[18] It should be noted that a static gesture can be directional as well (i.e. a static diagonal hand posture) and such cases were annotated in line with the gesture annotation scheme that I employed. However, they were extremely low in this data set, therefore here the static gestures connotates non-directional gestures.

the pattern observed in haptic graph condition, a significant difference was observed between the directional gestures with multiple directions (*N* = 15) and single directions (*N* = 52), $\chi^2(1) = 20.4$, $p<.05$, indicating that participants tend to describe the events with simple, one directional hand movements.

Visual graphs without data labels. Eight of the ten participants produced gestures accompanying their verbal description of the graph. For %52 of the protocols (26 of 50 protocols), speech-accompanying gestures (*N* =62) were observed. The results of a Chi-square test showed that more directional gestures (*N* = 49) were observed than static gestures (*N* = 13), $\chi^2(1) = 20.9$, $p<.05$. Similar to the visual condition with labels and unlike the pattern in haptic graph condition, a significant difference was observed between multiple directional gestures (*N* = 15) and single directional gestures (*N* = 34), $\chi^2(1) = 7.4$, $p<.05$, indicating that participants tend to describe the events with simple, one directional hand movements, although there is no data labels that make segmentation easier.

The effect of the sensory modality. In the haptic condition, the number of protocols accompanied by at least one gesture (%) is significantly higher than in the visual condition without labels ($\chi 2 = 6.85$, $p < .01$.). Furthermore, the results of a Pearson's Chi-square (see Figure 5-9) revealed that the difference between haptic graphs and visual graphs without labels in terms of the preference pattern for gestures (static versus directional) is not significant, $\chi 2 = 1.41$, $p > .05$. Finally, while the participants prefer to produce single directional gestures for the graphs presented in both visual conditions, participants in the haptic condition preferred to use multiple directional gestures as well as single directional gestures in their verbal descriptions, $\chi 2 = 4.58$, $p<.05$.

The effect of data labels. The number of protocols accompanied by at least one gesture in both visual conditions was similar $\chi 2 = .73$, $p > .05$, (see Figure 5-9). Moreover, the results of a Pearson's Chi-square test also revealed that there was no association between the existence of data labels and the gesture type (static versus dynamic (directional)) $\chi 2 = 3.32$, $p > .05$, in both conditions more directional gestures were produced than static gestures (see individual results presented above for each condition). A Pearson's Chi-square test also revealed that the association between the existence of data labels and the variety in the directional gestures (single or multiple) was not significant, $\chi 2 = 0.99$, $p > .05$ for the visual conditions.



(a) (b)

Figure 5-9 Comparison of (a) the number of protocols with speech-accompanying gestures (b) 1-directional and multi-directional gestures

This result implies that while there was no difference in terms of directional gesture types for haptic graphs, there was a significant difference in the visual graph exploration.

A.2 Qualitative Evaluation

A qualitative analysis which is more specific to gestural representation of metrical information was performed as well. A closer look to the data revealed that the sensory modality of the representation has an influence on the spontaneous speech-accompanying gestures. As an example, consider the gesture trajectories, the verbal descriptions and the original haptic stimuli presented in Figure 5-10.

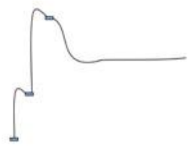
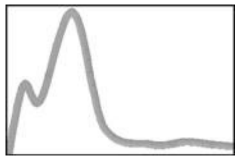
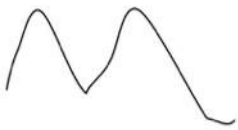
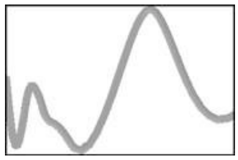
	Gesture Trajectory	Haptic Stimuli	Verbal Descriptions
(a)			"It is similar to the one before with the difference that it was a bit deeper. Then it had a small peak, than a large peak and somewhere in the midfield it levels off."
(b)			"The graph behaves like a damped cosine wave that means it begins at a high value to then stronger oscillating fluctuate with a further increasing amplitude over time."

Figure 5-10 Left column: The trajectories of the gestures accompanying the verbal descriptions; the rectangles indicate pauses on the gestural movement. Middle column: The haptic graph-stimuli. Right column: The speech accompanying the gestures.

The shape of the gesture trajectory in Figure 5-10(a) is similar to the original stimuli. Meanwhile, the metric properties of the two (i.e., the gesture trajectory and the original graph stimuli) are not similar. In particular, the almost-horizontal line segment in the middle part of the graph is much higher in the gesture trajectory compared to the stimuli. Figure 5-10 (b), the shape of the gesture trajectory and the shape of the graph line also have a high similarity. However, the peak locations, as described by the gesture trajectory, are not metrically correct. In the participant's drawing, the first peak has the same value with the second peak, which is quite different from the original graph. Those findings indicate that gestures convey qualitative information about graph shape at a level of coarse-granularity.

B. Interim Discussion for Speech-accompanying Gestures

In this experiment, the effect of the modalities was investigated using the gestures produced by the participants. In both the haptic and the visual conditions, more directional gestures were produced supporting the idea that line graphs emphasize trend conceptualization (Zacks and Tversky, 1999). Furthermore, during haptic graph comprehension, the production of multiple-directional gestures that highlight the general pattern in the segments or in the entire graph, was observed as well as the production of gestures that point out one directional segment in the graph. This kind of "as-a-whole" comprehension might have been facilitated due to sequential perception of the data. Still, more directional gestures were produced compared to static gestures during visual graph comprehension, but the difference is not as large as in the haptic version. Additionally, the participants tended to produce gestures with single direction (such as only an upward or downward movement) more than multi directional gestures, indicating that visual exploration enhances the segmentation of the events with respect to visually salient points. The comparison between the haptic and the visual modalities also shows that haptic exploration has an influence on the production of gestures during verbal description, possibly due to the alignment between shared spatial properties of gestures and haptic exploration.

5.3.1.3 Post-Exploration Sketches

A third experimental modality of analysis is the post-exploration sketches produced by the participants. The sketches represent what retains in the participants' memory after haptic or visual exploration. As such, they can be an indicator of how well the depiction matched the observers' mental image of the graph shape. The sketches produced in all conditions were analyzed based on their similarity rankings compared to their original counterpart in the graph set. Two coders ranked all sketches for each graph, this means that 31 sketches produced by 31 participants for each graph type were ranked from 1 (most similar) to 31 (least similar). For the analysis of the sketches, inter-rater reliability between the two raters was assessed using a two-way mixed, consistency average-measures ICC (Intra-class correlation). The resulting ICC (=.73) was in the "good range", as identified by Cicchetti (1994). A non-parametric Mann-Whitney U test indicated that the sketches produced for the visual graphs with labels (M=13.26, SD= 6.81) and the ones produced for the visual graphs without labels (M=15.54, SD=8.54), exhibited similar scores, U=1276.00, z=-1.41, p>.01. The effect of the sensory modality was also observed in the similarity rankings: The results showed that the sketches produced for the visual graphs without labels (M=15.54, SD=8.54) were ranked with lower scores indicating a higher similarity to the original graphs compared to sketches for the haptic graphs (M=19.91, SD=7.09), U=872.00, z=-2.53, p<.01, note that an increase in ranking corresponds to lower similarity.

5.3.1.4 Haptic Exploration Movements

Aside from the effects of the sensory modality and the existence of data labels, the results presented so far also highlight the contribution of the graph shape on haptic graph comprehension. Here, explanatory analysis on the haptic exploration movements performed during the exploration of shape landmarks and segments and also on the differences in the exploration speed when it comes to line segments with varying steepness values were conducted. The raw data set consisted of the recorded locations of the haptic stylus on the 2D horizontal plane (horizontal and vertical coordinates). As the first step, the raw data points were segmented and labeled with respect to the landmarks for each graph (see Figure 5-4 for Qualitative Ascriptions). After the labeling of each data point, the Euclidean distance between pairs of consecutive data point was calculated. The time spent for the exploration of each landmarks, and segment and the average speed were calculated, for this calculation since each line segment has different length, the data has been corrected w.r.t. time spent for 1-unit.

Shape Landmarks. Shape landmarks were categorized into three conceptual categories: (1) global or local maxima, (2) global or local minima and (3) other smooth points. Since exploration has to stop at end points (both left and right ends), the data that belongs to those landmarks were excluded from the analyses. Table 5-6 presents descriptive statistics of the average speed and the time spent on identifying the shape landmarks. The results showed that the average speed was significantly affected by the landmark type (F(2,18)= 10.96, p=.001, η^2=.55 (by using a Bonferroni correction). Pairwise comparisons revealed that explorers passed maximum landmarks with higher speed compared to minimum landmarks (F(1,9)=20.76, p<.001 η^2=.70) and also to other smooth points (F(1,9)=14.59, p<.01, η^2=.62). But there was no significant difference between the latter two (F(1,9)=1.53, p>.05). Shape landmarks correspond to instantaneous events therefore they are represented with very limited region on a graph. We can assume that they are length- invariant, therefore as wells as the exploration speed, the average amount of time spent were also calculated and compared w.r.t three types of conceptual landmarks. The results showed that the average amount of time spent was significantly affected by the landmark type (F(2,18)= 62.13, p<.001, η^2=.87. Explorers spent more time on landmark maxima than on minima (F(1,9)=77.99, p<.001, η^2=.90) and smooth points (F(1,9)=154.39, p<.001, η^2=.95). But, again, there was no significant difference between the latter two (F(1,9)=2.08, p>.05). These two results are in line with the observations regarding haptic exploration on the attended locations, namely explorer's repetitive fast back-and-forth movements on a particular location.

Table 5-6 Descriptive statistics for the average speed and time spent for shape landmarks

	Maxima	Minima	Other Points
Average Speed	M=1.03 (SD=.01)	M=1.0 (SD=0.1)	M=1.01 (SD=.01)
Time Spent (total)	M=4.45 (SD=.32)	M=3.1 (SD=0.37)	M=3.31 (SD=.19)

Shape Segments. Shape segments were categorized w.r.t their steepness (steep, moderate, slight and horizontal; see qualitative ascriptions table for details). Since the average time spent is tightly related to the segment length, the average speed is considered as an indicative parameter (see Table 5-7 for descriptive statistics). The results showed that the average speed of the explorer was affected by the steepness of the segment ($F(3,27)=72.23$, $p<.001$, $\eta^2=88$). Steep segments were explored with lower speed compared to moderate segments ($F(1,9)=106.22$, $p<.001$, $\eta^2=92$) and horizontal segments ($F(1,9)=105.90$, $p<.001$, $\eta^2=92$) but with higher speed compared to slightly steep segments ($F(1,9)=13.34$, $p=.005$, $\eta^2=60$). The average speed for slight segments was also significantly lower than that for the moderate ($F(1,9)=104.04$, $p<.001$, $\eta^2=92$) and the horizontal ones ($F(1,9)=121.42$, $p<.001$, $\eta^2=93$). However, there is no significant difference between moderate and horizontal values ($p>=.05$).

Table 5-7 Descriptive statistics for shape segments

	Steep	Moderate	Slight	Horizontal
Average Speed	M=.99, SD=.009	M=1.02, SD=.01	M=.98, SD=.008	M=1.02, SD=0.11

These results demonstrated that the exploration of global or local maxima was different than the exploration of other points on the graph. The qualitative ascriptions for those points correspond to the concavity on the domain of shape geometry. On the other hand, the exploration pattern for minimum points, corresponding to the convexity of the shape, was not distinctive than any other smooth point on the graph. Additionally, the systematic differences between the different types of shape landmarks and line segments were observed. This simple fundamental analysis indicated that haptic exploration patterns might provide useful information in the investigation of the effect of amodal properties. The geometric properties of the graphs were not among the controlled parameters in this experimental design. The next chapter provides a more comprehensive analysis on the effect of amodal geometric properties on haptic graph comprehension.

5.3.2 Interim Discussion for the Empirical Study-I

Table 5-8 summarizes all the results (together with the significance and the effect sizes) provided in this part. The study of the production of referring expressions, the analysis of the gestures and the analysis of the post-exploration sketches in the different sensory modalities is a necessary step for the development of verbally assisted haptic exploration systems. Verbally assisted haptic exploration has a dialog-like character, even if—as in the empirical studies so far—the haptic explorers do not communicate verbally with the assistant. Based on the analysis of the verbal descriptions that were produced in the graph description tasks, systematic relationships in the production of referring expressions under different experimental conditions were found. The findings indicate that both the sensory modality and the presence of data labels influence the production of referring expressions by humans.

The analyses performed in the context of this empirical study also provided evidence about the detection of important shape properties and how they are conceptualized (as states such as peak or sinus or as actions such as increase or decrease trends). The distinctive properties of shapes (such as shape landmarks or salient line segments) are helpful elements for segmentation. The results emphasize that shape is a basic element of verbal descriptions, and the geometric properties of shape segments (positive vs. negative, smooth vs. steep) seem to affect the way how the readers segment graphs. In particular, the saliency of shape landmarks has an effect on the haptic saliency, which plays a crucial role in the detection of these regions.

1-line graph comprehension was also investigated by a comparative analysis of gestures that accompany verbal descriptions of both visual and haptic graphs. The results showed that the modality of the representation has an influence on the gestures that accompany the descriptions.

Table 5-8 Significance and effect sizes for Experiment-1

5.3.1.1 Verbal Descriptions			
	VG with labels	**VG without labels**	**HG without labels**
A. The Reference Rate			
Line versus Landmarks	Sig. (χ^2=17.4)	Sig. (χ^2=64.9)	Sig. (χ^2=22.0)
Reference Rate for Landmarks	Sig. (χ^2=16.7)		
		Sig. (χ^2=9.9)	
Landmark Type			
Intermediate vs. End Points	Sig. (χ^2=19.3)	No. Sig.	No. Sig.
Intermediate Landmarks	Sig. (χ^2=5.6)		
		No. Sig.	
Term	Sig. (χ^2=9.1)		
		Sig. (χ^2=8.4)	
B Use of Modifiers			
Main Effect	No. Sig.		
		Sig. (U=584)	
Adjectives and Adverb	No. Sig.	Sig. (z=-2.53)	Sig. (z=-2.62)
5.3.1.2 Speech-Accompanying Gestures			
# of protocols with accomp. gesture	No. Sig.		
		Sig. (χ^2=6.85)	
Static vs. Dynamic	Dynamic	Dynamic	Dynamic
1-Directional vs. Multi-Directional	1-Directional	1-Directional	No. Sig.
Multi-Directional	No. Sig.		
		Sig. (χ^2=4.58)	
5.3.1.3 Speech-Accompanying Gestures			
Similarity Score	No. Sig.		
		Sig. (U=872.00)	
5.3.1.3 Haptic Exploration Movements			
Shape Landmarks	Not applicable		Sig. (η^2=.55)
Average Speed			Faster for Max.
Time Spent			Longer for Max.
Shape Segments			Sig. (η^2=.55)
Average Speed			Faster for Moderate and horizontal

5.4 Empirical Study-II on Haptic Line Graph Exploration: Visually Impaired Users

As emphasized in earlier chapters already, many factors take part in graph comprehension, ranging from bottom-up perceptual influences to high-level top-down effects. One of the most important top-down effects is the role of prior knowledge on how to read a specific graph type, namely the facilitation of appropriate graph schemata. In the previous study, blind-folded sighted participants were used in the investigation of haptic graph comprehension in order to keep this factor constant as much as possible. However, empirical studies are needed to understand the differences in terms of the graph readers' needs. The following section presents the findings from an experimental study on haptic line graph exploration with visually impaired participants[19].

[19] The results of this experiment regarding speech-accompanying gestures were published in Acartürk, Alaçam and Habel (2014).

5.4.1 Experimental Setup

Participants. Eight visually impaired participants, of which three were partially sighted, participated in the experimental study. The participants were teachers at an elementary school for the visually impaired in Turkey. The category of blindness was identified by various parameters, such as the onset-time of blindness (*congenital-early* or *late*) and level of deprivation (*total*, *severely blind* or *partially sighted*) (Vecchi and Cattaneo, 2011). Only congenitally blind participants with or without partial sight took part in the present study.

Procedure and Materials. The experiment was conducted in single sessions. Each session took approximately 1 hour. The sessions were audio and video recorded. The experiment was conducted in Turkish, the native language of all participants. Before the experiment, a warm-up session was conducted to familiarize the haptic explorers with the Phantom Omni® Haptic Device. In the experiment session, each participant was presented with five haptic line graphs with smooth edges (two additional graphs were employed for in the familiarization process). The graphs were presented in random order. The participants did not have any time limitation. After the experiment session the participants were asked to present single-sentence verbal descriptions of the explored graphs to the experimenter. They were also asked to fill in a spatial term survey (see APPENDIX-D) that contained the words or phrases used by the verbal assistant in the previous experiment (presented in *5.3*). The participants used a Likert scale to rate the meaningfulness of the terms, using 1 (less meaningful) to 5 (most meaningful). They were also asked to verbally explain the meaning of the terms in the questionnaire. Only basic assistance or alerts that aimed to help the participants locate themselves with respect to the reference frame (e.g., "you are at the start point", "you are out of the line") was provided by the experimenter during the course of haptic exploration. The graph set, which was employed in the previous experiment, was used here again.

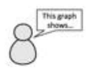
5.4.2 Results

In this section, I focus on I focus on the differences between totally blind participants and partially sighted participants in terms of their conceptualization of the events represented by line graphs (e.g., an increase in bird population). The motivation of the analysis is that the identification of the differences between those two conditions may provide information for the design of the verbal assistance content for the two user groups. Because of the small sample size for each group, the data gathered through this experiment was reported qualitatively.

5.4.2.1 Effect of an Irrelevant Feature

First of all, since the haptic representation of the line graph by the haptic device is implemented by the engraved line on a horizontal space, depth information is irrelevant, therefore it was kept constant. The depth only served the purpose of preventing the stylus of the haptic device from going outside the graph frame, thus facilitating the tracing action. The participants were informed about the workings of the haptic device and the structure of the task space. Two totally blind participants exhibited a tendency to interpret the depth information as a parameter for trend information. This is possibly due to participants' predominant perceptual habits. This finding suggests that a more detailed training session should be designed for totally blind participants so that the interface introduces itself in a coherent manner for the haptic exploration.

5.4.2.2 Verbal Descriptions



A qualitative analysis of all the exploration patterns and the utterances produced by the participants showed that the points of high curvature were easy to detect for all the visually impaired participants. The haptic line graph interface leads to a successful interpretation, most of the time, of trend changes (in terms of the changes in bird population) that are represented by those points of high curvature in the graphs (henceforth, landmarks). On the other hand, totally blind participants had difficulty detecting line segments compared to partially sighted participants. This finding was accompanied by less adequate interpretations of population events by the totally blind participants when compared to the partially sighted participants. This is possibly due to higher prior knowledge of the partially sighted participants with line graphs. egments with low trend changes.

Table 5-9 presents excerpts from verbal utterances produced by a single participant. The utterances show how the participant links the concept of “population increase” to haptic movement. In the graph presented in Excerpt.1a there exists an abrupt change in the direction of exploration movement from horizontal to vertical, and additionally the graph ends with a landmark that is also the global maximum of the graph. Therefore the participant instantly linked the “Loc1” point to population increase. However, in the exploration of the graph in Excerpt.1b, the participant first attempted to find a landmark to detect a possible maximum point. Since the graph line continues with an increase, she interpreted the distance between the upper border and the line by taking the middle point as a reference point (“Loc2”). This finding and similar findings that were obtained in the experiment show that landmarks are relatively easier to explore haptically than line segments with low trend changes.

Table 5-9 Locations of the haptic stylus on the explored graph and accompanying verbal utterances for two instance by the same participant (“E” is a haptic explorer and “A” is a verbal assistant)

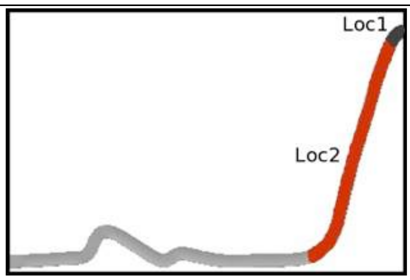
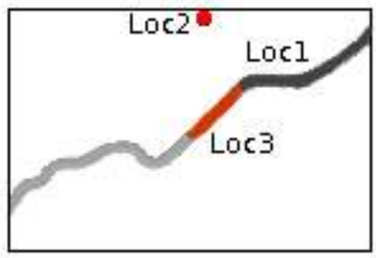
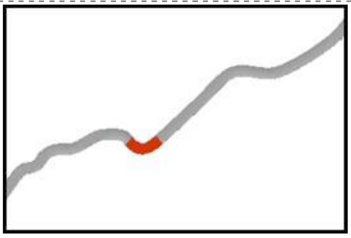
Excerpt.1a	**Excerpt.1b**
	
E: This is upside, right? (Loc1) *A*: Yes *E*: It is upside, then the population increases as it goes upward. (Loc2 and Loc1)	*E*: Lets look at the upside (Loc1) *E*: I am not out, right? (Loc2) *A*: Yes, you are out (Loc2 and horizontal movement on upper border). *E*: Then, there is no increase here (Loc1 and Loc3) *E*: I mean, if that is end point, then I am trying to find the upside since I don’t find a line upside, it means that the population did not increase (entire graph).

Figure 5-11 presents an excerpt from a partially sighted participant. Although there is a misconception about one of the local points (darker section in Figure 5-11), the details and scaling can be considered quite good. The verbal protocol shows that it is a challenging task to distinguish between global and local maxima and minima for haptic explorers, possibly due to the local sequential character of haptic perception.



"10% increase at the start. Then, it seems as if it gets 20% more. Then there is clutter. Then it sinks towards the bottom 5%. It increases from 5% to 40%, it continues on the 40% level. But, there is a slight thing here, %42 or %43. I believe there is a slight increase here. There is not, it is flat. Then it seems that it increases towards 45% and here, it is completed."

Figure 5-11 A participant's verbal description of the population in terms of percent changes

A closer look at the vocabulary of the utterances used by the haptic explorers also revealed similarities and differences between partially sighted and totally blinds. Both user groups employed a common vocabulary set, including the terms for the following concepts: "increase, decrease", "left, right, down, upside", "here, there, this, that", "slow, steep, small, big", and "something" (as a deictic pointer). Table 5-10 presents a list of additional vocabulary items, employed by the users during the course of haptic exploration.

Table 5-10 Vocabulary used by the totally blind and partially sighted participants

Total Blind	**Partially Sighted**	
Difference	Difference	Inclined towards down
Change	Change	Curve/bend
Increase or change	Point	Cavitation/ bump/bulge
Stay same	Horizontal line	Deep
Point(s)	Diagonal	Fork
Hole	Stable course	Letter J
Different point	From right to left	Hit the ceiling
Like a hill	Something which is not	Sink to bottom
Point	pyramid but a steep	With long intervals
With constant intervals	slope(geographical term)	At constant intervals
Narrow	Slope / incline	Border
	Slight round	

An analysis of the vocabulary terms revealed that totally blind participants verbally describe landmarks (i.e., trend changes) by using domain-independent terms, such as "difference", "change", without providing direction or polarity information. Partially sighted participants, on the other hand, employ a domain-dependent and graph-oriented terminology, in addition to the domain-independent terms. Moreover, the vocabulary of partially sighted participants involved "horizontal", "diagonal" and "stable", "inclined" for line segments, and "curve", "bend", "bulge", "cavitation", "deep" for landmarks. The vocabulary terms that belong the jargon of statistical graph were also observed (e.g., "hit the ceiling" and "sink to bottom").

5.4.2.3 Spatial Survey and Questionnaire

The answers to the spatial-terms questionnaire revealed another major difference between the two user groups in terms of the use of navigational terms. All participants rated the terms "upward" and "downward" with high scores (in terms of its relevance to graph exploration). The terms "backward" and "forward", however, received lower scores from the partially sighted participants. The haptic graph line is a directed line (i.e. the x-axis represented the year from the past up to now). Therefore the terms "backward" and "forward" correspond to left and right respectively. However, from the ego-centric point of view, it may refer to the back of the user as well. Possibly due to this ambiguity, they were rated with low scores by partially sighted participants who have more complete graph-domain knowledge. The totally blind participants, however, rated those terms with high scores but with wrong interpretation, namely assuming that "backward" refers to back of the participant.

5.4.3 Interim Discussion for the Empirical Study-II

The previous experiment has shown that haptic exploration of line graphs exhibits different patterns compared to visual exploration, thus leading to differences in in how difficult it is to conceptualize the events represented by the graphs. A verbal assistance system has the potential to facilitate haptic graph comprehension, as long as appropriate verbal assistance is provided to the haptic explorer. The design and development of a verbal assistance system should take into account different types of visual impairment because as stated by Cattaneo and Vecchi (2011) comprehension differs between totally blind users and partially sighted users.

In the present study, I aimed to find the similarities and differences between the two user groups, namely totally blind and partially sighted participants. The major finding of the study is that partially sighted users may have more developed prior knowledge of line graphs compared to totally blind users. Providing the same verbal assistance may lead to a suboptimal use of the system by the different user groups. Therefore, an effective assistive system that meets the different needs of the diverse target groups should adapt itself to a user's prior knowledge of line graphs. In particular, a verbal assistance system for totally blind users should introduce general graph comprehension concepts and the verbal assistance for the specific graph under exploration.

The findings obtained in the present study were also compatible with the previous blind-folded experiments in that smooth and low-curvature segments are haptically less salient regions of the graph, which are explored by smooth and relatively effortless movements with stylus. Therefore these regions can be considered as candidate locations for verbal assistance. The prior knowledge of line graphs is of crucial importance in haptic comprehension, like it is also in visual comprehension of line graphs. The findings suggest that the design of a haptic interface for the use of totally blind users should be more informative. This information can be provided by an introduction session about graph concepts and their relation to domain concepts (e.g., bird population) before the exploration starts.

Finally, the participants stated that they did not have difficulty in getting familiar with the device. All participants stated (during the experiment or after the experiment) that it became easier to stay within the engraved line as they proceeded with the exploration. Being less anxious about the device and the stylus, they were able focus on the task at hand. Overall, these finding suggest that a verbal assistance system for haptic line graph comprehension reveals a potential use and acceptance by visually impaired persons.

5.5 General Remarks

This chapter focuses on the perceptual and conceptual factors regarding perception and comprehension of haptic shapes from the perspective of graph comprehension. Accordingly, a total of two empirical studies were reported. In the first study, the effects of the sensory modality were examined to find differences and similarities in the conception of haptic graphs. Moreover, a common way of presenting graphs is to provide x and y-axes and data labels as well. However, providing haptic numerical data labels is a challenging design issue and this way of representation also has high risk of being distractive. Therefore, the effect of the presence of data labels was also investigated systematically by comparing two visual graph conditions. Systematic differences with respect to the sensory modality and the presence of data labels were found in the verbal descriptions, speech-accompanying gestures and post-exploration sketches. This line of research is particularly important for the design of verbally assistant system (from HCI perspective), namely in the decision of what kind of information should be provided in which modality. These results also highlighted the importance of facilitation of appropriate graph schemata.

In addition to differences due to modality and data labels, the findings also highlighted differences in the effects of graphical features, in particular global and local shapes. A graph line consists of a combination of various graphical entities that exhibit different properties, for example amodal properties like curvature polarity, curvature angle, size, shape and orientation. In the experiments reported on in this chapter, the umbrella term "shape" was chosen intentionally since the focus here is not on the individual contribution of those properties. The next section addresses the systematic analyses of those features and also of their role in event segmentation. Lastly, the findings were discussed in interim discussions throughout this chapter, but a more integrated interpretation of the results will be provided in Chapter 8.

CHAPTER 6

Event Segmentation and Description

The previous chapter focused on the modality-dependent differences and on the effects of data labels. While I still look at the effects of those two issues in this chapter, the particular focus is on the effect of amodal geometric features on haptic graph comprehension this time. A further issue that is considered in this chapter is how graph explorers (visual or haptic) segment events depicted in a graph, what kind of parameters govern this event segmentation process and how the graph explorers refer to those sub-events after the event segmentation has been performed[20]. This chapter virtually involves two parts: In the first part, I introduce existing literature on event segmentation and description, while the second part is dedicated to empirical research that was conducted to understand the effect of these important issues on the comprehension of haptic graphs.

As introduced in Chapter 1.2, the concept of events regarding haptic graph comprehension should be investigated in two layers: Firstly, the graphs, regardless of their representational modality, depict abstract events, i.e. a change occurring in the bird population over the years. Secondly, due to the sequential and local nature of haptic perception, the exploration of the haptic graphs yields an event-like perception and not an object-like perception, as it is the case with static visual graphs. This event-like perception resembles watching an animated object and trying to segment the event (the series of actions that the object performed in space and time) into meaningful parts. Therefore, event segmentation and description addresses both the segmentation of the abstract event and of the actively perceived static haptic object and then the production of referring expressions and gestures for these segmented entities. To my knowledge, graph comprehension has not been approached yet from this perspective in the literature. Furthermore, the similarities and differences between the comprehension of visual and haptic graphs depicting the same abstract event (in other words, the object-like versus the event-like perception/comprehension of the same entity) might also contribute to our understanding of the underlying mechanisms of object/event segmentation. Moreover, the empirical investigations conducted here also contribute to constructing design guidelines for a system that assists visually impaired people during haptic graph exploration.

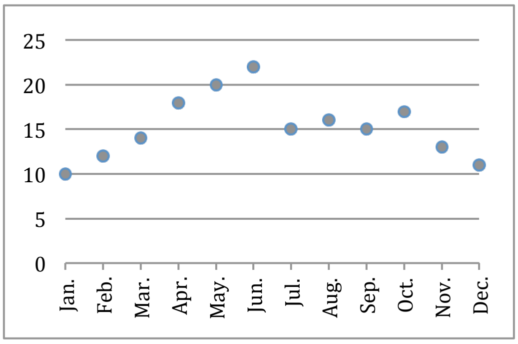
A considerable amount of literature on object segmentation and on event segmentation has been published over the past two decades (Swallow, Zacks and Abrams, 2009; Zacks, Kumar, Abrams and Mehta, 2009; Zacks, Speer, Swallow, Braver and Reynolds, 2007; Zacks and Swallow, 2007; Zacks, 2004; Zacks and Tversky, 2001; Habel and Tappe, 1999;

[20] The concept of the *event* slightly differs in philosophical and psychological perspectives, see *6.3.*

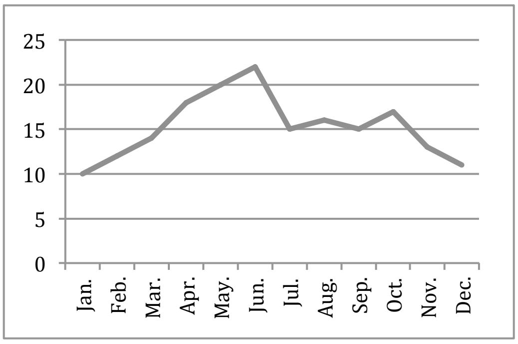
Rubin and Richards, 1985), however the previous research is mostly limited to the visual or auditory modalities. One of the aims of this dissertation is to provide some insights that may be extracted from cross-modal (visual and haptic) empirical studies. Statistical line graphs, as well as other kinds of pictorial representations, are designed to be presented in the visual modality. As such, the theoretical discussions and the empirical conclusions coming from the research focusing on the visual modality may provide a concrete basis for the investigation of event segmentation in haptic graph comprehension. Furthermore, in the investigation of event segmentation, mostly simple animations or more complex natural narratives were chosen as stimuli. Graphs are spatio-temporal representations that depict abstract events. The investigation of events represented by graphs might help to extend existing research and also provide fruitful experimental paradigm in the investigation of event segmentation. Besides, this is also a crucial design issue (from the human- computer interaction perspective) for a verbally-assistance system that aids blind people in accessing statistical graphs. Understanding event segmentation is important for providing beneficial verbal assistance. In order to have a successful communication over a graph between an explorer and an assistant, and in order to provide verbal descriptions automatically, the abstract-conceptual event represented in a concrete manner by means of a graph shape should be segmented into meaningful parts, and then the description of those segmented events should be provided in a convenient way (i.e. deciding on the details to be provided verbally and also adjusting the scope of the content for avoiding repetitive assistance regarding same part).

Figure 6-1, which is identical to Figure 1-1, illustrates two graph types ((a) a data point graph and (b) a line graph) depicting average daily temperatures. Each one has an x-axis representing the months and a y-axis indicating the temperature. In a data-point graph, dots represent the data values of quantitative features of some event depicted in a statistical graph. In a line graph, the dots are connected by a line. This property of line graphs invites readers to comprehend data points as a continuous line by employing visual Gestalt principles and make them focus on trends and such second order changes instead of focusing on the values of individual data points. Due to this spatio-temporal continuity, a line of the graph does not convey a spatial or temporal gap. In that case, an important question highly relevant for the current investigation is what governs the segmentation. As densely elaborated in the previous section, shape is the most dominant feature of a graph line. As such, the principles of shape segmentation can also be applied to this domain, suggesting that contour discontinues are the most prominent features to be used in event segmentation. The visual graphs provide static snapshots of dynamic events[21] that unfold in time; all the information is presented spontaneously to the reader. Thus they display object-like representations of the abstract events. On the other hand, a graph line is explored sequentially through actions performed by haptic explorer. Thus the line is perceived as an action differently from the object-like visual perception of graphs.

[21] Within the scope of this dissertation, I use “event” as a technical term that covers actions, states, processes. See the upcoming part (*6.3)* for more detailed information.



(a)



(b)

Figure 6-1 Average daily maximal temperature graphs presented in (a) a data-point graph and (b) a line graph

Furthermore, it is important that the graphical entities provide an effective and meaningful bridge between the original message (intended by the graph designer) and the message understood (by the graph reader). Temporal events such as population distributions, temperature etc. (on daily, monthly or yearly time scales) are converted by the graph producer into spatiotemporal representations to make trends and other second order relations easy to comprehend. Then, graph readers have to unfold (perceive and comprehend) the event depicted by the graph. For averaged data, the line between data points does not have an actual value; instead, they serve the purpose of highlighting second-order information such as trends between two data points. Although these lines are fictional added with a function of diminishing the saliency of data points to highlight the trends, this does not mean that misconceptualization regarding data points can be acceptable. Data-labels, one of the standard elements of graphs, help a reader in segmentation. A line graph without explicit data point marks (plain line graphs) but with labels can be still segmented (*in a discrete/categorical way*) or comprehended as a whole (*in a continuous scale*). On the other hand, a line graph that lacks such explicit marks and axis labels can only be considered and perceived as continuous. As a result, the graph representing averaged (i.e. monthly) data might be highly misconceptualized in such representations. The reference frame may help the reader make estimations about the numerical quantity concerning y-axis, at least in a proportionate way. However, x-axis mostly convey ordinal information as opposed to scale which makes the segmentation without cues harder, thus carrying this information is more important and crucial for x-axis labels. The lack of data labels in visual graphs or the availability of sequential perception only (as it is the case with haptic graphs) may hinder the properties, which are highly relevant and essential for event segmentation. To sum up, continuous line representation of data points foregrounds shape and haptic explorers are good at picking up shape information. Providing trend information with a continuous representation facilitates comprehension and also makes comprehension of relations easier. Therefore, event segmentation can benefit from loading hard to encode information such as data labels into appropriate verbal content.

Event segmentation studies and theories in the literature seem to always have tight links and references to object segmentation research. As aforementioned, both object and event segmentation are considered relevant for this dissertation and are therefore presented side-by-side in the rest of this chapter. The distinction between objects and events as well as their commonalities are well–investigated. The first basic distinction is in their nature of being: Objects exist and events occur or take place (Casati and Varzi, 2002; Cresswell, 1986). Another primary distinction is in their relation with time and space. Objects mostly have tight spatial boundaries and not so tight temporal boundaries. Meanwhile, events are more bounded in time. The other way around, spatial dimension highlights the sequence of states/objects that take part in events, and temporal dimension highlights the sequence of actions that occur in events (Zacks et al., 2001).

But despite their differences, there are researchers who tend to put objects and events together as they do not find these distinctions to be significant (Goodman, 1951). Both events and objects are considered as individual entities that possess spatially partonomical (part-whole) relations. Events display temporal structures on top of it. However, both objects and events can be "counted, compared, quantified over, referred to, and variously described and re-described" (Casati and Varzi, 2002). That is the reason why they are treated as entities of the same kind under some conditions, since "an object would simply be a "monotonous" event; an event would be an 'unstable' object" (Goodman, 1951).

The first virtual part that concentrates on existing research on event segmentation and event reference, addresses three main questions. First, the question of *why people segment* is elaborated on in 6.1. Then, the second part (6.2) investigates the question of *how people segment* by describing the theories on event segmentation, the factors that have an effect on the segmentation such as partonomic and taxonomic structures, and how event segmentation affects cognitive processes ranging from perception to memory encoding. The question of *how people describe events and sub-events* is discussed in the third part (6.3). After that, the comprehensive empirical investigation on event segmentation and reference is presented in 6.4.

6.1. Why Do People Segment Continuous Data Streams into Discrete Units?

Segmentation of events into sub-events elicits many benefits. These benefits can be inspected under two main topics. First, event segmentation had been known to help perception and second the research on this has also demonstrated its contributive role in language processing, memory, planning and action (Zacks, Tversky, Iyer, 2001; Reynolds, Zacks and Braver, 2007). Segmentation reduces a continuous stream of data into discrete events. Thus, it is considered to be an influential and economical perceptual operation (Zacks and Swallow, 2007). During this segmentation process, some features are highlighted while others are backgrounded in a way that makes this newly derived representation more useful for cognitive processes such as memorizing, and reasoning (e.g. Rosch, 1978; Tversky and Hemenway, 1984; Tversky, Zacks and Hard, 2008). The studies on object perception suggest that people generally do not perceive space as consisting of a continuous spectrum of features such as color and texture. Rather, they perceive space as segmented, as consisting of spatially coherent objects. As already suggested in several articles (Zacks and Swallow, 2007), this is not just limited to space; time also exhibits the same characteristic. As such, events are also perceived as discrete entities just like objects. In the very comprehensive review on event segmentation by Zacks and Swallow (2007), three conclusions about event perception have been drawn: First, event segmentation occurs automatically as a part of ongoing perceptual processing. Prior research on object recognition (e.g. Biederman, 1987), demonstrated that segmenting an object into spatial parts is crucial for recognition. Similarly, the empirical research on event segmentation also revealed that segmenting events in time is crucial for both understanding and recalling them for later use. Secondly, it has been suggested that the way how people remember events is highly dependent on how they segment it in real time in the first place. The underlying idea behind this is that segmentation packs the information and forms the units of memory encoding. Good memory and learning are achieved through correct event segmentation, while wrong segmentation leads to poor memory and learning. One of the roles assigned to event segmentation is separating "what is happening now" from "what just happened". In further research (Swallow et al., 2007), it also turned out that event segmentation influences the contents of both short-term and long-term memory, thus providing a structure for event memory. Similarly, the authors (Swallow et al., 2009) claimed that "perceptual event segmentation reflects a control process that regulates the contents of activity memory". Basically, they interpret this as an evidence for the boundaries in event perception being also boundaries of the memory. The final point of the review (Zacks and Swallow, 2007) mentions the neural mechanisms underlying event segmentation. It has been conclusively shown that there are specialized neural systems that process either perceptual or conceptual features introducing a change in a continuous data stream. These neural systems also play a role in identifying event boundaries by means of these feature changes.

6.2. How Do People Segment Events?

The results of numerous studies indicate that event segmentation is affected by low-level perceptual changes in a bottom-up manner (e.g., changes in action and movement characteristics) and high-level conceptual features of activity in a top-down manner (e.g. change in the goals of the actors) (Speer, Zacks and Reynolds, 2007; Zacks, 2004). When segmenting the continuous stream of data dispersed over time, people pack this information into discrete and meaningful units and form categories of events. Then, these categorized sub-events are organized with respect to their taxonomic (relation of kinds; e.g. Morris and Murphy, 1990) and partonomic relations (relation of parts; e.g. Tversky, Zacks and Hard, 2008; Zacks et al., 2001; Zacks and Tversky, 2001) The locations that an event is segmented in time have many different names in the literature, such as *event markers*, *event boundaries* or *breakpoints*. Though slightly different, they denote the same function. Thus, I use all three terms interchangeably. This sub-part concentrates on the literature on the role of breakpoints and the effect of partonomic and taxonomic relations on event segmentation.

6.2.1. Breakpoints (Event Boundaries)

A vast amount of empirical studies highlights the importance of breakpoints in the perception and understanding of activities. First of all, the studies showed that people mostly agree on the location of event boundaries when they are asked to segment an event. Besides, these boundaries seem to be hierarchically organized. When these boundaries are artificially deleted, the understanding of the event seems to suffer as they correspond to locations in time that contain rich, important and memorable information (e.g., Hard, Tversky and Lang, 2006; Schwan and Garsoffky, 2004; Zacks, 2004; Zacks and Tversky, 2001; Newtson and Enquist, 1976). The empirical evidence on the function and processing of breakpoints (Zacks et al., 2009; Hard et al., 2006; Zacks, 2004) indicates that event boundaries are loaded with lots of information, that is, they can be regarded as the most informative moments in an action stream. The probability of a drastic action change is greater at breakpoints than at ordinary moments in the action stream, making event boundaries more memorable than other entities of the event. This change may have important consequences for how people process activities in real time. Breakpoints or event boundaries are, function-wise, similar to inflection points on an object's contour: They correspond to discontinuities in the continuous stream of information just like inflection points in an object contour (Zacks et al., 2009). They separate events into subparts and they also bind them. Once the current action is completed a new action occurs and breakpoints function as connection points between two consecutive sub-events (Mennie, Hayhoe and Sullivan, 2007). Furthermore, they link one level of action to another, namely coarser actions to finer actions both on a partonomic and taxonomic level (Hard, Recchia and Tversky, 2011). In other words, breakpoints link perceptual features to conceptual features, that is, an observed action to its goal. Therefore, they are not only boundaries that separate actions but they are also a source of rich information formed with well-selected concrete and abstract features, which foreground the shape of ongoing action.

In related research, in order to measure segmentation, participants usually are asked to explicitly identify e*vent boundaries* (Newtson, 1973). One commonly used method was introduced by Newtson (1973): First, the participants watched a movie showing an event and pressed a button to mark an event boundary whenever one event ended and another one began. Several studies (Newtson, 1976; Zacks et al., 2001) using this methodology showed that boundaries between events demonstrate good consistency and reliability among participants (Newtson, 1973; Speer, Swallow and Zacks, 2003). Functional neuroimaging studies also suggested that (Speer et al., 2003; Zacks et al., 2001; Zacks, Swallow, Vettel and McAvoy, 2006), observers show a tendency to segment events at points of salient perceptual or conceptual changes.

6.2.2. Partonomy and Taxonomy of Events

A breakpoint separates an event into sub-units. This segmentation is not a random process. Both perceptual (i.e. the position of an object or the trajectory of action) and conceptual factors (i.e. the intention and the task) have an influence on it, and hierarchical relations among the features are one of the key components that govern event segmentation (Hard, Tversky, and Lang, 2006; Newtson, Engquist and Bois, 1977; Zacks, 2004; Speer, Zacks and Reynolds, 2007). Similar to how objects form hierarchies of kinds and parts in space, events have a relational structure in regards to time. Previous research indicated that events and objects have similar partonomic and taxonomic features and these features interact in the same ways (Zacks and Tversky, 2012; Zacks et al., 2001) Graphs in a physical form, conceptual events that are depicted by these graphs and their event-like sequential perception through haptic modality also display these two kinds of relations. Hence, the investigation of these relations may be crucial for the design of a verbal assistance system and, in particular, to understand how events should be segmented and described in different granularities.

Partonomic Relations: The hierarchical relationship between parts and subparts constitutes a partonomy and partonomic relations (Miller and Johnson-Laird, 1976; Tversky, 1990; Tversky and Hemenway, 1984). The literature on the partonomic relations concerning the object segmentation demonstrated that part decomposition is an automatic process in which contour discontinuities or maxima in local curvature play important roles for segmentation and recognition (Biederman, 1985; Hoffman and Richards, 1984). The locations at which parts mostly join, exhibit contour discontinuities and these distinctive physical features can be used to identify the shape of an object (Rosch, 1978). Segment identification also seems to be under the influence of changes in contour length, in turning angle (curvature), and in orientation (Cohen and Singh, 2007; Klatzky, Lederman and Reed, 1987).

As reported by several studies (Zacks et al., 2001; Tversky and Hemenway, 1984), parts have both perceptual and functional roles, i.e. different functions can be ascribed to similar perceptual parts. For example, the lowest value on the graph line can be conceptualized as one of the any data points, or as a global minimum". Or, if the general trend is also compatible with it, it can be used for predictions (i.e. the possible extinction of the species represented in the graph). Tversky and Hemenway (1984) investigated this effect of perceptual and conceptual features on segmentation with several experiments. They suggested that the parts' ability to carry these two features allows the perceiver to make inferences about the object's function based on its appearance. In other words, their study highlighted the parallelism between functional and perceptual features, concluding that functionally significant parts are also perceptually salient, such as the wheels of a car or the trunk of a tree (Tversky and Hemenway, 1984). In that sense, the research in event segmentation goes hand-in-hand exhibiting another similarity with the object segmentation domain. The empirical studies demonstrated that people tend to divide activity at locations that exhibit salient physical change when asked to identify event boundaries. These salient changes in the features attribute functions like contour discontinuities in objects.

Thibadeau (1986) proposed a computational scheme for identifying event boundaries in simple animations and hypothesized that second-order changes are important for detecting psychological boundaries. According to him, event boundaries do not occur at the first-order changes. Instead, they are detected based on the changes at the first-order changes and thus are called second-order changes. To exemplify, constant-speed does not highlight a boundary, but change in the speed (acceleration) does. Statistical graphs provide unique example for such cases.

Graphs are smart designs that carry specific messages and foregrounds conceptual information by employing Gestalt principles of visual perceptual organization. A line graph, for example, is usually used to highlight trend information. Thibadeau's second-order hypothesis in the graph domain can be exemplified by the depictions in Figure 6-2. First of all, while Thibadeau refers to real changes in the velocity of the objects presented in an animation, here, a visual graph provides these abstract changes, which the event depicted in that graph possesses, in a static way. Visual graph readers do not observe acceleration as an action; instead they see inflection points that imply change in acceleration. In these generic line graphs, change occurs at *t3* on the x-axis. In the graph in the upper left corner, *t3* has a different value from the values of the previous and next data points. Yet, since the rate of change remains the same, this point does not provide any clue for segmentation. In the graph in the upper right corner, the line segments between *t2-t4* still have an increasing trend but the rate of change is slightly different, and the data point for *t3* starts to form a landmark in this location. In the graph presented on the bottom left, the rate of change of the increasing trend is way more prominent compared to that of the first two graphs, so is the landmark. Finally, the graph in the bottom right corner presents not only the rate of change but also a direction change, forming a prominent landmark. In haptic exploration, a line graph that carries this information is explored actively in a sequential pattern; and the changes are explored dynamically. The dichotomy in the use of "event" term for haptic graph comprehension (the segmentation in the abstract event and the segmentation of a graph shape) should be kept in mind. But on any ground, the second-order changes provide cues for segmentation in haptic modality as in the case in visual modality.

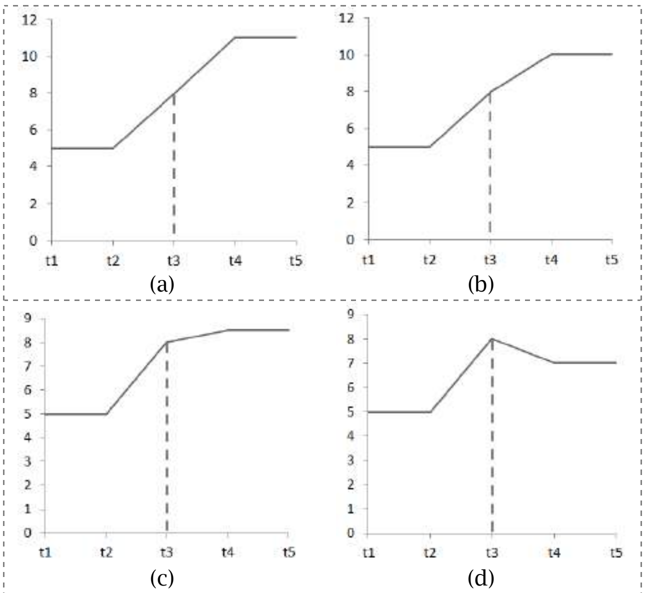


Figure 6-2 Illustrations of second-order changes in the graph domain

Constructing a partonomy for a graph line (as a static visual object or as dynamic haptic event-like exploration) is challenging because the boundaries of the parts are not well-defined. This may also lead to multiple reasonable spatial decompositions employing different landmarks as event boundaries and to the construction of partonomic relations in different temporal granularities.

Temporal Granularity (Coarse- and Fine-Grained Event Segmentation). It has been empirically demonstrated by Zacks and Tversky (2005) that people spontaneously segment events at different timescales in correspondence with the partonomic hierarchy.

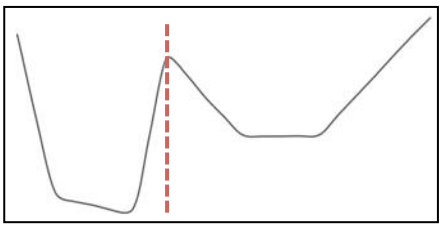
And, as mentioned earlier, the chosen event boundary plays a role in determining the temporal granularity. Temporal granularity in event segmentation is an issue that is particularly important due to the prominent role of temporal structures in the language modality. As proposed by Talmy (1983), they are a frequent and essential part of narratives as every sentence (verbal or written) contains this information implicitly or explicitly. Furthermore, it also helps organize memory retrieval.

People are good at establishing relations between events and sub-events on different timescales (Zacks and Tversky, 2012). Furthermore, they may use this information for perceiving, reasoning and communicating about, and remembering events. Another study conducted by Zacks et al., (2001) indicated that the human perceptual system actively uses partonomic hierarchies to encode ongoing activity. In this study, they presented observers with videotapes of four day-to-day, goal-directed, and ranging from very familiar to very unfamiliar, activities: making a bed, doing the dishes, fertilizing houseplants and assembling a saxophone. Each participant segmented all four activities with respect to two different timescales; segmentation into the natural and meaningful smallest units and into the largest units. Their results indicated that there is a hierarchical relationship such that large-unit boundaries were mostly in line with small-unit boundaries. They also found that this alignment was more prominent for the familiar events than for unfamiliar ones. Another relevant finding was that this hierarchical relation is more pronounced when observers were asked to produce an online description for an event compared to when they were just asked to segment it. The authors concluded that event segmentation is affected by both bottom-up perceptual information and top-down cognitive representation of events.

Apart from the investigation of the relation between different layers of temporal structure, the relation of each temporal structure with the represented event was also an important issue. The experimental research conducted by Zacks et al. (2009) indicated that fine-grained events are more perceptually determined and more strongly associated with actions on objects, whereas coarse-grained events are more conceptually determined and more associated with conceptual information about an event, such as action contexts (Zacks, 2004; Zacks et al., 2001) or goals and causes (Baldwin and Baird, 1999).

Table 6-1 illustrates two verbal descriptions for a graph on that are on different partonomic levels (these taken from the verbal descriptions produced for the experiment in the following part, in 6.4). The preferred event boundary chosen in the first sentence (1) separates the shape into two main parts. This is a good exemplification of a coarse-grained partonomic relation in the graph domain. On the other hand, one may also describe the same graph by dividing it into more segments based on actions, as illustrated with Sentence-2. As can be also seen from these examples, coarse-grained partonomic relations are suitable for object-like descriptions, while fine-grained descriptions mostly carry action information.

Table 6-1 Examples for different scales on the partonomic level

Coarse-Grained Scale	1. “There are two shapes look like water glasses, the first one is deeper.”	
⇕		
Fine-Grained Scale	2. “It decreases steeply, then starts to increase, then decrease again, goes stable and then it ends by increasing.”	

Conceptual Framing: Conceptual framing has also been shown to affect event comprehension and segmentation (Zacks et al., 2009; Massad, Michael and Newtson, 1979). A conceptual frame can be provided by giving an informative title or clues that highlight the context. It is essential since it activates semantic knowledge representations, called *event schemata*, which allow a reader or perceiver to integrate information about different aspects of the event. Reynolds, Zacks and Braver (2007) described event schemata as goal-directed hierarchical knowledge structures for event representation. Event schemata allow meaningful goal decompositions as well as inference about future or missing events. Furthermore, event schemata also affect how events are organized in regard to partonomic hierarchies. Thus, goal-directed event schemata are important for comprehending activities. It has been claimed that the temporal (partonomic) structure and goal information are brought together in event schemata (Zacks et al., 2009). The results of a study conducted by Reynolds et al. (2007) on computational models of event understanding also demonstrated that partonomical relations are crucial for event comprehension due to their prominent role on facilitating inference, summarization and question-answering. These are also the tasks where graphs are considered to be superior to the texts and tables.

In the graph domain, event schemata correspond to graph schemata (as introduced in Chapter-1) that provide graph readers with information on how to read a specific type of graph and the relations between the graphical entities and also on which features are important for graph comprehension or need to be ignored. Besides, graph schemata also highlight conceptual features that a graph represents, such as the fact that the closest point to the x-axis corresponds to a global minimum.

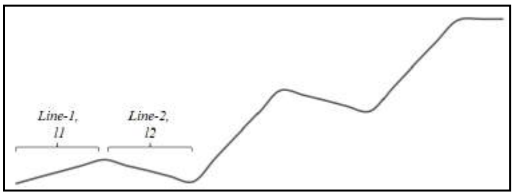
Consequently, providing a conceptual frame would be very useful to activate the appropriate graph schemata that facilitate comprehension. Event schemata relate conceptual knowledge with the partonomic structure of the physical shape. Furthermore, their role in the online processing of events and later memory are also substantial. Therefore, it is especially important for haptic graph comprehension since spatiotemporal reading can be affected negatively from the interference due to active exploration actions during right-to-left reading (this issue will be elaborated in Chapter 9.3).

Taxonomic Relations: In addition to partonomic relations, both objects and events can be organized based on "kind of" relationships, their so-called *taxonomic hierarchy*. One can use different taxonomic (*subordinate, basic and superordinate*) levels to describe an object or an event. As elaborated on in Zacks and Tversky's broad analysis on the temporal granularity of events (2012), people use the basic level as a default taxonomic level. Moreover, a generalized image for an object or a behavior routine for an event can be constructed easily from a basic level description. The empirical studies also showed that people are fast in naming events with basic level description (Rosch, 1978).

The descriptions on a subordinate level contain fine-grained taxonomic relations, basic level terms contain intermediate scale relations, and descriptions on the superordinate level contain coarse-grained relations. As exemplified in

Table 6-2, which presents samples of descriptions on different taxonomic levels, one can refer to the segment complex formed by the combination of *line-1* and *line-2* with the description "*there is a change*". This corresponds to the superordinate category. On the basic taxonomic level, more detailed information defining the change in the line can be presented, such as "*there is an increase and a decrease*". Finally, one can refer to the same line complex with more precise information by using the amount of change, as in "*there is a slight increase and decrease*".

Table 6-2 Examples for different scales in taxonomic level

Coarse-Grained Scale	1a. There is [a change]$_{l1+l2}$	
	1b. It [stays around low values]$_{l1+l2}$	
⇕	2. There is [an increase]$_{l1}$ and [a decrease]$_{l2}$	
Fine-Grained Scale	3. There is a slight [increase and decrease]$_{l1+l2}$	

Referring on different taxonomic layers is dependent on the concrete choice of a referent and whether there are alternative contrasting referents. The contrasting feature(s) that distinguish(es) the referent and distractors from each other form a level in a taxonomic hierarchy. Basic level descriptions exhibit more similarity in terms of physical features compared to descriptions produced on a superordinate or subordinate level. Similarly, basic level descriptions for events also exhibit more common motor movements regarding actions than the superordinate level does. And the subordinate level descriptions, in other words, fine-grained descriptions are strongly tight to actions (e.g. Zacks and Tversky, 2012; Morris and Murphy, 1990). Therefore, the choice of a referent has an influence on the choice of spatial, temporal and taxonomic levels. As suggested by Zacks and Tversky (2012), the information is processed based on these established levels of relations and the conceptual and perceptual representations are profoundly affected by the chosen taxonomic scale. In other words, as suggested by Zacks and Tversky (2012), referring to an object or an event at different taxonomic scales evokes different sets of contrasting objects or events. In general, the taxonomic organization promotes reasoning about intrinsic properties, whereas the partonomic organization promotes reasoning from the physical structure to the conceptual structure; the latter includes functions and causes. To sum up, all these partonomic and taxonomic relations are intertwined and have systematic relationships, since the choices made regarding the partonomic level affect the taxonomic level, and vice versa. Also, all these relations are affected by top-down and bottom-up features the referent and the reference space possess.

6.2.3. The Event Segmentation Theory

Reynolds, Zacks and Braver (2007) proposed an *Event Segmentation Theory* (*EST*) that aims to explain these relations systematically. The first claim of the EST (and other psychological accounts of event segmentation, see Newtson, 1976 and Zacks et al., 2007) is that segmentation occurs at locations where perceptually (such as changes in action patterns) or conceptually (such as changes in actor's intention) the highest changes occur. They also argue that these feature changes, especially the unpredicted ones, have a high probability of being chosen as an event boundary.

I should open a bracket here to refer back to the three layers in graph comprehension: a conceptual event, a static representation (a physical graph) and static or dynamic perception of the graph that represents the abstract conceptual event. It should be mentioned that a conceptual event can have unpredicted changes, but visual graphs present all changes simultaneously in a static way. As such, conceptual event segmentation during dynamic exploration of haptic graph is different than the segmentation in static-visual graph. Still, graph schemata tell that the regions that present a high amount of change in space correspond to events that have abrupt changes in its action stream. On the other hand, unlike the object-like perception of a conceptual event in the visual modality, haptic graph exploration yields an event-like perception of the static graph that represents that conceptual event. Therefore, graph schemata that allow this mapping (between the conceptual event and the physical shape) should be constructed and used for appropriate event segmentation.

The EST suggests a theoretical perspective on how the human cognitive system performs event segmentation (Zacks et al., 2007). According to this theory, segmentation controls and regulates the contents of short-term memory. In order to perceive the current situation, mental models of the situation are generated first and what the EST claims is that these mental models are formed at breakpoints. This model suggests that when an event boundary is formed, mental representations of the current situation are updated and actively maintained until the next boundary. Considering that the locations of event boundaries are chosen based on bottom-up perceptual features and top-down conceptual knowledge (*event schemata*), event models are assumed to be constructed by using currently available and relevant features (Glenberg, 1997; Johnson-Laird, 1989; Rumelhart, Smolensky, McClelland and Hinton, 1986; Zwaan and Radvansky, 1998). The EST suggests that memory for conceptual as well as perceptual information will be better for boundary objects than for nonboundary objects. Figure 6-3 shows a sketch of the model that was retrieved from Zacks (2004) and redrawn based on the original.

To sum-up, according to the EST, event boundaries are constructed at the convergence of major changes occurring in bottom-up perceptual information (motion-based) and in top-down conceptual information (knowledge structure). Event segmentation occurs as a result of holistic process of both type of information. Therefore choosing appropriate event boundaries by taking the partonomic and taxonomic relations into account with both perceptual and conceptual effects is crucial for event segmentation in regards to online comprehension of event and also retrieval of event information later. Another important detail in this model that should be highlighted is the association of the feature detectors with the range of temporal grains. As stated in Reynolds, Zacks and Braver (2007), coarse-grained feature detectors are more sensitive to knowledge structures, whereas fine temporal tuning is more associated with low-level sensory information.

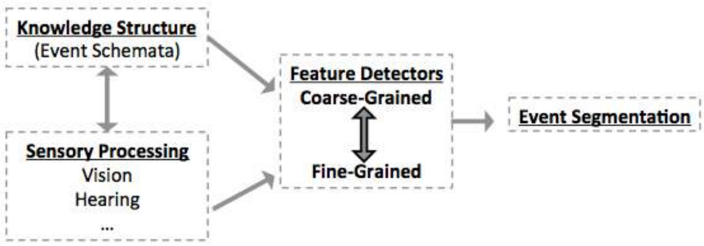


Figure 6-3 A model of the role of the movement features in event segmentation, (retrieved from Zacks, 2004, p.982 and redrawn)

6.2.4. Online and Offline Descriptions of Events

The results of Reynolds et al. (2007) support the view that people spontaneously construct partonomic event representations, and that these representations are important for producing descriptions of ongoing events. Their findings indicated that there were systematic differences between the coarse and the fine-grained descriptions. The coarse unit descriptions tended to mention objects (or object-like descriptions) more often than actions. Furthermore, they contained more semantically precise specifications of the objects than the fine unit descriptions did. On the other hand, the actions were more precisely mentioned in the fine unit descriptions than in the coarse ones. Also, while the coarse units were usually divided by the objects, the fine units were mostly divided by different actions. They interpret those to suggest that same partonomic structures that have an effect on the perceivers' segmentation of ongoing activity influence their descriptions as well. Additionally, as proposed in Zacks and Tversky (2012), changing the conceptual grain of description has an effect on both the granularity in the taxonomic details and the temporal grain of the events described (the partonomic relations). Additionally, it has been also suggested that people recognize visual details of the events better when they segment them at the fine-grained level instead of coarse grained (Hanson and Hirst, 1989)

It has also been suggested that the on-going segmentation affects memory encoding and off-line memory retrieval. Zacks et al. (2001) showed that the descriptions of an event from recent memory resemble the description of an ongoing activity. Their findings suggested that a partonomically structured representation that guides perception has also an influence on memory. Their experiment is particularly relevant for the issues related to the graph domain. In this experiment, the participants presented simple transcriptions of activities, and they were instructed to divide the list of transcriptions into groups based on the type of activity described (assembling a saxophone, fertilizing houseplants, washing dishes and making a bed). Their findings suggested that the participants were able to extract the partonomic structure of the original activity given the simple transcripts. They concluded that the hierarchical structure on the syntactic and semantic levels used by the producer of the transcripts were successfully transferred to the perceiver's conceptual representation of the described activity. The content of the verbal assistance for the haptic graphs may provide similar guidance too. Providing appropriately segmented assistance may further facilitate both online comprehension and memory recall for later use.

Another study presented by Swallow et al. (2011) provided evidence for the substantial role of event segmentation on memory retrieval and memory encoding. They proposed that memory updates occurring at the event boundaries may qualitatively constitute different context signals for consecutive events. These signals may be used to differentiate the events in the episodic memory (Polyn and Kahana, 2008). There are also several studies that demonstrated that physical characteristics of the event were recalled more from description with fine-grained segmentation (Hanson and Hirst, 1989, Lassiter and Slaw, 1991). All these results indicate that the principles governing online event segmentation have also an influence on the event segmentation process during verbal description of ongoing events and later memory as well.

6.3. Event Description

Event description is an integral part of the conceptualization of an event and it is also an issue critical for the design of the verbal assistance system. Therefore, in addition to the question of how graph readers segment events into smaller units, how they refer to those units (which is already introduced shortly in 6.2.2 under the title of taxonomic relations) was investigated. First, the different aspects of events and also the event related concepts should be examined. This part, which presents a philosophical discussion on what events are, is based on the well-combined collection of papers presented in Casati and Varzi (2008; 2002) and in Goldman (2007).

Goldman (2007) introduced various notions of events; these are the *common-sense*, *theoretically revised*, *scientific*, and *internal psychological* notions. According to him, the common sense-notion of an event is overdeterminate, and the overdetermination is explained by *two* underlying internal representations of events. First, there is a *perceptual* internal representation of events that is action-based (or change-based), therefore they exhibit mostly spatial and temporal features, yielding coarse-grained segmentation. Second there is a *conceptual internal* representation of events that is property based, yielding fine-grained segmentation.

Overdeterminate descriptions are very common when it comes to eventualities (Casati and Varzi, 2008). As discussed before in Section 2.3, overdeterminate expressions are acceptable for instructional settings (Koolen et. al, 2001, 2009). As has been exemplified in Casati and Varzi (2008), we can identify A with B, for example, if they belong to same category with showing little mostly gradual *physical feature* changes. In another context, we can identify A with C, if they have some kind of *temporal relation* again with some shared features. But this does not entail that B and C are the same. Together with the previous literature presented in this chapter, this would be the another exemplification of having two different labeling in sub-ordinate level for the entities that have same basic level label and the common-sense notion proposed by Goldman may corresponds to the basic level categories in psychological and computational domain. In the graphical communication domain, referring to two distinct event segments with a verbal description such as "it increases" does not mean that they are the same. They can differ in terms of their temporal aspects or spatiotemporal parameters (i.e. steepness). Figure 6-4 presents a graph with three main line segments that exhibit different slopes. Each of the three line segments, and even the line complex consisting of all segments, can be described as "it increases". However, such a description would not be discriminative enough to convey the small changes occurring in the line graph. The relation between the previous and the next entities in the graph shapes (i.e. contrasting or alternating features with respect to the adjacent and far neighbor segments) also affects how they are referred as.

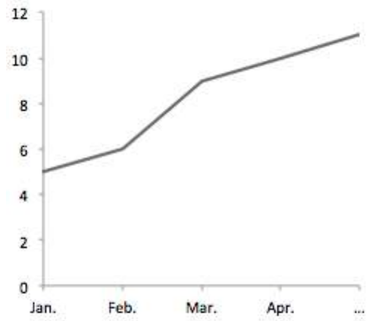


Figure 6-4 A graph sample with different values for the same basic action (i.e. increase)

The differences and similarities between objects and events have been discussed throughout this chapter. Here, first I will focus on a more detailed definition of events and event-related notions such as facts, properties and actions, which have been also focus of the graph comprehension research. These distinctions should be considered as different aspects instead of different distinct categories.

Event vs. Fact: As opposed to a more concrete distinction between objects and events, the distinction between events and facts is not that clear. Facts are distinguished from events with respect to the features of abstractness and temporality. Casati and Varzi (2002) provided a generic example for this: The event "Caesar's death took place in Rome in 44 B.C." is considered to be different from the fact that "Caesar died" (Ramsey, 1927). Moreover, the fact that Caesar died violently and the fact that he died can also be interpreted as two different events (Chisholm, 1970, 1971) as opposed to interpreting them as one and the same event described in two different ways (Davidson, 1969). From this point of view, the event of "a population reached a peak" is also different from the event "it has a peak" or "it reached a global maximum".

Event vs. Property: Events and properties are also contrasted with each other. Some philosophers consider an event to be a particularized property located at some region of space-time (Bennett, 1996). However, they are mostly regarded as two different kinds. Events are seen as individuals, whereas properties are seen as universals. According to Casati and Varzi's definition, individuals exist or occur, while universals recur. In the graph domain, three main changes can occur: A value represented on the y-axis can increase, decrease or remain stable with respect to time. These basic changes can be classified as events, the modifiers of those events can be categorized as a property. For example, in the description "it increases slightly", the adverb "slight" is a property, the verb "increase" is the main basic action and both of them constitute the event.

Event vs. Action: Another distinction that needs to be drawn is between events and actions. The basic actions are claimed to be atomic components of events. According to by Zacks and Tversky (2001), actions are performed intentionally by actors, thus they are less general than events and can even be regarded as sub-categories of events. In the graph domain, verbal descriptions such as “it increases” or “there is an increase” can be classified as events but only the former one is an action.

In addition to contrasting events with other kinds, an event can be inspected with respect to various internal nuances as well. The taxonomy of the events differs between four types of events: *activities*, *accomplishments*, *achievements*, and *states* (Ryle, 1949; Vendler, 1957). An activity such as “it increased”, can be considered as a homogeneous event. The trend may change from steep increase to slight increase but its sub-events still satisfy the main action (i.e. increase). An accomplishment, such as “it reached the global maximum” is not homogenous, and it may have a culmination. An achievement, such as “it is reaching the global maximum”, has a culmination and is instantaneous. And a state, such as “it has two peaks”, is homogeneous and may extend over time. Being *static* or *dynamic is* another aspect of events. According to some authors, static contents can be categorized as events since they do not involve any change (Ducasse, 1926). According to Casati and Varzi (2002), the distinction between static and dynamic events is similar to the difference between states and activities presented above. For example, a description like “it does not change for a while” can be classified as a static event because it shows change in the temporal but not in the spatial dimension. The description “it increases for a while” is an example of a dynamic event. It should be noted that static events are different from states. States generally do not exhibit a change in time and space, i.e. a description like “there is a peak” would be an example of this category.

These distinctions that need to be made when dealing with verbal descriptions can be also observed in speech-accompanying gestures. As introduced in Chapter 3.3, there are different categories of gestures in communication (such as *beat*, *interactive* and *representational* gestures). The representational gestures consist of more refined categories, such as deictic gestures that point out objects, places and people and iconic gestures that resemble the shape of an object or an action (Özçalıskan and Goldin-Meadow, 2005). Similarly, graphical communications also have similar “toolkits” that can carry these differences in the meaning by employing different graphical annotations such as points, lines and arrows (Acartürk, 2012). More specifically, graphical annotations such as arrows may highlight the conceptualization of processes and the punctual marks may highlight the conceptualization of states. Furthermore, the study conducted by Acartürk and Alaçam (2012) indicated that people produce more iconic and dynamic gestures while talking about entities highlighted with arrows in the graph, whereas they tend to produce mostly deictic and static gestures for the graphical entities marked with punctual annotations.

In this dissertation, I go along with the umbrella notion instead of refined concepts as discussed by Casati and Varzi, (2008) and by Goldman (2007). Here, I do not differentiate between these categories (i.e. action or state) although the verbal data was classified based on the distinctions outlined above (facts, properties, time, relations). To illustrate, all these following four points are called instances of an event that shows varying attributes:

- It decreases (basic action)
- There is a decrease (basic state)
- It decreases fast (basic action + modifier)
- It decreases and it is fast (basic action + a statement with modifiers)

On the other hand, this distinction was taken into account when handling graphical entities and speech accompanying gestures. Each graph shape was annotated with respect to its type; *shape landmark* or *shape segment*. Besides, the gestures were coded as non-directional (*static*) and directional (*dynamic*).

An investigation of the partonomic and the taxonomic relations in order to understand event segmentation requires thorough and systematic analyses of both the qualitative and the quantitative attributes of the events. Zacks et al. (2001) provides a methodology that addresses both aspects. The qualitative features of sub-events can be investigated through a linguistic analysis of general patterns in the verbal descriptions. As for the quantitative features of the sub-events, they can be investigated through a statistical analysis of the perceptual and the conceptual properties. Their findings through these analyses suggested that the semantic and the syntactic structure of the verbal descriptions depend on the partonomic and the taxonomic structure of the event representation. According to Talmy's (1975), a discourse about an event is governed by the structured representations of that event. This structure is consisting of an object (which is a nominal), an action (which is a verb), a path (which is a prepositional) and a ground (which is a nominal).

In the previously presented study (Chapter 5), I focused on the effect of the sensory modality on the referring expression production (REF). In another study reported in Chapter 9, the use of REs within dialogues in the context of collaborative haptic graph exploration activities is examined. Briefly worded, the former study reveal that certain aspects of graph segments are over- or underestimated in the haptic modality. Although the experiment was not designed to investigate event segmentation, it still provide intriguing examples in regards to this issue. To illustrate, one of the graphs in the stimuli set (Graph-II in Figure 4-6) showed an increasing trend with several fluctuations. The majority of visual graph readers (80% for the graph with labels, and 100% for the one without labels) produced REs that referred to the general pattern without applying segmentation in their post-exploration verbal descriptions. On the other hand, about 45% of the haptic explorers referred to the event as a whole. For another graph with a similar increasing pattern but with higher peaks rather than small fluctuations, a similar pattern in visual and haptic modalities, namely referring to the whole event, was observed (70% of the users in the "visual graphs with labels" condition, 64% of the users in the "visual graphs without labels" condition and 75% of the users in the haptic condition). The latter study indicated that the verbal assistants' production of graph-domain-specific REs (containing size, shape or relation) yielded more successful communications than the production of basic level REFs (i.e. type or direction). These results highlighted the significant role of specific graphical features in event segmentation and the role of event segmentation in the production of referring expressions. As previously proposed, event segmentation is affected by both knowledge structure and perceptual features. However, the saliency of the perceptual features and the accessibility of the conceptual features can be dependent on the perceptual modality's affordances. Although the same underlying principles, which state that event boundary occurs at the location that exhibit most conceptual or perceptual change, works for both modalities. Still the perceptual differences may cause differentiations in event segmentation as well. Based on prior research, this empirical investigation focuses on the effect of amodal geometric properties and conceptual features on event segmentation by conducting comparative analyses of REF production, speech-accompanying gestures and sketches across visual and haptic modalities.

6.4. Empirical Investigation For Event Segmentation And Event Description

In this empirical study, I focus on two main questions; (i) how the graph readers parse continuous stream of data into meaningful discrete events by using perceptual salience or conceptual significance, namely "event segmentation" and (ii) how they refer to segmented events, namely "event description". For this purpose, the data collected through Experiment-III was analyzed (see Chapter 4.4.3 for detailed information about the experiment).

6.4.1. Participants

Forty-eight participants (*Mage*=24.69, *SD*=5.85, 23 female) from Middle East Technical University participated to this experiment. 3 of the haptic explorers were left-handed and all subjects used the device with their dominant hand. Haptic explorers were blind-folded sighted people. The experiment was conducted in three conditions in a between-subject design. The experiment was conducted in Turkish, the native language of all participants.

6.4.2. Experimental Design

Following the same experimental design presented in the previous experiment, the experiment was conducted in three conditions in a *between-subject design* (16 participants for each condition). In the first condition, the participants explored line graphs haptically (without data labels). In the second condition, the graphs with data labels were presented on a computer screen, thus the participants had visual access to the graphs. In the third condition, the participants inspected the visual graphs without data labels. This condition served as a control condition for both former conditions. The effect of the sensory modality on event segmentation was explored by comparing the condition-1 and the condition-3. The data that belongs to these two groups are called *Modality-Group*. The effect of data labels on the event segmentation was investigated by comparing the condition-2 and the condition-3 and the data belongs to these two groups were named as *Label-Group*. In all conditions, after the graph exploration (visually or haptically) process has been completed, the participants were asked to present a single-sentence summary of the graph. Each graph in the both visual conditions was shown for 10 seconds on a computer screen, whereas haptic users did not have time limitation for their exploration. In the given instruction, the changes in the data were attributed to the seasonal climate changes or having a special event (like festivals etc.), so that the participants were not invited to make causal reasoning.

In the experiment session, each participant was presented twelve haptic or visual line graphs with smooth edges that present averaged monthly tourist visits for various cities in Turkey. The graphs consisted of systematically controlled graph entities in terms of steepness and angle. As illustrated in Figure 6-5, the steepness of the lines (*within-subject variable*) with respect to x-axis were restricted to four values to ensure a systematic analysis. Besides, the angle between the line and the x-axis were set to 0˚ (no change), 15˚ (slight change), 45˚ (moderate change) or 75˚(steep change) for both direction (such as an increase or an decrease).

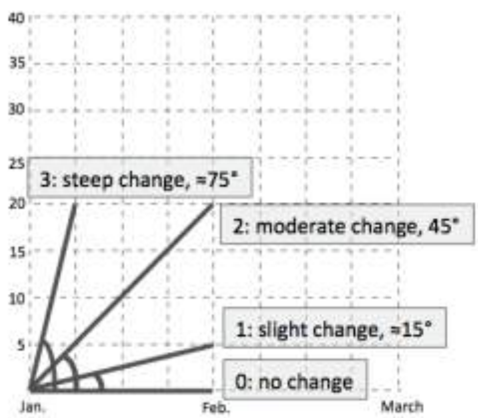


Figure 6-5 The illustration of slope values

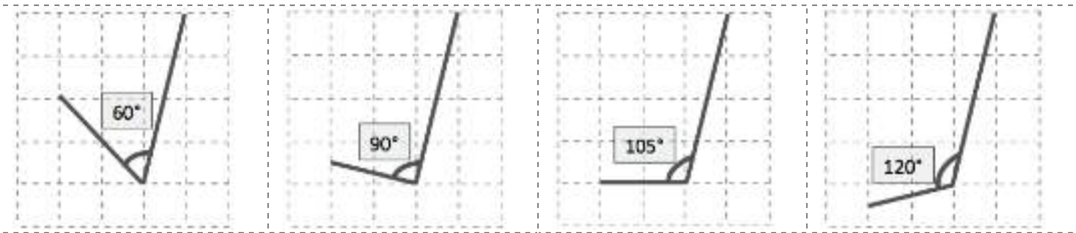
When designing visual graphs, several factors were kept constant for all graphs for the sake of a systematic analysis, these are as follows;

1. Y-axis labels (that correspond to the number of tourist visits)
2. X-axis labels (that correspond to the months),
3. the ratio between the width of the x-axis to the length of the y-axis

Due to this restriction, which is necessary for establishing systemically controlled environment, the four ordered-combination of line segments were not allowed in the data set, these are

1. a steep increase and a medium increase
2. a medium increase and a steep increase
3. a steep decrease and a medium decrease
4. a medium decrease and a steep decrease

In addition to the parameter of *steepness* of shape segments ("*slope-value*"), the turning angle at the intersection of two line segments, which have different steepness values from each other was also considered as one of the factors that had an effect on the perceptual saliency of the landmarks. However, it should be noted that the steepness of the line segments and the angle between them are closely tight to each other. In other words, as previously introduced in Chapter 1, whereas the slope corresponds to the second-order property, the angle corresponds to the third-order property. For example, as illustrated in Figure 6-6, the angle of a landmark that precedes a steep segment may take four different angle-values depending on the slope-value of the previous line segment. Here in this figure, only convex angles are presented, but same angles could also occur as concave.



(a) Medium decrease & steep increase | (b) Slight decrease & steep increase | (c) Horizontal segment & steep increase | (d) Slight increase & steep increase

Figure 6-6 The illustration of angle values

Figure 6-7 illustrates the angle values that can be formed from each possible combination of two line segments with varying slope-values. The *angle-values* were also categorized into four categories ranging from the most acute angles to the most obtuse angles. Figure 6-7a illustrates all possible angle values that can be formed at the intersection of two line segments with varying slopes. The slope-values are represented as 0 (*horizontal*), 1(*slight*), 2(*medium*) and 3(*steep*). Besides, the "minus" sign indicates that the segment has a decreasing trend. As depicted in Figure 6-7a, a landmark with at least one steep segment connection may have a value from three categories; A1, A2 and A3. On the other hand, a landmark with a horizontal or a slight segment connection may have a value from one of the A2, A3 and A4-value categories. Due to this dependency, additional statistical analyses regarding the angle values were conducted separately within each *slope-value,* besides the analysis on the main effect of angle.

	0	**1**	**-1**	**2**	**-2**	**3**	**-3**
0	0	-165	165	-135	135	-105	105
1	165	0	150	-150	120	-120	90
-1	-165	-150	0	-120	150	-90	120
2	135	150	120	0	90	NA	60
-2	-135	-120	-150	-90	0	-60	NA
3	105	120	90	NA	60	NA	30
-3	-105	-90	-120	-60	NA	-30	NA

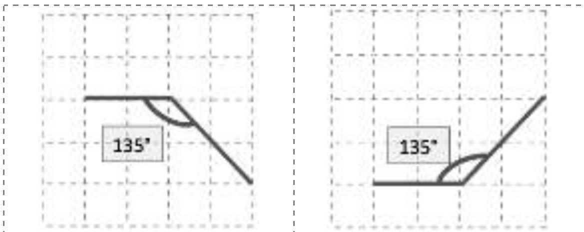
(a)

Angle-Value Categories	
Acute angles	**A1**: 30° & 60°
⇕	**A2**: 90° & 105°
	A3: 120° & 135°
Obtuse angles	**A4**: 150° & 165°

(b)

Figure 6-7 (a) Possible angle values formed at the intersection of two line segments with varying slopes, (b) Classification of *angle values*

The existing literature on shape segmentation already showed that the angle's polarity is an important factor for segmentation (Cohen and Singh, 2007). Therefore, regardless of the angle-value, the effect of polarity (*concave* versus *convex*) was also taken into account. The salient curvature landmarks of *positive maxima* and *negative minima* (Cohen and Singh, 2007) are related concepts of *convexity* and *concavity* on contours in closed curves. However, graph lines (of statistical graphs) are based on functions. Therefore, a graph line is neither closed, nor can it cross with itself, nor it can branch. Based on this, the convexity and concavity were elaborated under the umbrella notion of polarity, which is determined w.r.t horizontal axis, see Figure 6-8.



a) horizontal change & slight decrease

(b) horizontal change & slight increase

Figure 6-8 An example for different polarities (namely concave (left) and convex (right) angles)

As a result of different combinations of slope and angle values, twelve graphs, which are employed in this experiment, differ from each other in terms of their global shape. Besides, the number of the segments that the graph possesses may also have important influence on event segmentation. As a result of the possible combinations of angle and slope, the graphs in the stimuli-set can be categorized into three with respect to their *graph-segment count*; 5-segment, 6-segment and 7–segment graphs. Figure 6-9 depicts an illustration for a 5-segment graph.

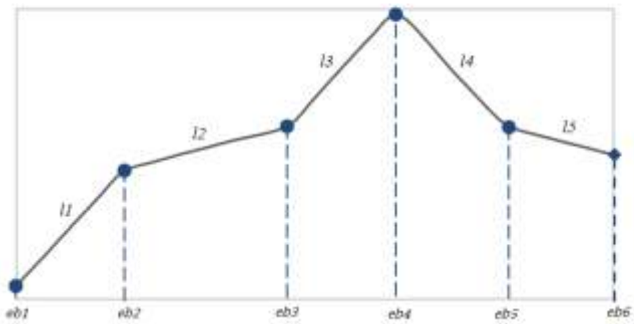


Figure 6-9 An illustration for a graph-segment Count

6.4.3. Segmentation And Annotation Procedure

Splitting verbal descriptions into phrases. A phrase (or a segment) was defined as any unit containing a predicate (i.e. a verb) that expresses a single sub-event or state by following Berman and Slobin (1994). Based on this definition, all post-exploration verbal descriptions were split into phrases (or sub-events). The investigation of event segmentation and description require refined linguistic analysis, hence both time and event denoting expressions were handled individually without breaking their relation. The annotation of time and event denoting expressions were conducted with respect to the annotation scheme introduced in *4.5.2*. Briefly, the annotation method for spatiotemporal information in this experiment was inspired by Schilder and Habel's approach (2001) that separates time-denoting and event-denoting expressions and also by the ISO-TimeML framework proposed by Pustejovsky and his colloquies. ISO-TimeML provides a method (a tagging system) to relate events to time. Both approaches are based on the Allen-like interval relations.

Table 6-3 gives descriptive statistics for the total number of the valid[22] protocols (the number of the participants x the number of graphs), the total number of the sub-events, and the number of time- and event denoting attributes.

Table 6-3 Descriptive Statistics

	Protocol	**Sub-Event Count**	**Time-denoting attributes**	**Event-denoting attributes**	**Total amount of Attributes**
5A	176	668	589	1537	2126
5B	185	823	606	2085	2691
5C	179	814	626	1964	2590
Total	540	2305	1821	5586	7407

The time-denoting expressions convey temporal information that comprises reference to a calendar system, such as on May, in summer, after several months etc. In this study, the scope was limited to these listed *basic* categories leaving the analysis of more refined level time-denoting attributes (i.e. the use of temporal prepositions) for further research. The list of attributes used in this analysis is as follows;

Basic Time-denoting Expressions

- Explicit Reference to Months (i.e. on May, or between June and August)
- Indexical Reference to Months (i.e. in third month)
- Seasons
- General (i.e. in general)
- Location (i.e. after the highest point)
- No temporal expression
- Vague Expressions (i.e. in a short time, for a while etc.)

All types of events have inherent temporal aspects, the *event-denoting* expressions also convey this information, but in an implicit way. On the contrary, information about the event, which is being referred to, is explicit in the content of event-denoting expression (such as an increase, a peak etc.). The verbal descriptions were annotated by using the <attribute, value> set approach (Dale & Reiter, 1995) to characterize the qualitative representations of the graph entities; the basic attributes in the semantic attribute scheme for the communication through graphs are *type* (state or action), *direction, shape*, *value*, *size*, *manner* and *relations* (see Chapter *4.5.2*).

Table 5-4 summarizes the list of dependent variables investigated in the upcoming analyses.

[22] * Some protocols were excluded from the analysis due to three main reasons;

1. If technical problems occurred during recording (regarding audio-video recording, eye-tracking or haptic device)
2. If the audio or video is not analyzable (i.e. very quiet speaking, or gesturing outside of the camera view)
3. If participant did not follow the experimental instructions or if they decide to skip "verbal description" step (for example, if they think that they have difficulty recalling the graph)

Table 6-4 The list of dependent variables

Verbal descriptions
• Expressivity Evaluation (for post-exploration descriptions) o Expressivity Scores o Matching Rates • Number of sub-events • Positioning rates (the location of event boundaries) • Reference rates • The choice of reference scope[23] • Number of attributes (in total and for each attribute individually) • Reference rate of event boundaries • Reference rates for shape segments • Landmarks' Type w.r.t qualitative properties • Error counts in Modifiers
Gesture
• The number of representational gestures • Gesture- Sub-event Mapping (Relation) • Static versus Dynamic Gestures o Directionality in gestures (1-directional or multi-directional) • Gesture-Negated Content Relation
Sketches
• Similarity scores for post-exploration sketches

RESULTS

In the following, firstly, short descriptive statistics for haptic exploration times for each graph are presented (*6.4.4*). The rest of this part presents the results of the statistical analyses in three main categories. The first part (*6.4.5*) contains the investigation of the expressivity evaluation of the outputs produced by the participants. Then, the findings that address how the participants segmented the graphs into sub-events and the effect of the amodal properties on event segmentation (as well as the data labels and the sensory modality) are reported in *6.4.6*. The last part of this result section involves the comprehensive analyses of how the participants refer to those sub-events in their verbal descriptions with particular focus on the analysis of speech accompanying gestures (*6.4.7*).

6.4.4. Exploration Time For Haptic Graphs

No time limitation for graph exploration was imposed to the haptic explorers due to the reasons explained beforehand. Figure 6-10 illustrates the average exploration times for each graph. The mean exploration time for all graphs was 81 sec. Although I did not apply a direct method to measure graph complexity, which is affected by many factors (in a range from perceptual to conceptual), the exploration time might be considered as one of them. The data of the mean exploration time for all graphs was split at the median (79 sec.) to form high and low groups. The group of the graphs which was explored longer than the median, namely the *long exploration* time group, contain the graphs 4, 5, 7, 10, 11 and 12, the group of the graphs which was explored shorter than the median, namely, the *short exploration* time group is consisting of the graphs 1, 2,3, 6, 8 and 9, (revisit Figure 4-7 for the visuals of the graphs).

[23] I used the term "the reference scope" to refer to the coverage of the referring expression w.r.t graph entities

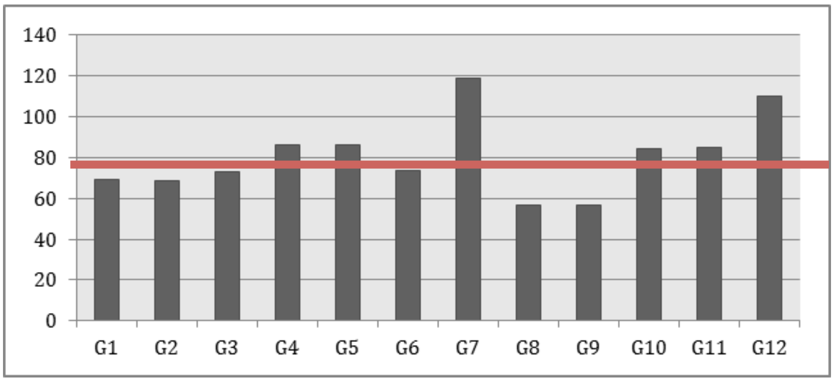


Figure 6-10 The average exploration time for the haptic graphs

This split highlights that having a steep segment(s) appears to have an effect on the exploration time. The presence of such a relation was tested by performing correlation analysis. The result showed that the exploration time is positively correlated with the presence of a steep segment[24] (r=.71, p (one-tailed) <.01), but no significant correlation was found between the exploration time and the other slope-values. Since the steep segments are longer than the other segments, it can be expected that the longer the graph path is the longer the exploration takes. The analysis confirmed that the exploration time is positively correlated with the total graph length as well (r=.71, p (one-tailed)=.01), see Table 6-5. Furthermore, the effect of different turning angles (of the landmarks) on the exploration time was also investigated. No correlation between the exploration and any angle-values was found. The slope-value seems to be more discriminative than the angle-value.

Table 6-5 The path lengths and the average exploration times for each graph (▲: the highest value in the column, ▼: the lowest value in the column)

Graph No	Path Length	Exploration Time
Graph 1	9.8	69
Graph 2	8.7	69
Graph 3	8.1	73
Graph 4	12.4	86
Graph 5	10.4	86
Graph 6	10.3	74
Graph 7	12.4	119▲
Graph 8	8.2	57▼
Graph 9	9.8	57▼
Graph 10	7.9 ▼	84
Graph 11	12.7	85
Graph 12	13.6▲	110

6.4.5. Expressivity Evaluation

6.4.5.1. Post-Exploration Verbal Descriptions

This analysis investigates the expressivity of the verbal descriptions produced by the participants in all three conditions. Two coders were asked to read the verbal descriptions one by one and to match each description with one of the twelve graphs. Correctly identified descriptions were counted as *match* in the calculation of the

[24] One can also argue that the presence of containing a steep segment increases the graph complexity. However, it is hard to pinpoint such main effects given 12-data point graph. In order to reach more conclusive results, further study with fewer line complexes (i.e. with 2-3 line segments) needs to be conducted

"*matching rate*"[25]. Then the coders rated verbal descriptions in terms of the expressivity by answering the question of "how well this description describes the graph in 5-point Likert scale". The score was called as "*expressivity score*", see Table 6-6 for the descriptive statistics regarding both parameters.

Table 6-6 The descriptive statistics for the correctly matched graphs and their expressivity ratings (▲: the highest value in the row, ▼: the lowest value in the row)

	Correctly Matched Graph (%)			**Expressivity Scores (1to5 Likert Scale)**		
Graphs	VG with labels	VG without labels	HG without labels	VG with labels	VG without labels	HG without labels
1	93.33▲	73.33	66.67▼	4.00▲	3.93	2.58▼
2	92.86	100▲	100▲	4.57▲	4.36	3.93▼
3	92.31▼	100▲	93.75	4.15	4.47▲	3.25▼
4	64.29	73.33	53.33▼	4.00	4.20▲	2.60▼
5	92.31▲	66.67▼	73.33	3.54	3.80▲	3.13▼
6	92.86▲	81.25	75▼	4.43▲	3.81	3.69▼
7	87.5▲	75	56.25▼	3.94▲	3.13	3.00▼
8	75	81.25▲	53.33▼	3.75	4.31▲	3.47▼
9	100▲	87.5	46.67▼	4.50▲	4.19	3.00▼
10	87.5▲	66.67▼	86.67	4.19▲	3.07▼	3.47
11	93.75	75	100▲	4.06▲	3.75	3.47▼
12	93.75	73.33	50▼	3.81▲	3.40	3.00▼

The expressivity scores lower than 4 underlines the uncertainty of the rater. Hence these descriptions were excluded in the interrater reliability analysis. The interrater reliability was calculated by Cohen's kappa. The results revealed a value of .74 that indicates substantial interrater agreement in the coder`s evaluations.

In the following, the analyses conducted within the Label-Group (the visual graph with/without labels) and within the Modality-Group were reported separately.

A. Effect of Data Labels

Several statistical analyses were conducted on the *expressivity scores* and the *matching rates* in order to explore whether there is a difference between the two visual graph conditions (with/without labels) in regard to *the global shape* of the graph and the *graph-segment count.*

Expressivity score. No significant difference was found between the visual graphs with labels and without labels, $p>.05$. There was also no significant difference in the expressivity scores of the verbal descriptions for different global shapes, $p>.05$. The graph-segment count also did not show any effect on the expressivity scores, $F(1.68, 50.33)= 2.39$, $p>.05$. This indicates that the descriptions for the visual graphs regardless of their shape properties were rated similar in terms expressivity.

Matching Rate: Likewise, no significant difference in the matching rates between the two conditions was found, $U=91.00$, $z=-1.42$, $p>.05$. Furthermore, there was no significant association between the matching rate and the global shape, $\chi_2(11)=15.47$, $p>.05$. The segment count did not exhibit any effect on the matching rates as well, $F(2,60)=2.45$,

[25] It should be noted that the raters were asked to match the verbal description among 12 graphs; therefore the distinctiveness of the graphical features that the graphs contain has also effect on the matching scores.

$p > .05$. Since there is no significant main effect, the pairwise comparisons were not analyzed.

B. Effect of Sensory Modality

Expressivity score. There was a significant main effect of the sensory modality on the verbal descriptions' expressivity scores, $F(1,21)=8.11$, $p<.05$, $\eta^2=.28$. The verbal descriptions for the graphs perceived via visual modality ($M=3.99$, $SE= .20$) were rated higher compared to that for the haptically perceived graphs ($M=3.11$, $SE=.23$). Additionally, the global shape also exhibited significant main effect on the expressivity scores, $F(11,231)=1.88$, $p<.05$, $\eta^2=.08$ with very small effect size. A significant main effect of the graph-segment count was observed as well, $F(1.56,46.93)=3.77$, $p<.05$, $\eta^2=.11$ (Greenhouse Geisser correction was applied since Mauchly's test of sphericity was significant). Follow-up analysis indicated that there was no significant difference in the expressivity scores between the 5-segment ($M=3.63$, $SD=1.01$) and the 6-segment graphs ($M=3.31$, $SD=.93$), whereas the scores for the 7-segment graphs ($M=3.18$, $SD=1.08$) were significantly lower than that for the 6-segment graphs, $F(1,30)=6.35$, $p<.05$, $\eta^2=.18$.

Matching Rate. The analysis conducted for comparing matching rates revealed that there was no significant difference between the two modalities, $U=94.50$, $z=-1.27$, $p>.05$. On the other hand, the difference among the global shapes was significant (χ_2 $(11)=25.70$, $p<.01$. The effect of the graph-segment count was also tested and the findings showed that the number of segments, which the graph contain, had a main effect on the matching rates, $F(2,60)=4.82$, $p<.05$, $\eta^2=14$. There was no significant difference on the matching rates between the 5-segment graphs ($M=.77$, $SD=.22$) and the 6-segment graphs ($M=.80$, $SD=.24$), $F(1,30)=.22$, $p>.05$. The number of correctly identified graphs from the verbal descriptions of the 7-segment graphs ($M=.64$, $SD=.32$) was lower than that of the 6-segment graphs $F(1,30)=7.96$, $p<.01$, $\eta^2=.21$ and it was also lower than that of the 5-segment graphs $F(1,30)=5.24$, $p<.05$, $\eta^2=.15$.

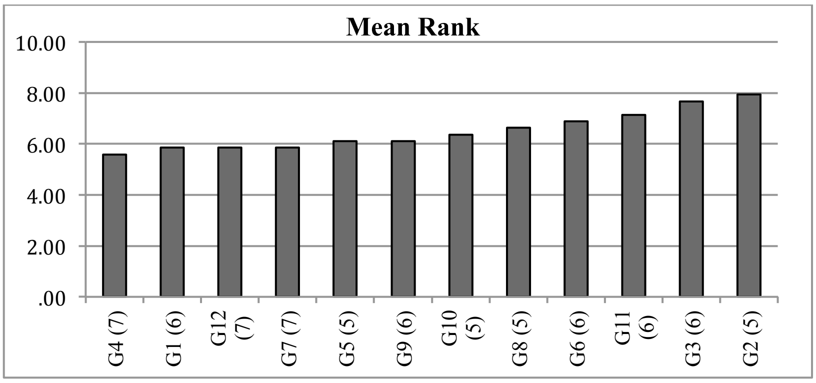


Figure 6-11 The matching rates for each graph (the numbers in parenthesis corresponds to the number of segments that the graph consist of)

6.4.5.2.Post Exploration Sketches

In this part, I focused on the similarity of participants' sketches to the stimulus-graphs as a performance measure. For the analysis of the sketches, the inter-rater reliability between the three raters was assessed using a two-way mixed, consistency average-measures ICC (Intra-class correlation). The resulting ICC was in the "excellent" range, identified by $ICC=.83$ ($N=551$) (Cicchetti, 1994).

The result of mixed design ANOVAs indicated that the sketches for the graphs with labels ($M=3.14$, $SD=0.95$) did not differ from the sketches of the visual graphs without labels ($M=3.46$, $SD=1.07$). However, the sensory modality had a significant effect (as expected) $F(1,24)=35.034$, $p<.001$, $\eta^2=.59$ and the score for the haptic graphs ($M=2.21$, SD=0.92) were significantly lower than that for the visual graphs (without labels). The global shape also showed weak significant effect F(1,24)=2.16, p<.05, see Figure 6-12 for the mean values of similarity scores for each graph.

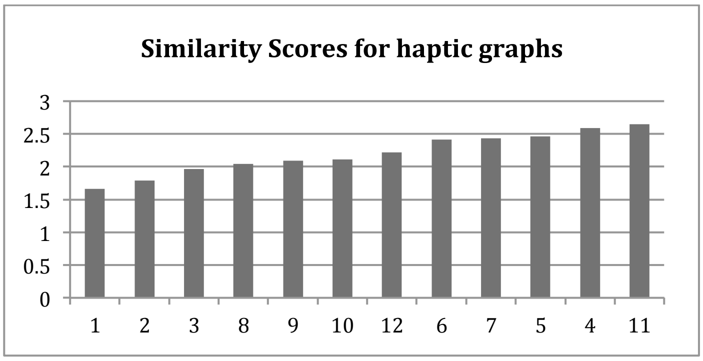


Figure 6-12 The similarity scores for the haptic graphs

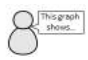
6.4.5.3.Interim Summary for The Expressivity

The findings of the analyses on the expressivity of the verbal descriptions indicated that regardless of having data labels or having different graphical features, the visual graphs had similar findings. As another indicator of how much resemblance there are between the graph and its conceptualization, sketches' similarity scores also exhibited the same pattern without showing difference between the two visual modalities. On the other hand, the difference between the visual and haptic graphs was significant indicating that the descriptions for the haptically perceived graphs were not clear as the visual counterparts were. Additionally, the graphical properties such as the global shape or the segment-count had an effect on the expressivity of the descriptions for the haptic graphs. Same pattern was also observed for the sketches' similarity scores. These results, first, tell us that the sketches and the verbal descriptions (as indicators of the graph conceptualization) produced for the haptically explored graphs in general had low scores and they are not expressive enough, pointing out that a conceptualization of the event provided through haptic modality might not be intact. Second, the verbal descriptions produced after exploring visual graphs can provide us valuable information, and the structure and the content of the verbal descriptions with high scores could be employed in deciding on the content given by verbal assistance system.

6.4.6. Event Segmentation

This section addresses the main topic of how graph readers parse continuous stream of data into meaningful discrete events by using perceptual salience or conceptual significance, namely "*the event segmentation*". In this part, the findings obtained from two experimental methods are reported. A series of analyses regarding the investigation of the post-exploration verbal descriptions is presented in 6.4.6.1. Afterwards, the segmentation of events on the level of speech-accompanying gestures is elaborated on in 6.4.6.2.

6.4.6.1.Event Segmentation In Verbal Descriptions



After the verbal descriptions produced by each participant were split into phrases (in other words, into sub-events) and annotated with temporal and event denoting attributes, two coders marked the corresponding event boundaries on the graphs (w.r.t the shape landmarks). This marking was utilized to anchor the verbal descriptions (*phrases*) to the corresponding graph segments. Figure 6-13 gives an example of this marking procedure; the utterance "it first increases, then goes stable"[26] refers to two shape segments on the

[26] The verbal descriptions provided in the examples were translated from Turkish.

graph. The time-denoting marker "*then*" separates the event into two sub-events, and that location is marked as an event boundary.

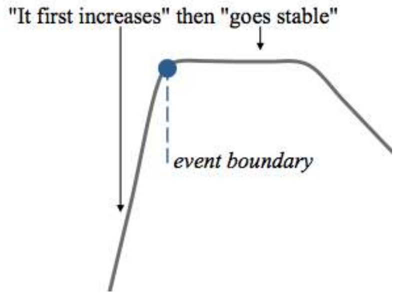


Figure 6-13 Illustration of event segmentation and event boundaries

Landmarks such as local and global points are natural event boundaries, however previous research already showed that during the communication over line graphs, graph readers tend to produce trend descriptions instead of focusing on the individual landmarks. The results of this current experiment, which is reported in the upcoming parts, also supported this conclusion. Therefore in the process of marking event boundaries, not only the shape landmarks but also the shape segments were taken into consideration. However, unlike the shape landmarks, the shape segments refer to dynamic events, which are not instantaneous. The participants do not necessarily convey explicit markers for the start and end points of each sub-event in their post-exploration verbal descriptions. Thus the verbal descriptions may not be clear about the location of the segmentation. To exemplify, the same verbal description presented in Figure 6-13 may be uttered to describe another combination of line segments. In that case, a graph comprehender may refer to just *line-1* or the combination of *line-1 (l1) and line-2 (l2)* by uttering "it increases" (see Figure 6-14a). Similar referring strategy can be also used for the utterance "goes stable" that may refer to just *line-3* or the combination of *line-2 (l2) and line-3 (l3)* (see Figure 6-14b). These kinds of regions are indicators of the expressions that have less expressiveness and the locations where the participants show non-uniformity. For such cases, the analysis of speech-accompanying gestures seem to be very effective tool in order to clarify the situation and to identify the end points (only if the participant produced gesture during his/her verbal description). See Figure 6-15 for the two examples of the gesture's contribution in understanding the ambiguities for the same sentence given previously.

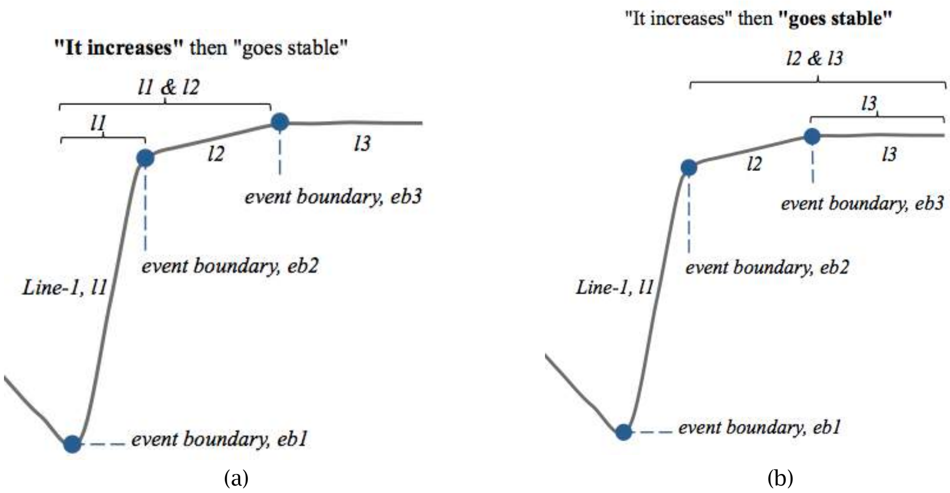
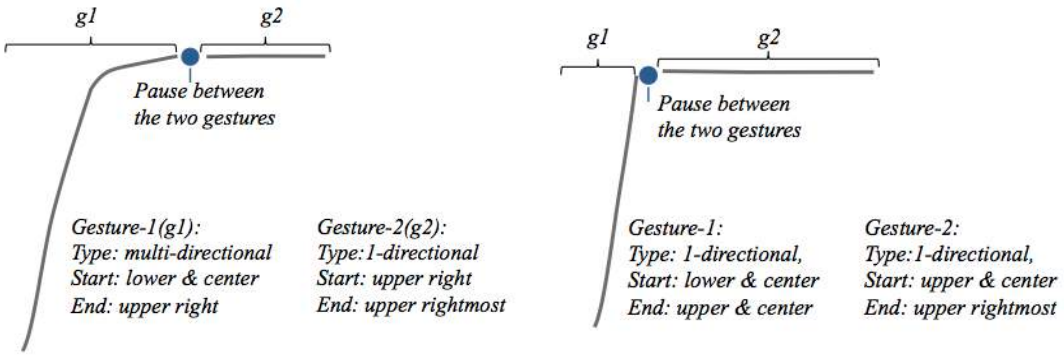


Figure 6-14 The illustration of the challenges in event segmentation and in defining event boundaries



(a) (b)

Figure 6-15 Gesture samples for the verbal descriptions presented in Figure 5

Furthermore, Figure 6-16 illustrates another example of the segmentation process that depicts how the descriptions were segmented into sub-events. Based on this description, *eb4* is the last mentioned event boundary. It, first, was mentioned as a shape landmark (*in the sub-event 5*) and secondly, it was implicitly referred as the start point of the shape segment (*in the sub-event 6*). In that case, the last utterance (*the sub-event 6*) specifies the end point, which does not match with the end point of the graph. Therefore the last points, which are referred by the participants, were also marked separately. For the identification of such regions, the post-exploration sketches also provides valuable information, since they help to understand whether the participant had incomplete conceptualization or just preferred to skip that piece of information in verbal description. Detection of these kinds of problematic regions and the analysis of the descriptions for these regions w.r.t their expressivity is highly crucial in order to provide distinguishing referring expressions by an effective assistance system.

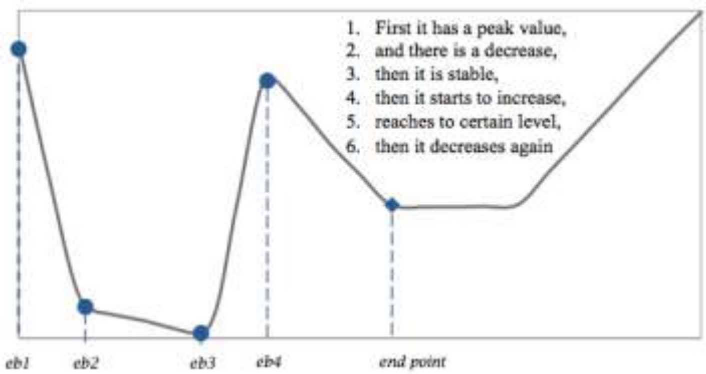
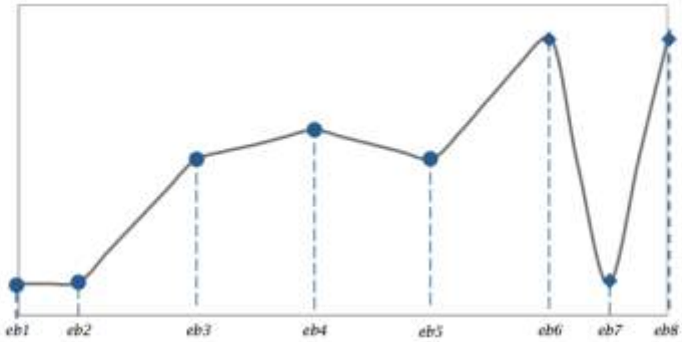


Figure 6-16 The "End of Global Event" Markers

Due to this fuzziness concerning the end points, the event boundaries were marked with respect to the references to the shape landmarks or to the start points of the shape segments.

Another parameter, which is used in this investigation, is *the reference rate*. The reference rate for each event boundaries (*the percentage of the participants who referred to*) was calculated for each condition and for each graph. Figure 6-17 presents a sample illustration of the corresponding markers on the graphs and the reference rates (as percentages). "*" sign shows that the corresponding event segment exhibits a significant difference (calculated by Pearson Chi-Square test) across conditions. The "end of global events" were also presented in the separate table in Figure 6-17. See Appendix-E for all statistics. The upcoming sub-sections contain detailed comparative statistical analysis. For a listwise summary, the readers are encouraged to take a glance at summary table (Table 6-28) that contain only significance values and effect sizes.



Effect of Data Labels							
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5****	***eb6***	***eb7***
VG- with labels	62.5	82.3	12.5	12.5	31.3	93.8	75
VG- without labels	86.7	73.3	26.7	26.7	80	80	60

Effect of Modality							
Event Start	***eb1****	***eb2***	***eb3***	***eb4****	***eb5***	***eb6***	***eb7***
VG- without labels	86.7	73.3	26.7	26.7	80	80	60
HG -without labels	43.8	81.3	12.5	62.5	56.3	87.5	75

End of Global Event	***eb6***	***eb7***	***eb8***
VG- with labels		25	75
VG- without labels		40	60

End of Global Event	***eb6***	***eb7***	***eb8***
VG- without labels		40	60
HG -without labels	6.3	6.3	87.5

Figure 6-17 A sample illustration and statistics for the reference rate of event boundaries (Graph-VII)

6.4.6.1.1. The Effect Of Data Labels

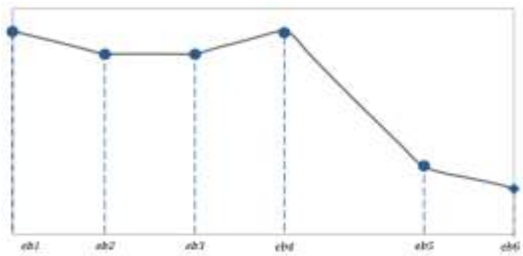
A. Number of the segmented sub-events

For each graph in the visual modality (with/without labels), how many segments the participants construct in their verbal descriptions (*the number of the segmented sub-events*) was calculated. First of all, the average number of the sub-events for all twelve graphs across two conditions was compared by conducting t-test. The results showed that there was no significant difference between the visual graphs with labels (M=3.89, SD=.90) and the visual graphs without labels (M=4.57, SD=1.40), $t(30)=-1.61$, $p>.05$, indicating that with or without labels, the visual graph readers split events into similar number of sub-events. The number of line segments (*the graph-segment count*) seemed to have significant effect on the number of segmentation applied by the participants, $F(2,58)=14.33$, $p<.001$, $\eta^2=.33$. As can be expected, if the graph has more line segments, in other words more trend changes, the participants segment the event depicted in the graph into more sub-events. Additional pairwise comparisons indicated no significant difference between the 5-segment graphs and the 6-segment graphs. On the other hand, the 7-segment graphs are significantly different from the 6-segment ($F(1,29)=14.442$, $p<.05$, $\eta^2=.32$) and from the 5-segments ($F(1,29)=22.55$, $p<.001$, $\eta^2=.44$.

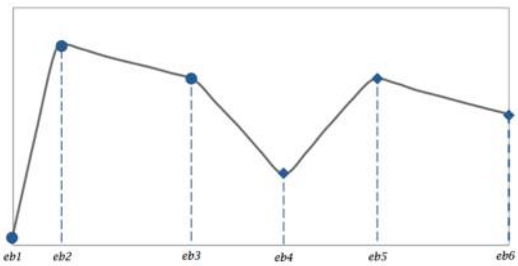
The stimuli-set is consisting of 12 graphs, resulting 12 different *global shape*s. A two-way mixed ANOVA (12 graphs and 2 conditions) was conducted to test the effect of the global shape on the number of the sub-events produced by the participants. The results indicated main effect of the global shape, $F(6.30, 132.36)=5.76$, $p<.001$, $\eta^2=.22$. However, there was no significant interaction between the global shape and the condition, $F(6.30, 132.36)=.93$, $p>.05$, indicating that for both conditions, the participants exhibited similar outcomes. In order to separate the effect of graph shape and the number of segment, additional analysis for the effect of graph shape within the each graph-segment count group (5-, 6- or 7-segments) were conducted. The findings indicated that the global shape had a significant effect on event segmentation for the 5 segment graphs, ($F(2.23,53.63)=4.57$, $p<.05$, $\eta^2=.17$), also for the 6 segment graphs ($F(4,92)=5.83$, $p<.001$, $\eta^2=.20$), but not for the 7-segment graphs ($F(2,50)=2.29$, $p>.05$). This may indicate that for the 5-segment and the 6-segment graphs the global shape may exhibit useful global features for event segmentation. For example, Graph-10 (Figure 6-18-left) is the one with less segmentation count within both the 5-segment graphs and all graph set. Since the contrasts between the local changes in the first three segments are less (a slight change, no change and again a slight change consecutively), these regions together may be conceptualized and mentioned as one sub-event. However the 7-segment graphs due to high number of inflection points that introduce more salient changes may lead focusing on the local features instead of the global features, see also Figure 6-19 and Figure 6-20 that represent least and most segmented graphs within each the 6-segment and the 7-segment graphs respectively.

Table 6-7 The number of shape segments and the average number of sub-events produced by participants for each graph in the stimuli set (▲: the highest value in the row, ▼: the lowest value in the row)

Graph No	Number of Shape Segments	Mean of Sub-Events divided by participants
Graph 1	6	4.36 (SD=1.55)
Graph 2	5	4.20 (SD=1.69)
Graph 3	6	3.68 (SD=1.87)
Graph 4	7	5.13 (SD=1.79) ▲
Graph 5	5	4.46 (SD=1.45)
Graph 6	6	4.16 (SD=1.53)
Graph 7	7	4.53 (SD=1.68)
Graph 8	5	3.90 (SD=1.76)
Graph 9	6	3.71 (SD=1.27)
Graph 10	5	3.29 (SD=1.57) ▼
Graph 11	6	5.09 (SD=1.57)
Graph 12	7	4.46 (SD=1.38)

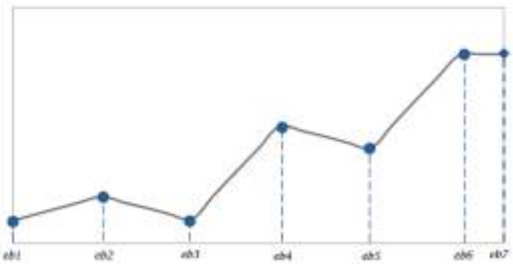


Graph-10, M=3.29

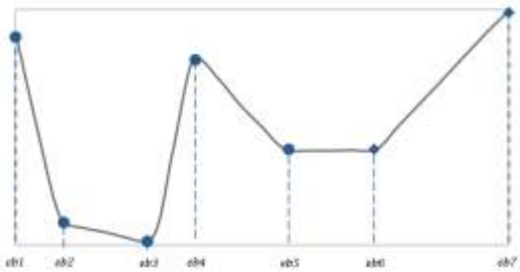


Graph-5, M=4.46

Figure 6-18 5-Segment Graphs (Left: less segmented, Right: More segmented)

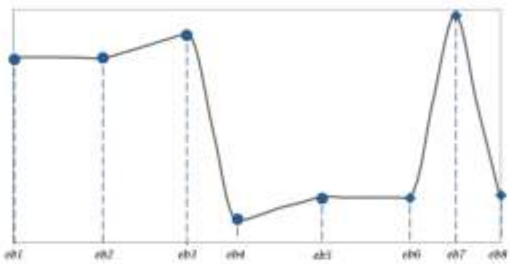


Graph-3, M=3.68

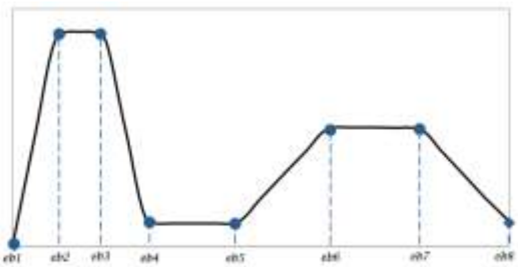


Graph-11, M=5.09

Figure 6-19 6-Segment Graphs



Graph-12, M=4.46



Graph-4, M=5.13

Figure 6-20 7-Segment Graphs

These results indicated that the graph-segment count has an effect on segmentation; however the prominence of the change at these possible breakpoints also seems to have influence on segmentation. Therefore more detailed analyses on the amodal geometric properties were conducted to understand the individual effect of each properties and also to reach a list of heuristics to be used for a verbally assistance system that segments a conceptual event automatically and effectively and produces verbal descriptions based on those segmented units.

B. Location of Event Boundaries – The Positioning Rate

In the following analyses, the users' positioning rate (reference rate) for each event boundary, as explained previously, was calculated concerning all graphs in the both conditions (visual graphs with/without labels). In the calculation of this parameter, first step was to code whether the event boundaries were explicitly referred by each participant, this was done for each event boundary on the graph set. Then the positioning rate for specific groups (with respect to each slope and angle values, i.e. the event boundaries before steep segments) was calculated by dividing the number of referred event boundary to the total amount of event boundaries under this specific group for each participant.

B.1 Main effect of slope

A two-way mixed ANOVA (2X4) was conducted to test the effect of presence of data labels (*the between-subject variable*), and the effect of segment's slope (*the within subject variable*). The results indicated that there was no significant main effect of the existence of data labels on the preferred location of event markers, indicating that the visual graph readers in both conditions locate the event boundaries in a similar way.

As explained before, the steepness of the graph segments (*the slope-value*) was set to four categorical values; (i) *no slope* – a horizontal line, (ii) *a slight slope* -15˚, (iii) *a medium slope* - 45˚ and (iv) *a steep slope* -75˚. A two-way ANOVA (2x4) was conducted in order to test the effect of slope and the effect of labels on the preferred location of the event markers. Table 6-8 presents the descriptive statistics. The findings indicated that the slope of the shape segment had a significant effect on whether that location is preferred as an event marker or not, $F(2.131, 66.076)=28.04$, $p<.001$, $\eta^2=.48$. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi2(5) =16.64$, $p < .05$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .71$).

Table 6-8 The reference rates w.r.t the slope-values

	Steep	Medium	Slight	Horizontal
VG with labels	0.83 (SD=.14)	0.68 (SD=.11)	0.32 (SD=.15)	0.40 (SD=.21)
VG without labels	0.80 (SD=.22)	0.74 (SD=.19)	0.49 (SD=.25)	0.57 (SD=.21)

Follow-up pairwise comparisons (see Figure 6-21) indicated that the landmarks located before the steep segments were employed as event boundaries more frequently than that before the medium segments $F(1,31)=11.042$, $p<.05$, $\eta^2=.26$. They were also more frequent than that before the slight segments $F(1,31)=46.70$, $p<.001$, $\eta^2=.60$ and that before the horizontal segments $F(1,31)=30.794$, $p<.001$, $\eta^2=.50$. The event boundaries before the medium segments were used more than that before the slight ($F(1,31)=33.83$, $p<.05$, $\eta^2=.52$) and the horizontal segments ($F(1,31)=20.18$, $p<.001$, $\eta^2=.39$). However there is no significant difference between the slight and the horizontal segments, $F(1,31)=1.45$, $p>.05$. This result suggests that the steeper the slope is, the higher the chance of the beginning of the segment to be selected as an event boundary is.

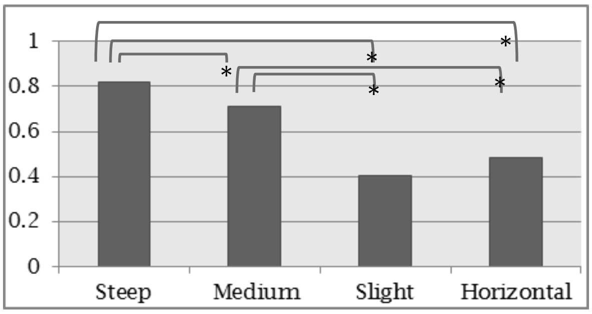


Figure 6-21 The preference rate for the each slope values

B.2 Main effect of angle and polarity

Not only the *slope-value* of a segment, but also *the turning angle* at the intersection of two line segments may have an effect on whether that landmark is chosen as an event boundary or not. Table 6-9 presents the four categories of angles ranging from the acute to the obtuse angles.

Table 6-9 Four categories of angle-values

Angle Categories	
Acute angles	A1: 30° & 60°
⇕	A2: 90° & 105°
	A3: 120° & 135°
Obtuse angles	A4: 150° & 165°

The results of a mixed design (2x4) ANOVA indicated that there was a main effect of angle (with a large effect size) on whether the corresponding landmark is chosen as an event boundary or not, $F(2.33, 69.92)=85.73$, $p<.001$, $\eta^2=.74$. The follow-up pairwise tests (see Table 6-10) indicated that while the difference between the A1 and the A2 value angles was not significant, the positioning rate for each of them was higher than that for the A3 ($F(1, 30)=27.59$, $p<.001$, $\eta^2=.48$, $F(1,30)=38.42$, $p<.001$, $\eta^2=.56$) and the A4 value angles, ($F(1, 30)=112.50$, $p<.001$, $\eta^2=.78$, $F(1, 30)=220.88$, $p<.001$, $\eta^2=.88$) respectively. Furthermore, the difference between the A3 and the A4-values were also significant, ($F(1, 30)=92.54$, $p<.001$, $\eta^2=.75$).

Table 6-10 The Positioning Rates for different angle values

	A1	**A2**	**A3**	**A4**
VG with labels	0.83	0.80	0.57	0.14
VG without labels	0.77	0.80	0.61	0.40

The interaction between the condition and the angle-value was also significant, $F(2.33, 69.92)=6.14$, $p<.01$, $\eta^2=.17$. Pairwise testing indicated that this interaction was mostly due to the difference between the A4-value landmarks to other angle-values. Overall, it can be concluded that for the visual graphs with labels, the participants did not prefer to use landmarks with very obtuse angle (such as 150° & 165°) as an event boundary, see Figure 6-22(a). Furthermore, the polarity of a landmark (whether it is concave or convex) did not have significant main effect ($p>.05$), indicating that the convex landmarks were referred in a similar rate as the concave landmarks were referred. On the other hand, the interaction between the angle and the polarity is significant $F(3, 90)=8.06$, $p<.001$, $\eta^2=.21$.

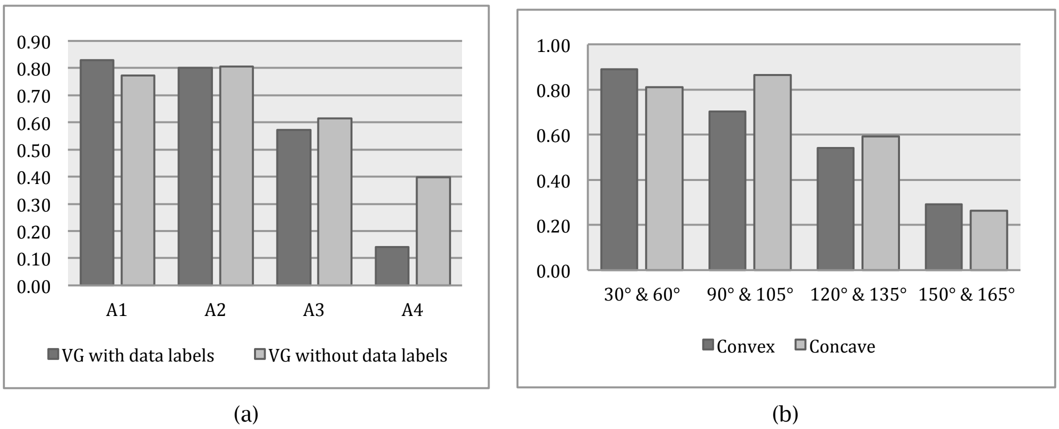
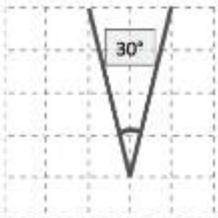


Figure 6-22 The positioning rates for (a) different angle values and (b) the interaction between polarity and angle

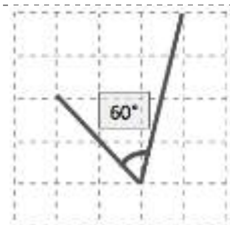
B.3 Effect of angle within each slope-value category

This analysis was conducted to see the effect of the angle within the each slope category. The angle is inherently dependent on the slope of the two segments that constitute it. Therefore, the contribution of the angle was investigated by keeping the slope of the following line segment constant. In other words, the contribution of the angle was isolated by conducting separate analyses for each slope-value.

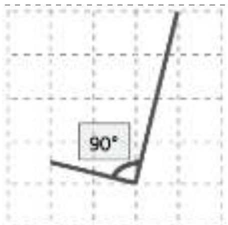
B.3.1 Landmarks before steep segments. All possible angle-values for a landmark before steep-slope segments are depicted in Figure 6-23. The landmarks that have at least one steep segment connection (in both increasing and decreasing directions) may have an angle value within three categories (A1: 30&60, A2: 90&105, and A3: 120&35). A two-way ANOVA was conducted to test the effect of angle of the landmarks preceding a steep segment on the preferred location as an event boundary (Table 6-11). The results showed no significant effect of the angle (F(1,60)=.11, p>.05) indicating that the angle does not have discriminatory power within the steep slope category for visually perceived graphs.



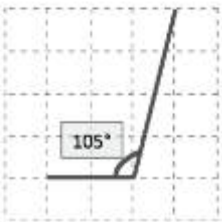
a) Steep decrease & steep increase



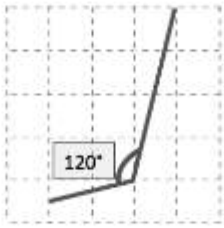
(b) Medium decrease & steep increase



(c) Slight decrease & steep increase



(d) Horizontal segment & steep increase



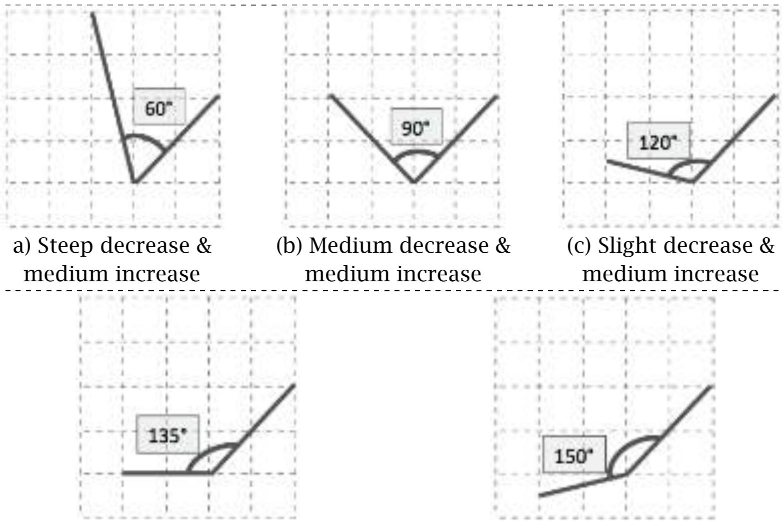
(e) Slight increase & steep increase

Figure 6-23 The illustration of the angle values with increasing steep segment connection

Table 6-11 The positioning rates for the angles with steep segment connections

	A1: 30˚, 60˚	A2: 90˚,105˚	A3: 120˚,135˚
VG with labels	0.84 (SD=.20)	0.84 (SD=.20)	0.81 (SD=.40)
VG without labels	0.70(SD=.29)	0.77(SD=.36)	0.75(SD=.45)

B.3.2 Landmarks before medium segments. All possible angle-values for a landmark located before the medium-slope segments are depicted in Figure 6-24. Unlike the landmarks that precede a steep segment, there was a main effect of the angle on the positioning rates for the landmarks that precede a medium segment, ($F(2.26,65.66)=31.05$, $p<.001$, $\eta^2=.52$). The interaction between the angle and the condition was also significant, ($F(2.26, 65.66)=6.12$, $p<.001$, $\eta^2=.17$) with a small effect size. Table 6-12 summarizes the average values for each angle-value. The reference rate for the event boundaries with the A1 and the A2 angles did not exhibit significant difference, $F(1,29)=3.87$, $p>.05$. However, the event boundaries with the A1-angles were mentioned more than the A3-angles ($F(1,29)=8.21$, $p<.05$, $\eta^2=.22$) and more than the A4-angles ($F(1,29)=32.89$, $p<.001$, $\eta^2=.53$). Similarly, the event boundaries with the A2-angles were mentioned more than the A3-angles ($F(1,29)=28.74$, $p<.001$, $\eta^2=.50$) and more than A4-angles (($F(1,29)=64.21$, $p<.001$, $\eta^2=.69$). The A3-angles also showed significant difference and were mentioned more than the A4-angles, $F(1,29)=22.53$, $p<.001$, $\eta^2=.44$). These results indicated that for the medium slope segments, the angle between the corresponding and the previous segment contributes to its saliency and have an effect on whether this point is selected as an event boundary or not.



a) Steep decrease & medium increase

(b) Medium decrease & medium increase

(c) Slight decrease & medium increase

(d) Horizontal segment & medium increase

(e) Slight increase & medium increase

Figure 6-24 The illustration of the angle values with an increasing-medium segment connection

Table 6-12 The positioning rates for the angles with medium segment connections

	A1: 30°, 60°	**A2: 90°,105°**	**A3: 120°,135°**	**A4: 150°,165°**
VG with labels	0.81 (SD=.25)	0.97(SD=.13)	0.67 (SD=.15)	0.22 (SD=.33)
VG without labels	0.83 (SD=.31)	0.87 (SD=.23)	0.71 (SD=.24)	0.58 (SD=.35)

B.3.3 Landmarks before slight segments. Figure 6-25 depicts all the possible angle-values in this category. The angle showed significant effect for the landmarks preceding slight segment connections, $F(2,58)=37.20$, $p<.001$, $\eta^2=.56$. The interaction between the angle and the condition was not significant ($F(2, 58)=.69$, $p>.05$). The follow-up pairwise comparisons showed that the A2-value angles were mentioned more than the A3-value $F(1,29)=20.21$, $p<.001$, $\eta^2=.41$ and more than the A4-value angles (see Table 6-13). The landmarks with the A3-value angles were also mentioned more than the landmarks with the A4-value angles, $F(1,29)=17.72$, $p<.001$, $\eta^2=.38$. These results indicated that the angle is an important factor in choosing a landmark that precedes a slight segment as an event boundary.

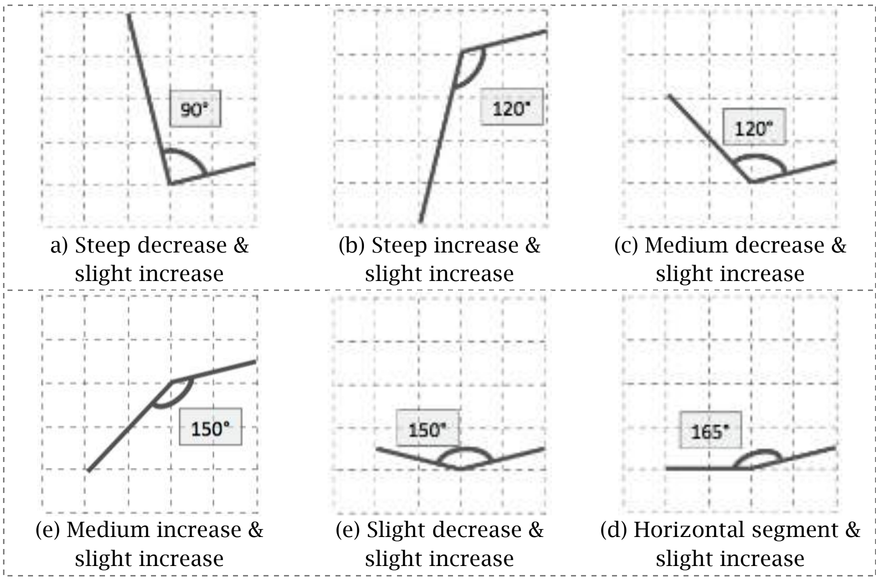


Figure 6-25 The illustration of the angle values with an increasing-slight segment connection

Table 6-13 The positioning rates for the angles with slight segment connections

	A2: 90°,105°	**A3: 120°,135°**	**A4: 150°,165°**
VG with labels	0.73(SD=.32)	0.41 (SD=.24)	0.17 (SD=.17)
VG without labels	0.78 (SD=.31)	0.54 (SD=.38)	0.35(SD=.27)

B.3.4 Landmarks before horizontal segments. Figure 6-26 also presents all the possible angle-values for this category. The angle-value of the landmarks that precede horizontal segments also exhibited the same pattern with those that precede slight segments. The angle showed main effect on whether this location is chosen as an event boundary, $F(2,60)=61.38$, $p<.001$, $\eta^2=.67$. The landmarks with the A2-value angles were mentioned more than those with the A3-value angles, $F(1,29)=40.44$, $p<.001$, $\eta^2=.57$ and more than those with the A4-value angles (see

Table 6-14). The A3-value angle landmarks were also referred more than the landmarks with the A4-value angle, $F(1,29)=25.48$, $p<.001$, $\eta^2=.46$.

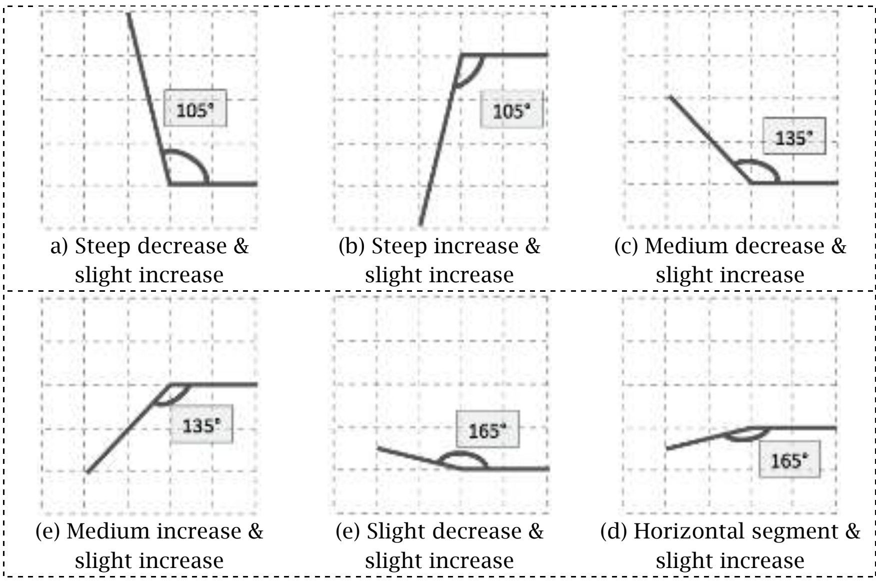


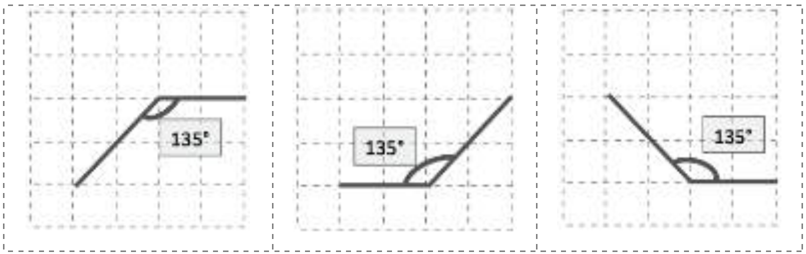
Figure 6-26 The illustration of the angle values with an increasing-horizontal connection

Table 6-14 The positioning rates for the angles with horizontal steep segment connections

	A2: 90˚,105˚	A3: 120˚,135˚	A4: 150˚,165˚
VG with labels	0.65 (SD=.31)	0.42 (SD=.29)	0.05 (SD=.14)
VG without labels	0.81 (SD=.17)	0.45 (SD=.33)	0.29 (SD=.30)

B.4 Landmark´s Type with respect to the qualitative ascriptions

The steepness (*the slope-value*) of the segments is related to the angle between the line and the orthogonal axes, namely the x- and y-axes. Therefore, the reading of how steep the segment is performed w.r.t the spatio-temporal reference frame of the graph. On the other hand, *the angle* of a landmark is less dependent on spatio-temporal perspective. In other words, a very steep line segment (i.e. with the slope of 75˚) can have only two variations (increasing or decreasing). On the contrary, wide variety of distinct landmarks can be formed from the same angle-value. Figure 6-27 depicts only three variation of one angle-value. Therefore the effect of another parameter, which is more sensitive to spatio-temporal characteristics of a statistical graph, namely *the qualitative landmark type,* was also explored. The qualitative ascriptions can be grouped into two: (1) the spatio-temporal ascriptions and (2) the conceptual ascriptions. As discussed in Chapter 5.2, the qualitative ascriptions of the graphical entities (the landmarks and the segments) play crucial role in event segmentation. This analysis addresses the question of whether these ascriptions affect the location of an event boundary.



(a) (b) (c)

Figure 6-27 Different looking landmarks with the same angle value

Spatio-temporal ascriptions for landmarks. Concerning statistical line graphs, the change in what is represented in the y-axis, lets say *the tourist visit count* displays three main different types of a change; the tourist visit count may (i) increase, (ii) decrease or (ii) remain stable. Based on this classification, three types of landmark can emerge w.r.t the direction of the change in the previous and the next data points. Consider Figure 6-28, which illustrates samples for each type. First, a landmark may belong to an intersection of two line segments, which have a connection to landmarks with higher values, like *the sp2* represents. This case represents an exemplar of *the direction-change, where* both the previous (*sp1*) and the next (*sp3*) line segments have different values than the intermediate landmark (*sp2*) and this landmark connects two line segments that represent two opposite directions (downward and upward respectively). In a formal way, the qualitative relation of the landmark with its previous and next landmarks can be represented with a notation like, *relation (source, goal).* For this example given above, the consecutive landmarks have *higher (sp1, sp2)* and *lower(sp2, sp3)* relations. The landmarks in the second category (i.e. *sp3*) do not exhibit a direction change w.r.t the x-axis. Both previous and next data points have different values and the line segments are in the same direction (in this case, both of them have upward direction). The formal notations for this would be *lower (sp2, sp3)* and *lower(sp3, sp4).* This category represents a change within a trend (*trend-change*). The landmarks in the last category have only one neighbor data point that has different value therefore it exhibits only one directional change w.r.t the x-axis. The formal notations *lower (sp3,sp4)* and *same(sp4,sp5)* would represent the relation for this case. This category (*both types of changes*) lies between the former two categories, since the change represented here is neither solely the direction change nor solely the trend change.

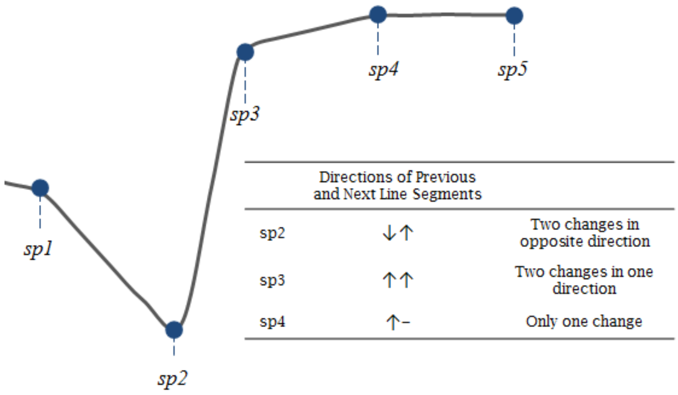


Figure 6-28 The Qualitative Landmark Types

A 2X3 ANOVA (the condition x the landmark type) indicated that there was no significant difference between the two visual graph conditions, on the other hand there was a significant main effect of a landmark type on the positioning rate, $F(1.59, 47.97)=11.89$, $p<.001$, $\eta^2=.28$. The follow-up tests indicated that the positioning rate for the landmarks that introduce a direction change was higher than that for the landmarks with a trend change ($F(1,30)=17.17$, $p<.001$, $\eta^2=.36$) and higher than that for the landmarks with both change $F(1,30)=12.80$, $p<.001$, $\eta^2=.30$. The latter two were also significantly different from each other, $F(1, 30)=4.75$, $p<.05$, $\eta^2=.14$, see Table 6-15. The interaction between the condition and the landmark type was not significant.

Table 6-15 The positioning rates for the different landmark types (revise it)

	Direction Change	**Trend Change**	**Both Changes**
VG with labels	.58	.39	.47
VG without labels	.63	.50	.57

Conceptual ascriptions for landmarks. Additional analyses on the qualitative ascriptions of landmarks were conducted to investigate whether the conceptual ascriptions have an effect on event segmentation. A landmark, in the graph domain, may correspond to a global point (the global maximum or minimum), to a local point (the local maximum or minimum) or it may correspond to the other kinds of smooth points which introduces a change in the trend as introduced earlier. The results of a two-way ANOVA (2x3) showed that there was no significant difference between the two visual graph conditions. Table 6-16 presents the descriptive statistics for each category across the conditions. However, the difference among the three types of landmarks (the global points, the local points and the other) showed significant main effect with a large effect size, $F(1.66, 49.79)=108.34$, $p<.001$, $\eta^2=.78$. The follow-up pairwise tests revealed that the global points ($M=.78$, $SD=.15$) were chosen more as an event boundary than the local points ($M=.54$, $SD=.20$) $F(1,30)=165.72$, $p<.001$, $\eta^2=.85$) and more than the other points ($M=.41$, $SD=.23$) ($F(1,30)=176.88$, $p<.001$, $\eta^2=.85$). The positioning rate for the local points were also higher than that for the other points, $F(1,30)=19.25$, $p<.001$, $\eta^2=.39$). Furthermore, the interaction between the modality and the landmark type was also significant, $F(2,60)=4.55$, $p<.05$, $\eta^2=.13$).

Table 6-16 The descriptive statistics for the positioning rates of different landmark types

	Global	**Local**	**Other**
VG with labels	0.78 (SD=.11)	0.47 (SD=.17)	0.33 (SD=.18)
VG without labels	0.79 (SD=.18)	0.60 (SD=.22)	0.49 (SD=.25)

Additional 2x2 ANOVAs also indicated that there was no significant difference in the positioning rates between the global minima and maxima, p>.05. Same result was obtained for the difference between the local minima and the local maxima, p>.05, see Table 6-17 for the descriptive statistics.

Table 6-17 The descriptive statistics for the positioning rates of global and local points

	Global Points		**Local Points**	
	Min	**Max**	**Min**	**Max**
VG with labels	0.78 (SD=.13)	0.77 (SD=.11)	0.45 (SD=.22)	0.49 (SD=.15)
VG without labels	0.82 (SD=.19)	0.76 (SD=.19)	0.59 (SD=.25)	0.62 (SD=.29)

(B) Interim Discussion (*for the label-group*)

The analyses presented here focused on the location of the event boundaries. The findings indicated that the visual graph readers in both condition (with and without labels) locate event boundaries in a similar way. In order to investigate the effects of the different components of a graph, various perceptual (such as the slope and the angle) and conceptual (the landmark type, i.e global max.) features were taken into consideration. First, the slope of a segment had a strong effect on whether that location (the landmark that precedes it) is chosen as an event boundary or not. In other words, the steeper the slope is, the higher the chance of the beginning of the segment to be selected as an event boundary is. Furthermore, the angle of a landmark at the intersection of the two line segments is another factor that effects event segmentation. The results indicated that more the angle acute is, the higher the chance of being selected as an event boundary is. On the other hand, the polarity of a landmark (concave or convex) seems to have no contribution in event segmentation. Additional analyses conducted to explore the effect of angle-value within each slope category also revealed that if the following line segment is steep (with 75°), the angle of the landmark does not exhibit additional effect on whether this location is marked or not. However, the angle of landmarks shows significant differences within the rest of the three angle-value categories, exhibiting the same pattern of overall (main) angle effect. It should be noted while the positioning rate did not show any difference w.r.t the presence of data labels, the positioning rates for the A4-values were extremely low in the "*with-label*" condition and this result was consistent for all slope value categories. The positioning rate w.r.t the qualitative ascriptions suggested no effect of the data labels. In this data-set, the landmarks that introduce a direction change were used as an event boundary more, and the positioning rate for the landmarks that introduce a change in trend was the lowest. Last, according the findings of the conceptual ascriptions on the positioning rate, the global points were chosen as an event boundary more, then the local points followed. A closer look to the difference in the positioning rates between the minimum and the maximum points did not show any difference, indicating that both of them are equally distinct in terms of event segmentation.

6.4.6.1.2. The Effect Of The Sensory Modality

A. Number of Sub-Events

Same statistical analyses that were applied to the Label-group were also performed for the Modality-Group, and reported in the same order. Firstly, the two modalities were compared by conducted t-test on the mean of sub-event count overall, The results showed that there was no significant difference between the visual graphs (M=4.57, SD=1.40) and the haptic graphs (M=4.55, SD=1.26), t(30)=.34, p>.05. A two-way mixed ANOVA (12 graphs and 2 conditions) was conducted to test the effect of the global shape on how many sub-events produced by the users. As a result, significant main effect of the global shape was observed, F(6.66, 146.65)=7.75, p<.001, η^2=.26. However, there was no significant interaction between the global shape and the condition, F(6.66, 146.65)=1.76, p>.05.

The number of line segments (*the graph-segment count*), which the graph contains, exhibited a significant effect on the event segmentation, $F(2,60)=27.19$, $p<.001$, $\eta^2=.47$, see Table 6-18. The average segmentation count for the 5-segment events was significantly lower than the 6-segment events $F(1,30)=10.59$, $p<.05$, $\eta^2=.26$. Moreover, the segmentation for the 7-segment events was significantly higher than the one for the 6-segment ($F(1,30)=19.20$, $p<.001$, $\eta^2=.38$) and for the 5-segment graphs ($F(1,30)=47.53$, $p<.001$, $\eta^2=.61$. No significant interaction effect of modality and the segment count was observed. Further analyses indicated that the global shape had significant effect on event segmentation for the 5-segment graphs, ($F(3,75)=7.31$, $p<.001$, $\eta^2=.23$), for the 6-segment graphs ($F(2.94, 67.58))=3.12$, $p<.05$, $\eta^2=.12$), and also for the 7-segment graphs ($F(2,52)=5.81$, $p<.05$, $\eta^2=.18$), These result, as in line with the results of data-label group, indicate that not just the graph-segment count but the global shape has an effect on the segmentation.

Table 6-18 The number of shape segments and the average number of sub-events produced by the participants for each graph (▲:the highest value in the row, ▼: the lowest value in the row)

Graph No	Number of Shape Segments	Mean of Sub-Events divided by participants
Graph 1	6	4.37 (SD=1.98)
Graph 2	5	4.43 (SD=1.94)
Graph 3	6	4.61 (SD=1.64)
Graph 4	7	5.63 (SD=1.71) ▲
Graph 5	5	4.80 (SD=1.47)
Graph 6	6	4.43 (SD=1.62)
Graph 7	7	4.75 (SD=1.58)
Graph 8	5	3.74 (SD=1.71)
Graph 9	6	4.06 (SD=1.36)
Graph 10	5	3.46 (SD=1.54) ▼
Graph 11	6	5.19 (SD=1.70)
Graph 12	7	5.31 (SD=1.77)

B. Location of Event Boundaries – The Positioning Rate

B.1 Main effect of slope

A two-way mixed ANOVA (2X4) was conducted to test the effect of sensory modality (*the between-subject variable*), and the effect of the steepness (*the within subject variable*). The findings revealed that there was no significant main effect of the sensory modality on the preferred location of the event markers ($F(1,29)=.01$, $p>.05$) indicating that both the visual graph readers and the haptic graph readers segment the graph lines in a similar way, see Table 6-19 for the descriptive statistics. The steepness of the shape segments had a significant effect on whether that location is preferred as an event marker or not, $F(3,87)=67.45$, $p<.001$, $\eta^2=.70$ with large effect size. Additionally, there was a significant interaction effect between the slope-value and the modality with a small effect size, $F(3,87)=4.55$, $p<.001$, $\eta^2=.14$.

Table 6-19 The reference rates w.r.t. slope-values

	Steep	Medium	Slight	Horizontal
VG without labels	0.80 (SD=.22)	0.74 (SD=.19)	0.49 (SD=.25)	0.57 (SD=.21)
HG without labels	0.89 (SD=.10)	0.75 (SD=.18)	0.50 (SD=.22)	0.43 (SD=.20)

Additional pairwise testing indicated that the landmarks before the steep segments were employed as event boundaries more than those before the medium segments $F(1,29)=14.08$, $p<.05$, $\eta^2=.33$), more than those before the slight segments $F(1,29)=102.47$, $p<.001$, $\eta^2=.78$ and more than those before the horizontal segments ($F(1,29)=95.53$, $p<.001$, $\eta^2=.77$). The event boundaries before the medium segments were used more than those before the slight segments ($F(1,29)=91.08$, $p<.001$, $\eta^2=.76$) and more than those before the horizontal segments ($F(1,29)=64.05$, $p<.001$, $\eta^2=.69$). However there was no significant difference between the slight and the horizontal segments, $F(1,29)=.36$, $p>.05$.

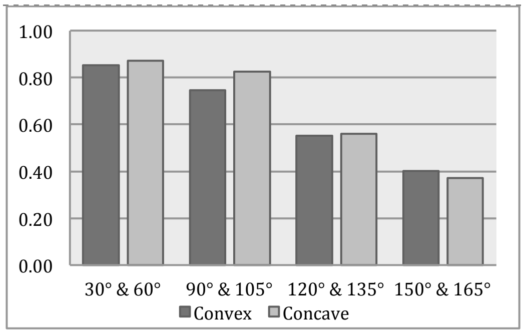
B.2 Main effect of angle and polarity

The findings revealed that there was a main effect of angle (with a large effect size) on whether the corresponding landmark is chosen as an event boundary, $F(2.27, 68.21)=52.56$, $p<.001$, $\eta^2=.64$.

Follow-up pairwise tests indicated that while the difference between the A1 and the A2 was not significant, the positioning rate for each of them was higher than that for the A3 ($F(1, 30)=44.36$, $p<.001$, $\eta^2=.59$, $F(1,30)=37.79$, $p<.001$, $\eta^2=.56$) and higher than that for the A4 ($F(1, 30)=73.74$, $p<.001$, $\eta^2=.71$, $F(1, 30)=85.84$, $p<.001$, $\eta^2=.74$) respectively. Furthermore, the difference between the A3 and A4-values were also significant, ($F(1, 30)=24.95$, $p<.001$, $\eta^2=.45$). However, the interaction between the angle and the condition was not significant, see Figure 6-29(a). The polarity of the landmark did not have significant main effect, $p>.05$. The interaction between the angle and the polarity was also not significant, $p>.05$, see Figure 6-29(b).

	A1	A2	A3	A4
VG without labels	0.77	0.80	0.61	0.40
HG without labels	0.90	0.79	0.52	0.37

(a)



(b)

Figure 6-29 (a) Positioning Rates for different angle values and (b) the interaction between polarity and angle

B.3. Effect of angle within each slope-value category.

B.3.1. Landmarks before steep segments. The landmarks with at least one steep segment connection may have angle-values from three categories (A1, A2, A3), see Figure 6-23 for all the possible slope and angle combinations that a steep segment can take. A two-way ANOVA was performed to test the effect of angle of the landmark preceding a steep segment on the preference of the event boundary location. The results showed a significant effect of the angle $F(1.50, 45.00)= 4.99$, $p<.05$, $\eta^2=.14$). The overall reference rates for each angle-value are summarized in Table 6-20. While there was no significant difference in the positioning rates for the A1-value and the A2-value angles, the difference between the A2-value and the A3-value angles was significant.

Table 6-20 The angle values of a landmark with a steep segment connection

	A1: 30˚, 60˚	**A2: 90˚,105˚**	**A3: 120˚,135˚**
VG without labels	0.70 (SD=.29)	0.76 (SD=.36)	0.75 (SD=.45)
HG without labels	0.86 (SD=.22)	0.92(SD=.15)	0.47(SD=.49)

The interaction between the angle and the modality was also significant, $F(1.50, 45.00)= 5.40$, $p<.05$, $\eta^2=.15$). One could interpret this as the landmarks with the acute angles (A1 and A2) are preferred more as an event boundary in the haptic modality, while the landmark with the A3 angle shows opposite trend. The positioning rate for the landmarks that precede a steep segment were almost similar for all angle-values across conditions except the A3-value in the haptic modality. Therefore, a closer look for the A3–value is required to understand this interaction. Figure 6-30 depicts a case that represents a landmark with the A3-value that preceded a steep segment. This significant interaction presented above suggests that for the visual graph readers, this landmark has high saliency and is a potential candidate to be chosen as an event boundary similar to landmarks with more acute angles. On the contrary, for the haptic graph explorer, this distinction was not that sharp. From the action perspective, when exploring actively this region, the haptic explorers perform upward movement, and this landmark introduces a change only in the general trend without a change in the basic direction such as an increase (see Figure 6-23 for the illustrations of all possible combinations). This may be the reason of why these kinds of regions were chosen less as event boundary in the haptic condition.

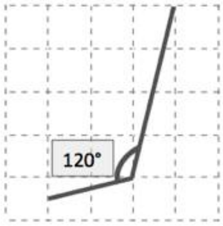


Figure 6-30 An illustration of the A3 angle-value that precedes a steep segment

B.3.2. Landmarks before medium segments. For the landmarks that precede a medium segment, there was a significant main effect of angle, ($F(1.88, 54.50)=11.93$, $p<.001$, $\eta^2=.29$). Table 6-21 presents the descriptive statistics for this analysis. The difference between the A1-value and the A2-value angles was not significant, ($F(1,29)=.00$, $p>.05$). However, the landmarks with the A1-value angles referred more than the landmarks with the A3-value angles ($F(1,29)=16.14$, $p<.001$, $\eta^2=.36$) and more than the A4-value angles ($F(1,29)=15.58$, $p<.001$, $\eta^2=.35$). Similarly, the referring rate for the landmarks with the A2-value angles were also significantly higher than the A3-value angles ($F(1,29)=8.53$, $p<.001$, $\eta^2=.23$) and higher than the A4-value angles ($F(1,29)=15.32$, $p<.001$, $\eta^2=.35$). Finally, the landmarks with the A3-value angles were referred more than those with the A4-value angles ($F(1,29)=7.67$, $p<.05$, $\eta^2=.21$). The interaction between the angle and the condition was not significant, ($F(1.88, 54.50)=.67$, $p>.05$).

Table 6-21 The angle with a medium segment connection

	A1: 30˚, 60˚	**A2: 90˚,105˚**	**A3: 120˚,135˚**	**A4: 150˚,165˚**
VG without labels	0.83 (SD=.30)	0.87(SD=.22)	0.71 (SD=.24)	0.58 (SD=.35)
HG without labels	0.94 (SD=.17)	0.91 (SD=.27)	0.73 (SD=.24)	0.50 (SD=.40)

B.3.3. The landmarks before slight segments. The angle at the event boundaries also showed a significant effect for the landmarks with a slight segment connection, $F(1.63, 49.01)=29.37$, $p<.001$, $\eta^2=.49$. Additionally, the landmarks with the A2-value angles were chosen more as an event boundary than the landmarks with the A3-value angles ($F(1,30)=16.99$, $p<.001$, $\eta^2=.36$ and with the A4-value angles. There was also a significant difference in the referring rates of the landmarks with the A3-value angles and the A4-value angles, ($F(1,30)=17.17$, $p<.001$, $\eta^2=.36$. The interaction between the angle and the condition was not significant ($F(1.63, 49.01)=.96$, $p>.05$).

Table 6-22 The angle values for the landmarks with a slight segment connection

	A2: 90˚,105˚	**A3: 120˚,135˚**	**A4: 150˚,165˚**
VG without labels	0.78(SD=.31)	0.54 (SD=.37)	0.35 (SD=.27)
HG without labels	0.78 (SD=.36)	0.51 (SD=.27)	0.36(SD=.22)

B.3.4. The landmarks before horizontal segments. The angle for the landmarks that precede a horizontal segment also exhibited same pattern with those precede a slight segments $F(1.66, 49.72)=29.18$, $p<.001$, $\eta^2=.49$. The landmarks with the A2-value angles were chosen more as an event boundary than the landmarks with the A3-value angles, F(1,30)=43.23, p<.001, η^2=.59 and more than that with the A4-value angles. However, there was no significant difference between the landmarks with the A3- value and the A4-vale angles, (F(1,30)=3.64, p>.05.

Table 6-23 The angle values for the landmarks with a horizontal segment

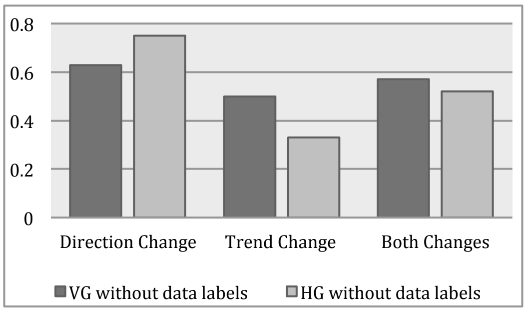
	A2: 90°,105°	**A3: 120°,135°**	**A4: 150°,165°**
VG without labels	0.80 (SD=.17)	0.45 (SD=.33)	0.29 (SD=.30)
HG without labels	0.57 (SD=.25)	0.34 (SD=.36)	0.26 (SD=.20)

B.4 Landmark´s Type

The effect of landmark’s qualitative type as explained earlier was also tested for the Modality-Set. The findings indicated that there was no significant difference between the two modalities, however there was a significant main effect of landmark type on the positioning rate, F(1.35, 40.48)=25.72, p<.001, η^2=.47. The follow-up tests indicated that the positioning rate for the landmarks that introduce a direction change was higher than that for the landmarks with a trend change (F(1,30)=32.56, p<.001, η^2=.52) and to that for the landmarks that have both partially, F(1,30)=38.15, p<.001, η^2=.56. The latter two were also significantly different than each other, F(1, 30)=10.89, p<.05, η^2=.27, see Figure 6-31. The interaction between the condition and the landmark type was also significant, F(1.35, 40.48)=7.28, p<.01, η^2=.19. Pairwise comparisons indicated that this significant interaction effect was resulted due to the differences in the “direction change” and in the “trend change” across modalities.

	Direction Change	**Trend Change**	**Both**
VG without labels	.63	.50	.57
HG without labels	.75	.33	.52

(a)



(b)

Figure 6-31 The positioning rates for the spatio-temporal ascriptions of landmarks

Additional analyses on the qualitative ascriptions of landmarks were conducted to investigate whether the conceptual ascriptions have an effect on the event segmentation. No significant difference was found between the two modalities. However, the difference among the three types of the landmarks (the global points, the local points and the others) showed a significant main effect with a large effect size, F(2, 60))=64.28, p<.001, η^2=.68. The follow-up pairwise tests revealed that the global points (M=.79, SD=.15) were chosen more as an event boundary than the local points (M=.66, SD=.22) (F(1,30)=29.06, p<.001, η^2=.49) and the other points (M=.45, SD=.23) (F(1,30)=134.94, p<.001, η^2=.82). The positioning rate for the local points were also higher than that for the other points, F(1,30)=33.56, p<.001, η^2=.53).

Table 6-24 The positioning rates for the conceptual ascriptions of landmarks

	Global	**Local**	**Other**
VG without labels	0.79 (SD=.18)	0.60 (SD=.22)	0.49 (SD=.25)
HG without labels	0.79 (SD=.12)	0.71 (SD=.22)	0.42 (SD=.21)

Furthermore, the interaction between the modality and the landmark type was also significant, F(2,60)=4.62, p<.05, η^2=.13). Again, follow-up pairwise tests indicated that this difference is due to the differences between the global and the local points and also between the other and the local points. Furthermore, the global minimums were mentioned more than the global maximums, $F(1,30)=13.20$, $p<.01$, $\eta^2=.31$. However, no difference between the local minimums and the local maximums was observed, see Table 6-25.

Table 6-25 The positioning rates for the global and local points

	Global Points		**Local Points**	
	Min	**Max**	**Min**	**Max**
VG without labels	0.82 (SD=.19)	0.76 (SD=.19)	0.59 (SD=.20)	0.62 (SD=.30)
HG without labels	0.84 (SD=.13)	0.74 (SD=.12)	0.66 (SD=.25)	0.76 (SD=.23)

(B) Interim Discussion ***(for the modality-group)***

The analyses focused on the location of the event boundaries indicated that both the visual and the haptic graph users locate event boundaries in a similar way. Following the same analyses conducted in the investigation of the effect of data labels, the effects of the perceptual features (such as slope and angle) and of the conceptual features (the landmark type, i.e. the global max.) were analyzed. First, the slope of a segment had a strong effect on whether its beginning is chosen as an event boundary or not. In other words, the steeper the slope is, the higher the chance of the beginning of the segment to be selected as an event boundary is. Furthermore, the angle of a landmark at the intersection of two line segments is another factor that has influence on the event segmentation. The results indicated that more the angle acute is, the higher the chance of being selected as an event boundary is. On the other hand, the polarity of a landmark (concave or convex) seems to have no contribution in the event segmentation. Additional analyses conducted to explore the effect of angle-value within each slope category also revealed the angle of landmarks showed significant differences within the all four angle-value categories, exhibiting the same pattern of overall (main) angle effect. This indicates that both slope and angle-values had an effect on event segmentation. The positioning rate w.r.t qualitative ascriptions suggested no effect of the modality. In this data-set, the landmarks that introduce changes in direction were used as an event boundary more, and the positioning rate for the landmarks that introduce trend changes was the lowest. Last, according to the findings of the conceptual ascriptions on the positioning rate, the global points were chosen as an event boundary more. The local points were employed as an event boundary as well. A closer look to the difference in the positioning rates between the minima and maxima points showed that the global minimum was mentioned more than global maximum; however no difference was found between the local points. The overall interpretation of the results is discussed in interim summary for the "event segmentation".

6.4.6.2.Event Segmentation In Speech-Accompanying Gestures

In addition to the analysis of the verbal descriptions, the analysis of speech-accompanying gestures provides valuable information about how a graph reader parse the event represented in a graph. To that end, two parameters were taken into account, (i) the number of gestures produced during the description of each graph across conditions and (ii) the relation between the sub-event description and its accompanying gesture(s). Only representational gestures were counted excluding the beat gestures.

A. Average Amount of Representational Gestures

The effect of data label on gesture production was tested by performing a non–parametric Mann-Whitney U test, since the assumption of equality of variances was violated. The results (see Table 6-26 for the descriptive statistics) revealed that the number of the representational gestures produced for each graph was not affected by the presence of data labels ($U=94.50$, $z=-1.27$, $p>.05$). Although the difference between the two visual graphs with or without labels seemed different visually as seen in Figure 6-32, the difference was not significant statistically.

Furthermore, the sensory modality also did not have any effect on the number of the gestures produced during verbal descriptions, t(30)=.21, p>.05). Additionally, the gesture production was not affected by the global shape, F(6.51,201.82)=1.82, p>.05.

Table 6-26 The descriptive statistics of the speech-accompanying gestures for each condition

	Total	**#of Participants who produced gesture**	**Average Gesture Amount (per description)**
VG with labels	273	12 of 16	M=1.64 (SD=1.75)
VG without labels	555	13 of 16	M=3.00 (SD=2.84)
HG without labels	591	15 of 16	M=3.19 (SD=2.20)

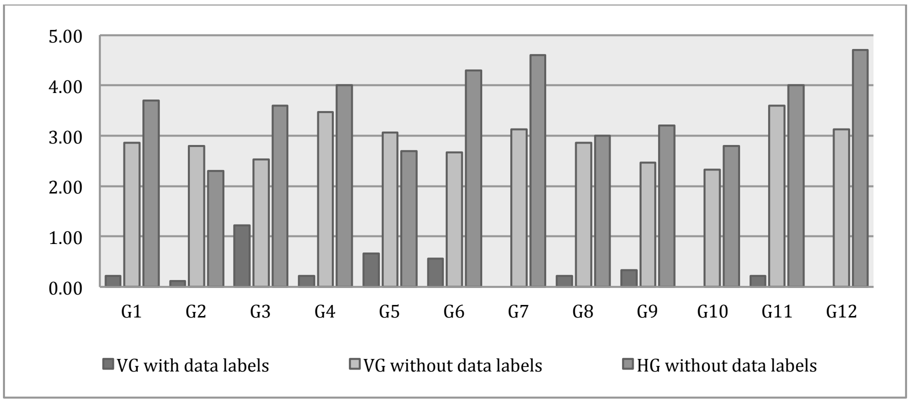


Figure 6-32 Average amount of gestures for each graph and for each condition

B. Gesture & Sub-Event Description Relation (Mapping)

As aforementioned, as well as exploring the event segmentation by looking at the verbal descriptions, investigating how this segmentation in the linguistic layer reflects itself in another, namely the gesture layer, would reveal valuable insights about the underlying mechanisms both of event segmentation and also of gesture-language relation. This analysis addresses the relation between the gesture and its accompanying speech part. For this purpose, the gestures were coded w.r.t their accompanying sub-event descriptions in terms of three relation types; first relation denotes "*one-to-one"* relation between the gesture and the sub-event description, in other words, if only one gesture was produced for a sub-event description (or *phrases*), then the relation was classified as "one-to-one". Second category corresponds to "*many-to-one*" referring to the cases with more than one gestures produced for a sub-event. Last category is called as "*one-to-many*", and this category represents a relation for the cases of having one gesture that accompanies to more than one sub-event description[27]. The examples for each category were illustrated below in detail.

(i) One-to-one relation: one gesture for one sub-event

Verbal description of the sub-event:	**Gesture properties**	
"This graph, at the beginning, stays at the stable value" (In Turkish: Bu grafik başta sabit bir degerde kalmış	• *Gesture type*: Dynamic 1-directional • *Direction:* Horizontal • *Hand*: Left	• *ASL-Letter:* 5 • *Start-Location:* Center • *End-location:* Center • Movement:Arm

[27] It should be noted that a full graph description may contain more than one sub-event, therefore a full description that is divided into more than one sub-events may contain all kinds of relations.

(ii) Many-to-one relation: more than one gesture for one sub-event

Verbal description of the sub-event		Gesture properties	
"[to the bottom value] [after it has fallen]" "after it has fallen to the bottom value" (In Turkish: en son noktaya düştükten sonra	**Gesture-1**	• *Gesture type*: Dynamic • 1-directional • Direction: Diagonal • *Hand*: Left • *ASL-Letter:* 5	• *Start-Location:* Center • *End-location:* Bottom • Movement: Arm
	Gesture-2	• *Gesture type*: Dynamic 1-directional • *Direction: Diagonal* • *Hand*: Left	• *ASL-Letter:* 5 • *Start-Location:* Center • *End-location:* Center • Movement: Arm

(iii) One-to-many relation: one gesture for more than one sub-event.

	Verbal description of the sub-event	Gesture properties	
Sub-Event 1	"Then, it increases a little" (In Turkish: Sonra çok az artıyor.)	• *Gesture type*: Dynamic Multi-directional • *Direction:* increasing diagonal and decreasing diagonal • *Hand:* Left	• *ASL-Letter:* Q • *Start-Location:* Center • *End-location:* Center • Movement: Arm
Sub-Event 2	"There is another decrease again." (In Turkish: Tekrar bir azalış oluyor.)		

The comparisons were conducted on the average amount of gestures for each relation type. The data belongs to the protocols in which at least one representational gesture was observed. The results indicated that there was a significant difference in the numbers of produced gestures with respect to the types of the gesture-event relation with a very large effect size, $F(2,74)=222.42$, $p<.001$, $\eta^2=.86$. As seen in Table 6-27, "*one-to-one*" gestures were produced more than the "*many-to-one*" $F(1,37)=114.03$, $p<.001$, $\eta^2=.75$ and more than the "*one-to-many*", also the "*many-to-one*" gestures was significantly higher than the "*one-to-many*" $F(1,37)=51.94$, $p<.001$, $\eta^2=.586$.

The effect of data labels was also investigated with further analysis in the Label-Group. The results showed no significant difference between the two conditions for the cases that exhibit *one-to-one* relation, $U=58.00$, $z=-1.37$, $p>.05$ and for the cases that contain *many-to-one* relations, $t(23)=-1.33$, $p>.05$. Whereas same pattern was observed for the *one-to-one* ($t(26)=-.41$, $p>.05$) and the *many-to-one* ($t(26)=.39$, $p>.05$) relations across the visual and the haptic graphs, only the haptic graph explorers produced gestures that accompany to more than one event exhibiting "*one-to-many*" relation, ($U=45.50$, $z=-3.00$, $p<.05$. This indicates that event segmentation on gesture and language modalities may exhibit differences, and these differences seem to be originated from the differences due to sensory modality.

Table 6-27 The mean values for each gesture/sub-event relation

	One-to-one	Many-to-one	One-to-many
VG with labels	M=.98(SD=.04)	M=.28(SD=.23)	No cases
VG without labels	M=.94 (SD=.09)	M=.45 (SD=.36)	No cases
HG without labels	M=.92 (SD=.09)	M=.50 (SD=.30)	M=.16 (SD=.20)

6.4.6.3. Summary For Event Segmentation

The experiment was conducted in three conditions. First group explored the graph haptically, and the haptic graphs did not have data labels on it. The participants in the second group explored the visual graph with labels. Third group explored the visual graphs without labels. The graphs presented to the first and the third groups were informationally equivalent, however, they differed in terms of the perceptual sensory modality. The effect of this perceptual difference on event segmentation was tested by conducting pairwise comparisons (*the modality group*). On the other hand, the visual graphs with labels contain more information w.r.t the visual graphs without labels, therefore although they are perceived through the same modality. The effect of this informational difference on event segmentation was investigated by comparing the results of the second and third groups (*the label group*). First and second groups were not compared and the overall comparison among three groups was also not conducted. The investigation of the individual effect of data labels, of the modality and of the amodal properties in a systematically controlled way is of the main interest of this analysis. Table 6-28 summarizes all the results (*with significance and effect sizes*) reported in this part. As well as summarizing it, this table also provides information for performing overall meta-analysis. For instance, if the main effect of the global shape shows significant main effect in the both data sets[28] regarding a particular parameter, then it can be concluded that this effect is not dependent on the data-set (*or sampling*). This is helpful for making an inference about its generalization. Comparing the effect size is another trend in statistics, and it can be utilized as a useful tool to understand how much the data set has an influence on the dependent parameter.

In this analysis, the positioning rate was employed as an indicator of event segmentation. Overall, it can be concluded that how the graph readers/explorers locate event boundaries (looking at the verbal descriptions) seems not to be affected by the presence of data labels. Additionally, the sensory modality of graph reading also does not have a main effect on the event segmentation. On the other hand, the graphical features, both the perceptual (i.e. the slope and the angle) and the conceptual, exhibited a large effect on the graph segmentation. Although a main of effect of sensory modality was not observed, the results showed that there were significant *interaction effects* of modality regarding both qualitative ascriptions, when we had a closer look at the spatio-temporal ascriptions, we see that this difference was originated from the low positioning rates for trend changes in haptic modality. Together with the interpretation of significant effect of angle within the steep-value category, we can infer that for the haptic explorers a direction change is more prominent compared to other types of landmarks, and trend changes are not preferred as event boundaries as much as the visual graph readers prefer. The results are in line with the claims of event segmentation theory (EST) that suggests that the event segmentation is not a random process. According to EST, both perceptual and conceptual factors take part in event segmentation, and events are segmented on the locations where the salient change occurs, as it is the case with i.e. steep segments or acute angles.

Another finding indicated that the polarity (being *convex* or *concave*) does not play any role in the haptic graph comprehension. But, interestingly, the conceptual polarity (global minima versus global maxima) had an effect. Combining this result with the previous research that indicated that the negative minimum points are preferred more as a segmentation point, one can say that the saliency of the points are not just affected by their perceptual properties, their functions (the conceptual saliency) is another factor that determines whether they are chosen as an event boundary as well, supporting the existing literature (on functional taxonomy in e.g. Tversky, 1990). However, further analyses regarding the polarity issue by systematically controlling it would provide more confirmative results.

---

[28] The fact that the data sets share one common group, namely *visual graph condition with labels*, should be taken into account when the overall comparison is intended.

Regarding the analysis of speech accompanying gestures, the main results can be summarized as follows; the effect of the sensory modality on the number of gestures was not observed. Although this parameter (*the gesture count per stimuli/description* etc.) is one of the most frequent parameter used in the gesture research, it may not be well fit to the comparison between the visual and haptic modalities, since the gestures exhibit differences in form with respect to sensory modality of exploration as it is elaborated on in upcoming sub-section 6.4.7.3. Furthermore, the gesture-sub event relation provided interesting case indicating that event segmentation on gesture and language modalities may exhibit differences depending on the sensory modality of graph reading.

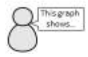
Table 6-28 The significance and the effect sizes for all tests in "Event Segmentation" (CD: condition dependent effects)

Verbal Descriptions			
	Label-Group	**CD**	**Modality-Group**
a. Number of sub-events			
Condition (C)	No. Sig.		No. Sig.
Graph-Segment Count	Sig. (η^2=33)		Sig. (η^2=47)
5-Segment vs 6-Segment	No. Sig.	*	Sig. (η^2=26)
6-Segment vs 7-Segment	Sig. (η^2=32)		Sig. (η^2=38)
5-Segment vs 7-Segment	Sig. (η^2=44)		Sig. (η^2=61)
Graph Pattern	Sig. (η^2=22)		Sig. (η^2=26)
Graph Pattern x C	No. Sig.		No. Sig.
Graph Pattern x 5-Segment	Sig. (η^2=17)		Sig. (η^2=23)
Graph Pattern x 6-Segment	Sig. (η^2=20)		Sig. (η^2=12)
Graph Pattern x 7-Segment	No. Sig.	*	Sig. (η^2=18)
b. Positioning rates as an event boundaries			
Condition (C)	No. Sig.		No. Sig.
Slope	Sig. (η^2=48)		Sig. (η^2=70)
Slope x Condition	Sig. (η^2=15)		Sig. (η^2=14)
Angle	Sig. (η^2=74)		Sig. (η^2=64)
Angle x Condition	Sig. (η^2=17)	*	No. Sig.
Polarity	No. Sig.		No. Sig.
Polarity x Angle	Sig. (η^2=21)	*	No. Sig.
Angle x Steep Slope	No. Sig.	*	Sig. (η^2=14)
Angle x Medium Slope	Sig. (η^2=52)		Sig. (η^2=29)
Angle x Slight Slope	Sig. (η^2=56)		Sig. (η^2=49)
Angle x No Slope	Sig. (η^2=67)		Sig. (η^2=49)
Angle x Steep Slope x C	No. Sig.	*	Sig. (η^2=15)
Angle x Medium Slope x C	Sig. (η^2=17)	*	No. Sig.
Angle x Slight Slope x C	No. Sig.		No. Sig.
Angle x No Slope x C	No. Sig.		
Spatio-temporal ascriptions	Sig. (η^2=28)		Sig. (η^2=47)
Spatio-temporal ascriptions x C	No. Sig.	*	Sig. (η^2=19)
Conceptual ascriptions	Sig. (η^2=78)		Sig. (η^2=68)
Conceptual ascriptions x C	Sig. (η^2=13)		Sig. (η^2=13)
Minimum versus Maximum	No. Sig.	*	Sig. (η^2=31) for global pt.
Speech-Accompanying Gestures			
	Label-Group		**Modality-Group**
# of representational gestures			
Condition (C)	No. Sig.		No. Sig.
Graph Pattern	No. Sig.		No. Sig.
One-to-one Relation	No. Sig.		No. Sig.
Many-to-one Relation	No. Sig.		No. Sig.
One-to-Many Relation	NA.		Sig. (z=-3.00)

6.4.7. Event Description

In addition to how graph readers segment the events concerning partonomic relations, how they refer to these segmented events, namely the taxonomic relations, is also important both for understanding haptic graph comprehension and for deciding the content provided by the verbal assistance system. This part addresses the investigation of event descriptions by analyzing segment-based reference rates, the use of modifiers and error rates in verbal descriptions and finally the speech accompanying gestures are discussed.

6.4.7.1.Reference To Segments



As densely elaborated in the literature section, the partonomic and taxonomic relations are highly intertwined, and have an influence on each other. Therefore although this section is dedicated to the topics concerning *event description* (mostly addressing taxonomic relations), it inherently contains investigation of the partonomic structures as an extension to previous sub-section as well.

As shown in Figure 6-33, the participants may prefer to refer segments in different temporal granularities. A description of a sub-event may refer to one individual line segment on the graph or it may also refer to the broader content by containing reference to two or more line segments. Therefore, the segmentation can be observed in different temporal scales in the partonomic level. In this analysis, two levels of temporal granularity were used; (i) the fine-grained (one or two segments) and (ii) the coarse-grained (three or more segments). The comparisons were conducted on the average amount of the sub-events for each category.

Figure 6-33 (a) gives an example of the case where the graph readers prefer to use the fine-grained temporal units, while the participants who explore the graph in Figure 6-33(b) prefer to use the coarse-grained units.

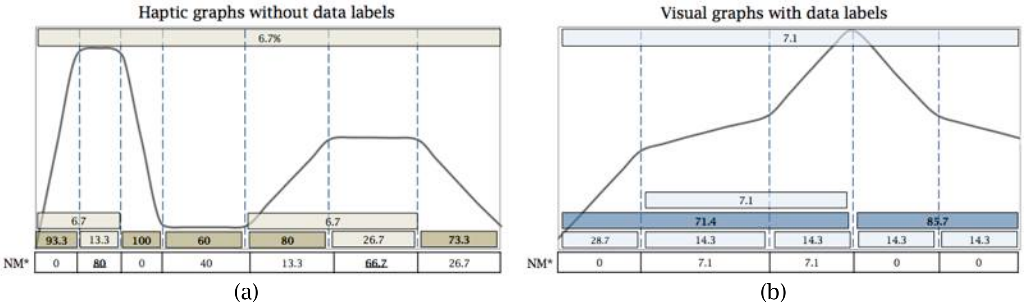


Figure 6-33 Illustrations for different temporal granularities (NM: none-mentioning rate)

The analyses reported thus far addressed the parameters regarding full event description. In this section, I focused on the investigation of the factors that have a role in sub-event referring. The effects of the presence of data labels, of the sensory modality and also of the amodal graphical properties on reference production was in the focus of this part. In this analyses, users' reference rate for each segment or segment complex as explained previously for each graph in each condition was used.

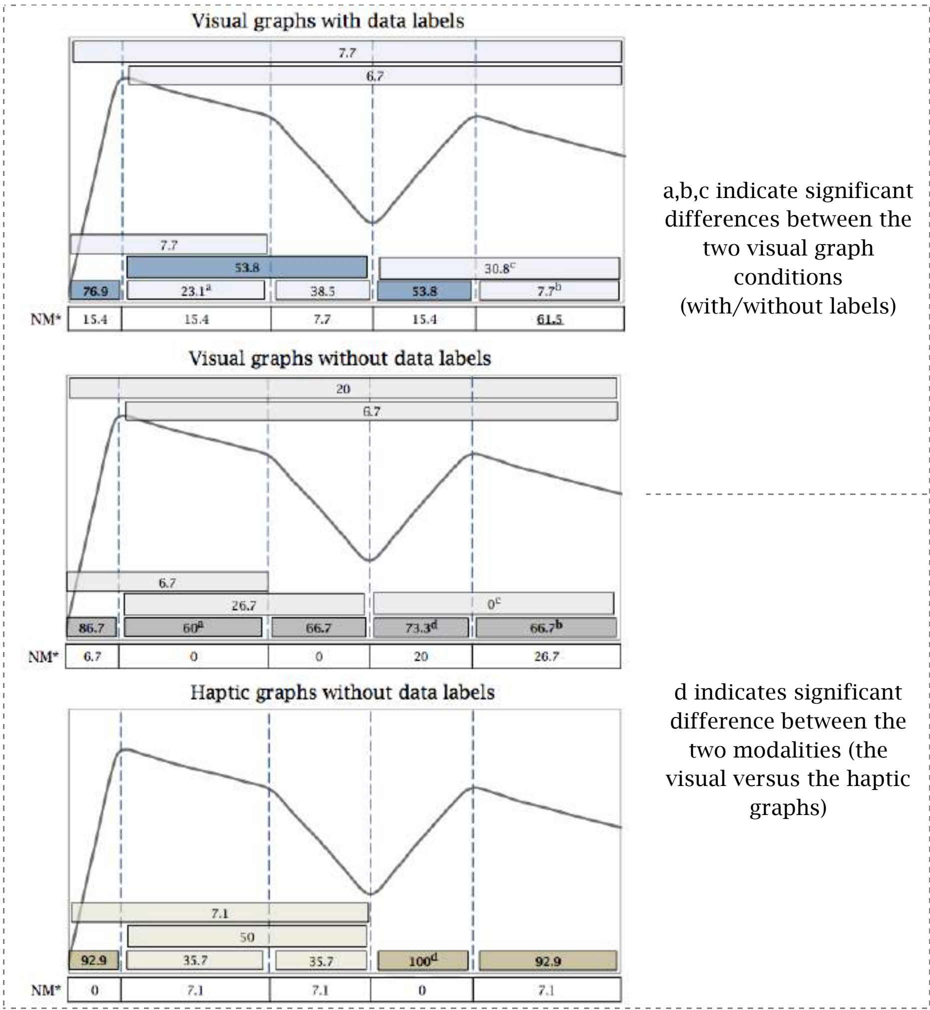
A. The Choice of Reference-Scope

In this section, the participants' way of referring to the shape segments were analyzed. *The choice of reference scope*[29] was categorized into four. First, a graph reader may prefer not to refer the segment at all ("*no-mention*"), second s/he may prefer to refer to the particular segment explicitly and individually ("*separate-mentioning*"), third s/he may

[29] I used this term as kindred to concept of "Reference Scope", but here I preferred to narrow it down into four distinct categories for the sake of the analysis, and names as "the choice of reference scope".

prefer to mention the segment in a combination with other segments ("*combined-mentioning*"), or s/he may refer to the segment separately and then in a combination with other segments ("*both separate & combined*") in his/her verbal description. As shown in the illustration, the size of event segments (*the fine-grained units* or *the coarse-grained units*) and also the referring rates for each unit may exhibit difference between the conditions. Appendix F presents the overall referring rates for each graph in each condition. I hypothesized that the choice of reference-scope is affected by the saliency of the amodal graphical entities. In specific, the salient segments are expected to be referred as separately, and the less salient segments are expected to be combined with other segments in the verbal description. The upcoming sub-sections contain detailed comparative statistical analysis, for a listwise summary, readers are encouraged to take a glance at summary table (Table 6-50) that contain only significance values and effect sizes.

Table 6-29 An illustration of segmentation patterns for each condition and the percentages for each segment, NM: none-mentioned



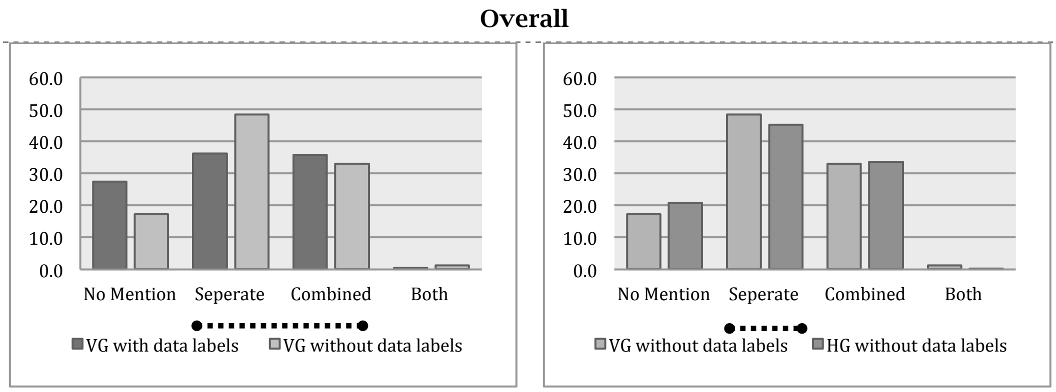
A.1. Overall Analysis

Effect of Data Labels. The findings indicated that there was a significant main effect of the reference-scope choice $F(1.57, 47.16)=40.41$, $p<.001$, $\eta^2=.57$. There was no significant interaction between the choice of scope and the condition. Additional pairwise comparisons (as illustrated and summarized in Table 6-30 and Figure 6-34) indicated that the "separate-mentioning" and the "combined-mentioning" were preferred more and there was no significant difference between their reference rates, $p>.05$. The reference rate for the separately mentioned segments were significantly higher than the none-mentioned segments ($F(1, 30)=12.42$, $p<.01$, $\eta^2=.29$) and higher than the "both separate and combined" type $F(1, 30)=149.91$, $p<.001$, $\eta^2=.83$. Similarly, the reference rate for the "combined-mentioning" segments were significantly higher than the none-mentioned segments ($F(1, 30)=12.64$, $p<.01$, $\eta^2=.30$) and higher than the "both separate and combined" type $F(1, 30)=179.25$, $p<.001$, $\eta^2=.86$. Finally, the rate for the "no-mention" segments was significantly higher than the "both separate and combined", $F(1, 30)=66.80$, $p<.001$, $\eta^2=.69$.

Effect of Sensory Modality. The findings indicated that there was a significant main effect of the choice of reference-scope $F(1.53, 45.89)=42.90$, $p<.001$, $\eta^2=.59$. There was no significant interaction between the choice of reference scope and the condition. Additional pairwise comparisons indicated that the "separate-mentioning" were preferred more. The reference rate for the "separately mentioned" segments were significantly higher than the "none-mentioned" segments ($F(1, 30)=21.93$, $p<.001$, $\eta^2=.42$), higher than the "combined-mentioning" segment ($F(1, 30)=5.42$, $p<.05$, $\eta^2=.15$) and higher than the "both separate and combined" type $F(1, 30)=157.01$, $p<.001$, $\eta^2=.84$. Similarly, the reference rate for the "combined-mentioning" segments were significantly higher than the "none-mentioned" segments ($F(1, 30)=17.14$, $p<.01$, $\eta^2=.36$) and higher than the "both separate and combined" type $F(1, 30)=146.58$, $p<.001$, $\eta^2=.83$. Finally, the rate for "no-mention" segments was significantly higher than the "both separate and combined", $F(1, 30)=47.87$, $p<.001$, $\eta^2=.62$.

Table 6-30 Referring types for each condition

	No Mention	Separate	Combined	Both separate and combined
VG with labels	27.5	36.3	35.8	0.5
VG without labels	17.3	48.5	33.0	1.2
HG without labels	20.8	45.3	33.6	0.3



(a) (b)

Figure 6-34 The reference scope choices for each condition

A.2. Effect of slope on the choice of reference scope

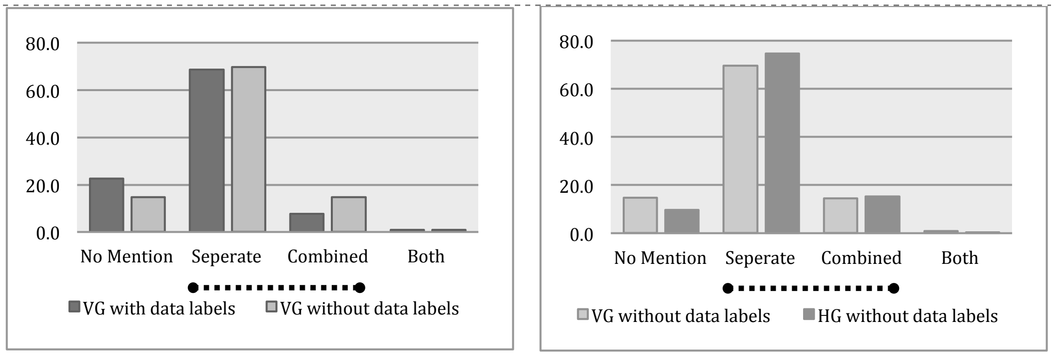
Additional analyses were conducted to investigate whether the participants prefer one reference-scope choice over others in their sub-event descriptions within the each slope-value category. The choice of reference scope is a categorical variable, thus each reference can fit into only one category. Here, I looked at whether there is a difference in the distribution patterns.

A.2.1. Reference Scope for Steep Segments

Effect of Data labels. The results revealed that there was a significant difference in the rates for the different reference-scope types with a large effect size $F(1.53, 46.16)=89.55$, $p<.001$, $\eta^2=.75$, see Figure 6-35. The follow-up pairwise comparisons indicated that the "separate-mentioning" was the most preferred way for the steep segments, and the rates for this category was significantly higher (with very large effect sizes) than the "combined-mentioning" $F(1, 30)=98.29$, $p<.001$, $\eta^2=.77$, higher than the "no-mention" category $F(1, 30)=58.92$, $p<.001$, $\eta^2=.66$ and also higher than the "both separate and combined" category $F(1, 30)=286.91$, $p<.001$, $\eta^2=.91$. The "both separate and combined" was the least preferred way of referring for the steep segments. Its reference rate was significantly lower than the "none-mention" category $F(1, 45)=30.18$ $p<.001$, $\eta^2=.50$, and lower than the "combined-mentioning" category $F(1, 30)=16.72$, $p<.001$, $\eta^2=.36$. The difference in the reference rates for the "none-mention" and the "combined-mentioning" was also significant with a small effect size, $F(1, 30)=4.42$, $p<.05$, $\eta^2=.13$. On the other hand, the interaction between the condition and the reference scope was not significant, $p>.05$.

Effect of Sensory Modality. There was a significant difference in the reference rates for different reference-scope types $F(1.61, 48.26)=91.56$, $p<.001$, $\eta^2=.75$. As shown in Figure 6-35, the follow-up pairwise comparisons indicated that the "separate-mentioning" was the most preferred way for the steep segments, and the rates for this category was significantly higher (with very large effect sizes) than the "combined-mentioning" $F(1, 30)=91.82$, $p<.001$, $\eta^2=.75$, higher than the "no-mention" category $F(1, 30)=76.86$, $p<.001$, $\eta^2=.72$, and also higher than the "both separate and combined" category $F(1, 30)=286.22$, $p<.001$, $\eta^2=.91$. Similar to the case in the label-group, the "both separate and combined" was the least preferred way of referring for the steep segments in the modality-group. Its rate was significantly lower than the "none-mention" category $F(1, 30)=12.35$ $p<.001$, $\eta^2=.29$, lower than the "combined-mentioning" category $F(1, 30)=31.11$, $p<.001$, $\eta^2=.51$. However, there was no significant difference in the reference rates for the "none-mention" and the "combined-mentioning", $p>.05$. Furthermore, the interaction between the condition and the reference scope was not significant, $p>.05$.

Steep Segments



(a) (b)

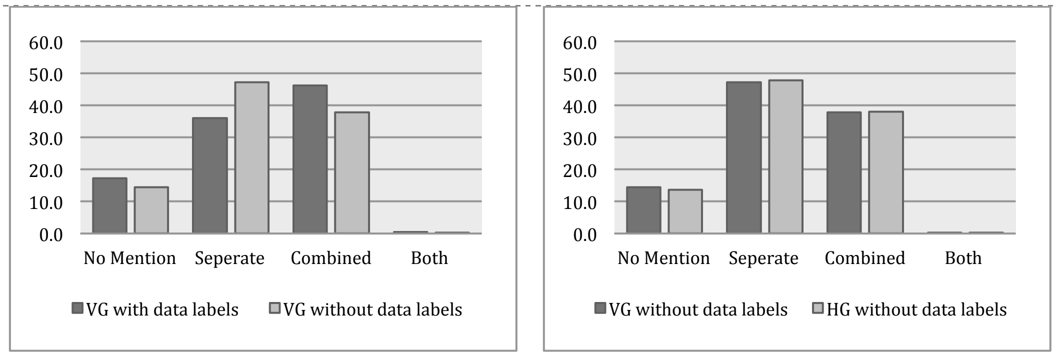
Figure 6-35 The choices of reference-scope for the steep segments

A.2.2 Reference scope for Medium Segments.

Effect of Data labels. Regarding the medium segments, there was a significant difference with a large effect size in the rates for the different reference-scope types $F(1.54, 46.18)=38.76$, $p<.001$, $\eta^2=.56$. As also illustrated in Figure 6-36, the additional pairwise tests revealed that the participants mostly preferred two ways of referring ("the separately-mentioned" or "in a combination" expressions). The difference between these two categories was not significant, $p>.05$. The reference rates for "the no-mention" category was in general very low and was significantly lower than "the separate-mentioning", $F(1, 30)=18.31$, $p<.001$, $\eta^2=.38$, and lower than "the combined-mentioning", $F(1, 30)=44.96$, $p<.001$, $\eta^2=.60$. "The both separate and combined" was the least preferred way of referring for the medium segments (significantly lower than the "no-mention" category $F(1, 30)=36.46$ $p<.001$, $\eta^2=.55$, lower than the "separately-mentioned" $F(1, 30)=103.68$, $p<.001$, $\eta^2=.78$, and lower than the "combined-mentioning" category. F(1, 30)=178.95, $p<.001$, $\eta^2=.86$. Furthermore, the interaction between the condition and the reference scope was not significant, $p>.05$.

Effect of Sensory Modality. There was a significant difference in the rates for different reference-scope types $F(1.54, 46.32)=38.96$, $p<.001$, $\eta^2=.57$. The additional pairwise tests revealed that the participants mostly preferred two ways of referring ("the separately-mentioned" or "in a combination" expressions) in referring to the medium segments; the difference between these two categories was not significant, $p>.05$. The reference rates for "No-mention" category was in general very low and was significantly lower than "the separate-mentioning", $F(1, 30)=27.58$, $p<.001$, $\eta^2=.48$, and "combined-mentioning", $F(1, 30)=33.35$, $p<.001$, $\eta^2=.53$. The "both separate and combined" was the least preferred way of referring for the medium segments. Its rate was significantly lower than the "no-mention" $F(1, 30)=25.81$ $p<.001$, $\eta^2=.46$, lower than "the separately-mentioned" $F(1, 30)=123.48$, $p<.001$, $\eta^2=.81$ and also lower than "the combined-mentioning" category. $F(1, 30)=130.47$, $p<.001$, $\eta^2=.81$. Furthermore, the interaction between the condition and the reference scope was not significant, $p>.05$.

Medium Segments



(a) (b)

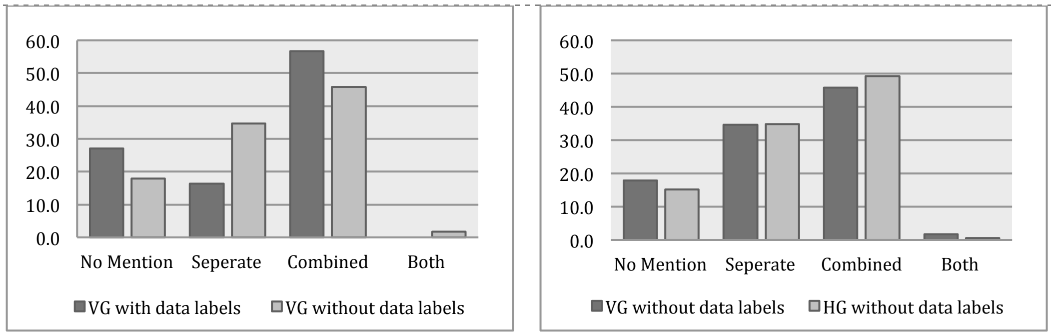
Figure 6-36 The choices of reference-scope for the medium segments

A.2.3 Reference Scope for Slight Segments.

Effect of Data labels. Similar to the results concerning steep and medium segments, there was a significant difference with a large effect size in the rates for different reference-scope types concerning slight segments $F(1.88, 56.57)=37.48$, $p<.001$, $\eta^2=.56$. The additional pairwise tests revealed that in referring to the slight segments, the participants tended to refer the slight segments as a part of a bigger combination instead of referring it as a separate entity $F(1, 30)=18.75$, $p<.001$, $\eta^2=.39$, or as not referring at all ($F(1, 30)=32.91$, $p<.001$, $\eta^2=.52$) or as the "both separate and combined" $F(1, 30)=49.03$, $p<.001$, $\eta^2=.62$. There was no significant difference in the reference rates for "none-mention" and the "separate-mentioning" categories, $p>.05$. Again, the "both separate and combined" was the least preferred way of referring and significantly lower than the "none-mention" category $F(1, 30)=41.17$, $p<.001$, $\eta^2=.58$ and lower than the "combined-mentioning" $F(1, 30)=224.43$, $p<.001$, $\eta^2=.88$. Furthermore, the interaction between the condition and the reference scope was significant with a small effect size, $F(1.88, 56.57)=5.49$, $p<.05$, $\eta^2=.16$.

Effect of Sensory Modality. Similarly, the findings also revealed there was a significant difference with a large effect size in the rates for different reference-scope types $F(1.78, 53.48)=29.70$, $p<.001$, $\eta^2=.50$. As illustrated in Figure 6-37 participants preferred two ways of referring ("separately-mentioned" or "in a combination" expressions) in referring to the slight segments; the difference between these two categories was not significant, $p>.05$. The reference rate for the "combined-mentioning" choice was significantly higher than that for the "no-mention" category $F(1, 30)=36.54$, $p<.001$, $\eta^2=.55$ and also than that for the "both type" $F(1, 30)=152.02$, $p<.001$, $\eta^2=.84$. Similarly, the "separate-mentioning" was also preferred more than the "no-mention" $F(1, 30)=6.81$, $p<.05$, $\eta^2=.19$, and more than the “both types” $F(1, 30)=66.92$, $p<.001$, $\eta^2=.69$. Again, “both separate and combined” was the least preferred way of referring and significantly lower than “none-mention” $F(1, 30)=20.03$, $p<.001$, $\eta^2=.40$. Furthermore, the interaction between the condition and the reference scope was not significant, $p>.05$.

Slight Segments



(a) (b)

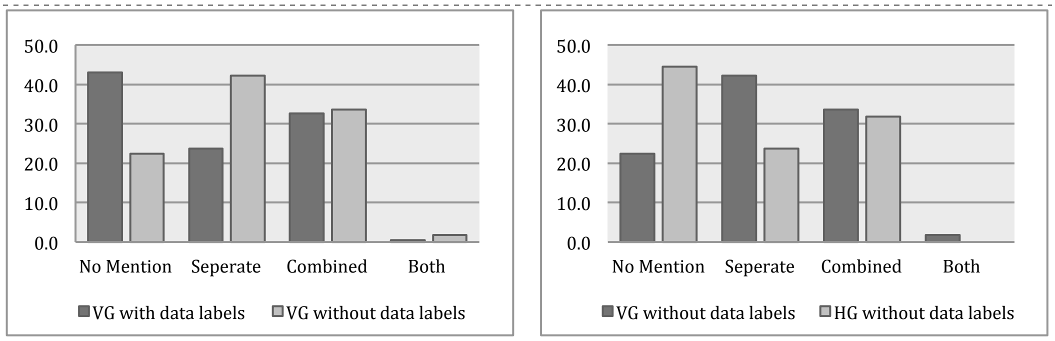
Figure 6-37 The rates of referring-type for slight segments

A.2.4 Reference scope for Horizontal Segments.

Effect of Data labels. Similar to the result of the steep and the medium segments, there was a significant difference with a large effect size in the rates for different reference-scope types $F(1.62, 48.74)=24.72$, $p<.001$, $\eta^2=.45$. When referring to the horizontal segments, there was no one favorite method. In both conditions, “none-mention”, “separate-mentioning”, and “combined-mentioning” were preferred similarly. As shown in Figure 6-38, the interaction between the condition and the reference scope was significant, $F(1.62, 48.74)=6.30$, $p<.01$, $\eta^2=.17$. Similar to the other steepness values, the “both separate and combined” category was the least preferred one and significantly lower than “none-mention” $F(1, 30)=99.80$ $p<.001$, $\eta^2=.78$, and lower than the “separate-mentioning” $F(1, 30)=73.20$, $p<.001$, $\eta^2=.71$ and also lower than the “combined-mentioning”, $F(1, 30)=132.98$, $p<.001$, $\eta^2=.82$.

Effect of Sensory Modality. Similarly, the findings also revealed that there was a significant difference in the rates for different reference-scope types $F(1.65, 49.58)=26.63$, $p<.001$, $\eta^2=.47$. When referring to the horizontal segments, there was no one preferred method. In all three conditions, “none-mention”, “separate-mentioning”, and “combined-mentioning” were preferred similarly, however their distribution differed w.r.t sensory modality, see Figure 6-38. The interaction between the condition and the reference scope was significant, $F(1.65, 49.58)=7.28$, $p<.01$, $\eta^2=.20$. Similar to the other slope values, the “both separate and combined” category was the least preferred one and significantly lower than the “none-mention” $F(1, 30)=137.31$ $p<.001$, $\eta^2=.82$, lower than the “separate-mentioning” $F(1, 30)=77.25$, $p<.001$, $\eta^2=.72$ and also lower than the “combined-mentioning”, $F(1, 30)=115.99$, $p<.001$, $\eta^2=.80$.

Horizontal Segments



(a) (b)

Figure 6-38 The rates of referring-type for the horizontal segments

A.3. Effect of angle on the choice of reference scope

Following the same method employed in the investigation of event boundary's location, the effect of angle was also taken into consideration by performing individual tests within each steepness value. Here, I present the results in tables and only highlight the main effects and significant differences for ease of reading.

A.3.1 A1-Value (Acute) Angles: 60° & 30°). The participants preferred to refer to the segments with the A2 angle connections as a separate segment in all conditions.

Table 6-31 The significance values and the effect sizes for the A1 angle-value tests

	Label-Group	Modality-Group
The choice of Reference-Scope	$p<.001$, $\eta^2=.49$	$p<.001$, $\eta^2=.48$
"No-Mention" versus "Separate"	$p<.001$, $\eta^2=.35$	$p<.001$, $\eta^2=.40$
Separate" versus "Combined"	$p<.01$, $\eta^2=.32$	$p<.001$, $\eta^2=.40$
"Combined" versus "Both"	$p<.001$, $\eta^2=.57$	$p<.001$, $\eta^2=.53$
Separate" versus "Both"	$p<.001$, $\eta^2=.82$	$p<.001$, $\eta^2=.82$
"No-Mention" versus "Both"	$p<.001$, $\eta^2=.54$	$p<.01$, $\eta^2=.32$
"No-Mention" versus "Combined"	Non Sig.	Non Sig.
Scope X Condition	Non Sig.	Non Sig.

A1-Value (Acute) Angles: 60° & 30°

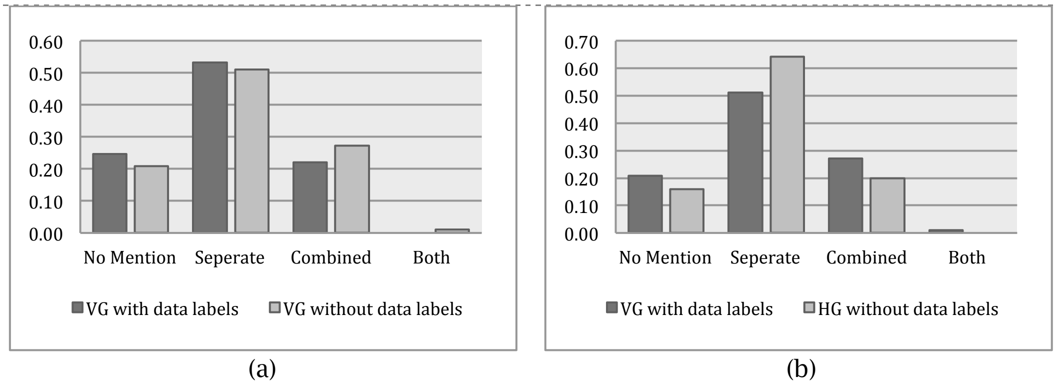


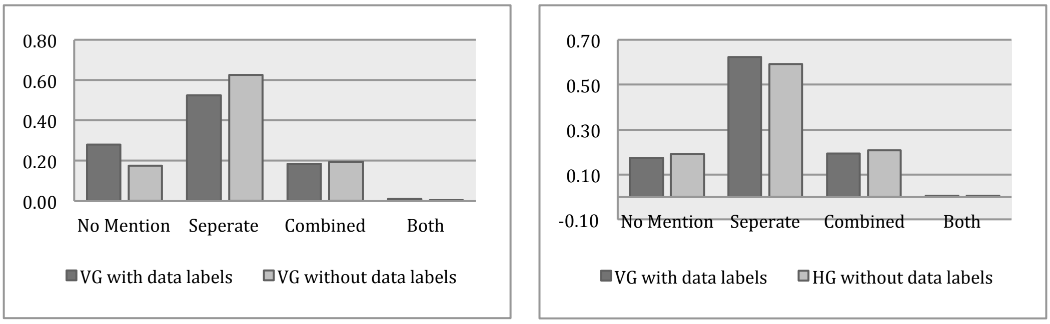
Figure 6-39 the rates of referring-type for A1-value angles

A.3.2 A2-value Angles: 90° & 105° ((nearly) Right angles. The results also indicated that the participants preferred to refer to the segments with the A2 angle connections as separate segments.

Table 6-32 Significance values and effect sizes for the A2 angle-value tests

	Label-Group	Modality-Group
The choice of Reference-Scope	$p<.001$, $\eta^2=.63$	$p<.001$, $\eta^2=.66$
"No-Mention" versus "Separate"	$p<.001$, $\eta^2=.45$	$p<.001$, $\eta^2=.59$
Separate" versus "Combined"	$p<.001$, $\eta^2=.59$	$p<.001$, $\eta^2=.56$
"Combined" versus "Both"	$p<.001$, $\eta^2=.68$	$p<.001$, $\eta^2=.60$
Separate" versus "Both"	$p<.001$, $\eta^2=.89$	$p<.001$, $\eta^2=.87$
"No-Mention" versus "Both"	$p<.001$, $\eta^2=.58$	$p<.001$, $\eta^2=.59$
"No-Mention" versus "Combined"	Non Sig.	Non Sig.
Scope X Condition	Non Sig.	Non Sig.

A2-value Angles: 90˚ & 105˚ ((nearly) Right angles



(a) (b)

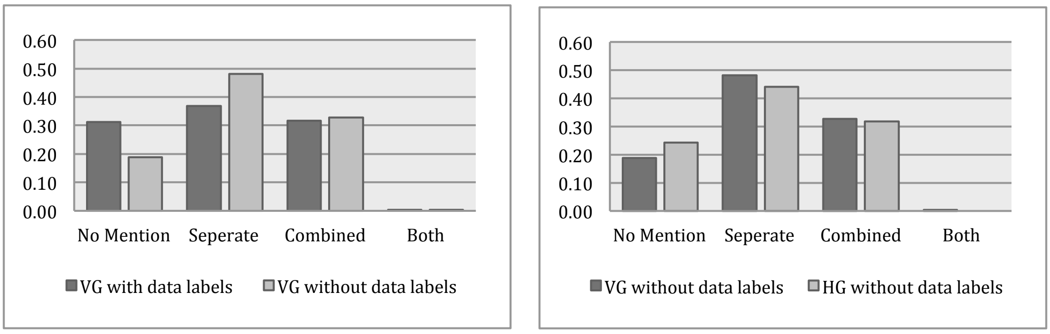
Figure 6-40 The rates of referring type for the A2-value angles

A.3.4 A3-value Angles: 120˚, 135˚. The results also indicated that the participants preferred to refer to the segments with the A3 angle connections either as a separate segment or as a parts of bigger segment combinations. Table 6-33 presents the significance values and the effect sizes for the tests concerning the A3-value angles.

Table 6-34 The significance values and the effect sizes for the A3 Angle-Value tests

	Label-Group	**Modality-Group**
The choice of Reference-Scope	$p<.001$, $\eta^2=.46$	$p<.001$, $\eta^2=.48$
“No-Mention” versus “Separate”	$p<.05$, $\eta^2=.20$	$p<.01$, $\eta^2=.30$
Separate” versus “Combined”	Non Sig.	Non Sig.
“Combined” versus “Both”	$p<.001$, $\eta^2=.75$	$p<.001$, $\eta^2=.75$
Separate” versus “Both”	$p<.001$, $\eta^2=.77$	$p<.001$, $\eta^2=.78$
“No-Mention” versus “Both”	$p<.001$, $\eta^2=.69$	$p<.001$, $\eta^2=.63$
“No-Mention” versus “Combined”	Non Sig.	$p<.05$, $\eta^2=.16$
Scope X Condition	Non Sig.	Non Sig.

A3-value Angles: 120˚, 135˚



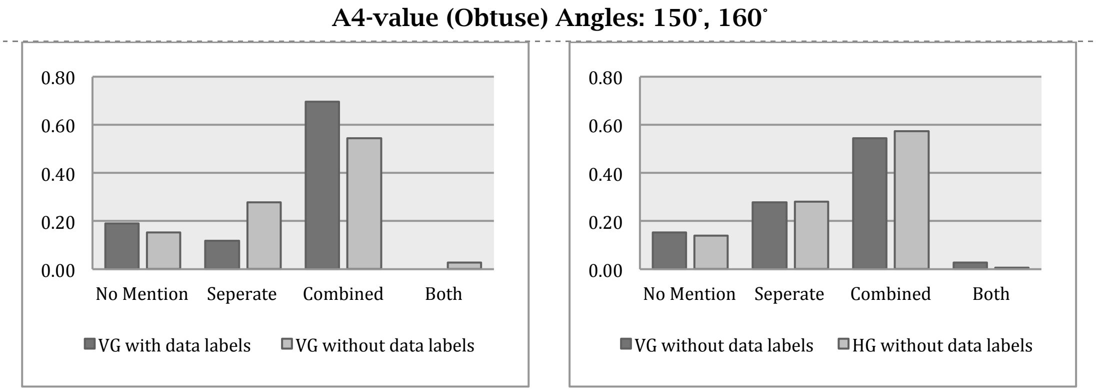
(a) (b)

Figure 6-41 The rates of referring type for theA3-value angles

A.3.3 A4-value (Obtuse) Angles: 150˚, 160˚. The participants preferred to refer to the segments with obtuse angle connections as a part of bigger segment combinations, than other types, see Table 6-35 and Figure 6-42.

Table 6-35 The significance values and the effect sizes for the statistical tests (A4-Value)

	Label-Group	**Modality-Group**
The choice of Reference-Scope	$p<.001$, $\eta^2=.69$	$p<.001$, $\eta^2=.61$
“No-Mention” versus “Separate”	Non Sig.	$p<.05$, $\eta^2=.17$
Separate” versus “Combined”	$p<.001$, $\eta^2=.60$	$p<.001$, $\eta^2=.35$
“Combined” versus “Both”	$p<.001$, $\eta^2=.89$	$p<.001$, $\eta^2=.85$
Separate” versus “Both”	$p<.001$, $\eta^2=.54$	$p<.001$, $\eta^2=.69$
“No-Mention” versus “Both”	$p<.001$, $\eta^2=.47$	$p<.001$, $\eta^2=.38$
“No-Mention” versus “Combined”	$p<.001$, $\eta^2=.73$	$p<.001$, $\eta^2=.68$
Scope X Condition	$p<.05$, $\eta^2=.12$	Non Sig.



(a) (b)

Figure 6-42 The rates of referring-type for A4-value angles

A.4. Angle and Slope Interaction on the reference scope

In addition to the investigations of each value of the slope and angle parameters, their interaction effect on the event description is also regarded as important, since these properties co-occur in graphical representations and are dependent on each other. Same method, as utilized in the investigation of the event boundary location, was applied. The effect of angle within the each slope-value for all condition were calculated and compared statistically. Here, Figure 6-43 depicts the color-coded representation of most favorite choices of reference scopes. Each cell represents statistically most preferred type for a particular angle within a particular slope value. The interpretation of the results was included in the interim summary in the following part.

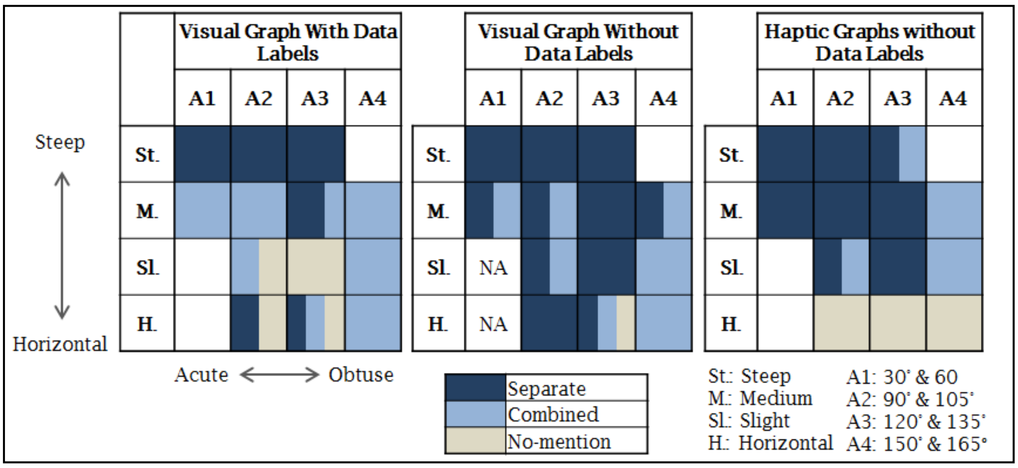


Figure 6-43 Angle x Slope interaction for the choice of reference scope

A.5. Interim Summary

Overall analyses indicated that both in the Label-Group and also in the Modality-Group, significant differences among the reference scope preferences were found with similar and large effect sizes. In general, in the group where the participants explore visual graphs (with/without labels), the segments were referred separately or in a combination with other segments forming line complexes. On the other hand, in the modality-group, the participants' first choice was the "separate mentioning", and the rate for this category was significantly higher than the other types. A closer look at the mean distributions showed that this difference was not because the haptic explorers chose "separately mentioned segments" more than the visual graph readers (in fact there was no difference), it is because the participants who explored the visual graphs with labels also preferred the "combined" category as well as the "separate mentioning", see Figure 6-34. Furthermore, the "none-mention" category was also observed although its rate was not as high as other two types. There was a consisting finding that indicates that referring to a segment in both separate and combined way in the same description was extremely rare.

As has been investigated for the location of the event boundaries, the effects of the slope and the angle were also tested in the investigation of the reference scope. When referring to the steep segments, in all conditions, the participants preferred to refer to them separately. The rate for the combined mentioning and the none-mentioning were low (around 18%). When referring to the medium segments, the participant again in all conditions exhibited similar pattern. They in general either prefer to refer them separately or in a combination. But preferring not to mention them at all was low. The opposite pattern, of which was observed for the reference of the steep segments, was observed for the slight segments. While the participants in the modality group refer usually either as separate or combined, the participants in the label group mostly preferred the combined type. It should be noted that data labels seem to facilitate grouping of the slight segments. Furthermore, while the reference rate for the "none-mention" category was similar to the "separate-mentioning" rate in the Label-Group, for the modality group, the rate of the none-mentioned segments was still low.

In the modality group, the rates for the steep, medium and slight segments went hand-in-hand across both modalities, while in the label-group, the differences due to the presence of data labels start to become apparent from steep to slight segments. And for the horizontal segments, in both the data label and the modality groups exhibited different patterns across conditions, see Figure 6-35 to Figure 6-38 for more details. Finally, the horizontal segments were the segments with less consistency. First of all, there was no one favorite way of referring for the horizontal segments in overall. But closer look at the mean distributions shows that the participants who explore the visual graphs with labels prefer not to mention them at all, the haptic explorers exhibited the same pattern. On the other hand, the participants who explore the visual graphs without data label preferred to refer to them separately.

Same analysis was also applied to different angle values. The segments with the A1 (*the acute angles*) and the A2 (*the right or almost right angles*) angle value connections were referred separately. The A3 values were referred as either separate or combined. But especially the participants in the visual condition with labels preferred not to mention them at all. The segments with the A4-value connection (*the obtuse angles*) were referred as in a combined way in the modality group showing no significant difference across groups. The participants in the label-group also preferred the combined way; however the pattern across conditions was different indicating that the data labels have an effect on referring to the segments with the A4-value angles.

Final analysis involved the investigation of the interaction between the slope and the angle across conditions. Figure 6-43 illustrates the color-coded visualizations of most preferred types of reference scopes for all conditions. The slope seems to be more discriminative than the angle for both visual groups, while the haptic group exhibits more homogenous distribution. In general, results suggest that the data labels facilitate the combinability by providing solid landmarks (such as *in January*, or *between April and June*) that do not require additional explanation, in other words, the expressions with explicit references rarely display ambiguities. On the other hand, in the haptic condition, two issues needs to be touched upon. First, in general, the participants preferred to use separate references for the segments, in other words, they produce fine-grained descriptions in the partonomic level. On the contrary, in the visual graphs, it seems that this decision was not just based on the segment`s slope or on the angle solely. In particular, the presence of data labels yields more coarse-grained descriptions. The results also provided support for the existing literature that states that the fine-grained units are more perceptually determined and they usually address actions. When the event was perceived through the actions as it happened in the haptic exploration of a graph, the participants tended to produce more fine-grained descriptions. Second issue is about the reference to the horizontal segments. In the visual conditions, depending on their connections with other segments, they may visually seem different and w.r.t that, different way of referring can be preferred. On the other hand, in the haptic condition regardless of the angle at the intersection points with the previous segment, they were usually ignored. The horizontal segments refer to *no-change* in the value represented in the y-axis, while time passes. Haptically the participants explore no-change w.r.t y-axis, and the results already showed that a direction change w.r.t the y-axis is more salient for the haptic users. Although we, as a graph reader, know that there is always a change in the temporal dimension, we can interpret that segment as *no-change* region by referring to change in the y-axis. But in the haptic condition, it seems that these regions are not found important to be mentioned in verbal descriptions, this means that the connection between the conceptual invariance and the perceptual invariance is not well established as opposed to visual counterpart. This may be due to immatureness of the graph schemata triggered by the haptic modality.

A successful segmentation does not mean that all individual differences should be emphasized in the verbal description by the verbal assistance system. Instead, it should be conducted in a way that it facilitates online comprehension and also later recall. Therefore, dividing the graphs into meaningful units by taking hard-to-notice differences, *if they are conceptually relevant*, into consideration and also reducing unnecessary information, which are easy-to-notice, should be the goal of the verbal assistance system. These results provide valuable heuristics in that respect.

B. Use Of Modifiers For Separately-Mentioned Segments

Here I focused on the use of modifiers for the "separately-mentioned" line segments which were introduced in the previous analysis. It should be noted that the descriptions for the horizontal segments do not show much variety (mostly referred like "*it continues*", "*stay around*" or "*remain stable*" without having a modifier). Hence the uses of modifiers for the slight, medium and steep segments were investigated in this analysis. To illustrate, a sub-event can be referred (i) with a type-attribute without using any modifier that describes the slope of the segment or (ii) both with type and additional attribute(s) that modify the segment.

(i)	**Without modifier**	"It decreases" "There is a decrease"
(ii)	**With modifier**	"It decreases <u>fast</u>" "There is a <u>steep</u> decrease"

In order to test whether there is an association between the uses of modifiers for each slope value, a series of Pearson's Chi-Square tests for main effects and also the pairwise comparisons were performed.

Effect of Data Labels. Overall, the use of modifiers for the separately-mentioned segments differed w.r.t whether the visual graph had data labels or not, $\chi2(1)=36.53$, $p<.01$, $N=704$, see Figure 6-44a. While the participants preferred to refer to the segments of a graph that has data labels without using a modifier, the opposite of this pattern was observed for the visual graphs without labels. Subsequently, the use of modifiers was investigated within the each slope-value category. The findings indicated that the use of modifiers showed significant differences across the two visual conditions for the slight segments ($\chi2(1)=4.27$, $p<.05$, $N=188$), for the medium segments ($\chi2(1)=15.77$, $p<.01$, $N=243$) and for also the steep segments ($\chi2(1)=15.26$, $p<.01$, $N=273$). When we look at the interactions between the slope-values and the condition reported above, the main interaction seem to be originated from the differences found for the steep and medium segments.

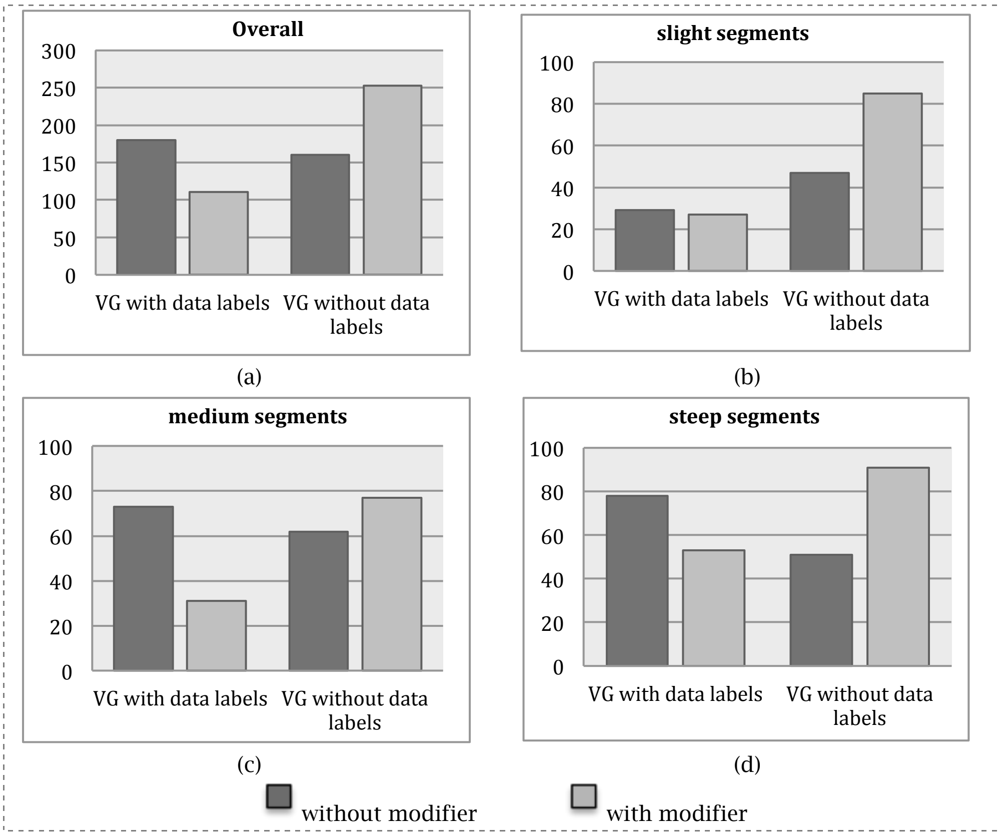
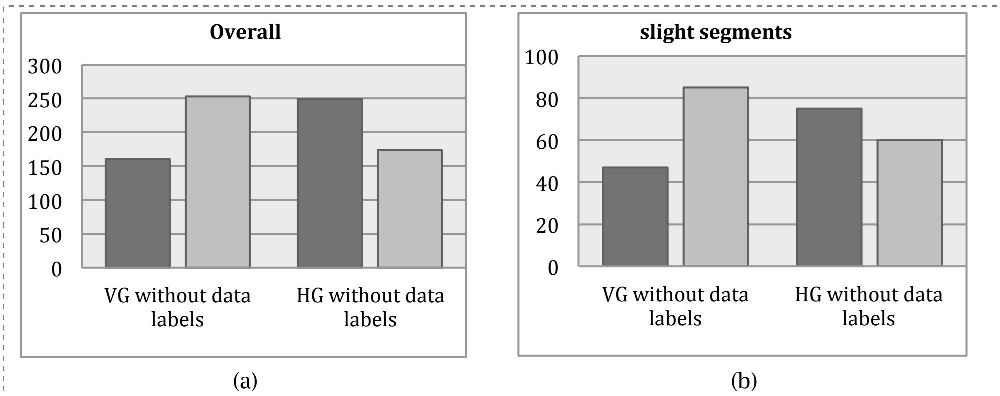


Figure 6-44 The reference rates for the each slope category

Effect of Sensory Modality. Overall, the use of modifiers for the separately-mentioned segments also differed between the two modalities (the visual versus haptic graphs) $\chi 2(1)=33.87$, $p<.01$, N=836, see Figure 6-45. While the participants preferred to refer to the segments of visual graphs by using modifiers, the participants in the haptic condition exhibited opposite pattern. Additional tests also revealed that the use of modifiers showed significant differences across two modalities for the slight segments (($\chi 2(1)=10.70$, $p<.01$, N=267)), for the medium segments (($\chi 2(1)=10.78$, $p<.01$, N=286) and also for the steep segments ($\chi 2(1)=12.39$, $p<.01$, N=283). The patterns in interactions between the modality and the slope value were consistent for all three slope-values.



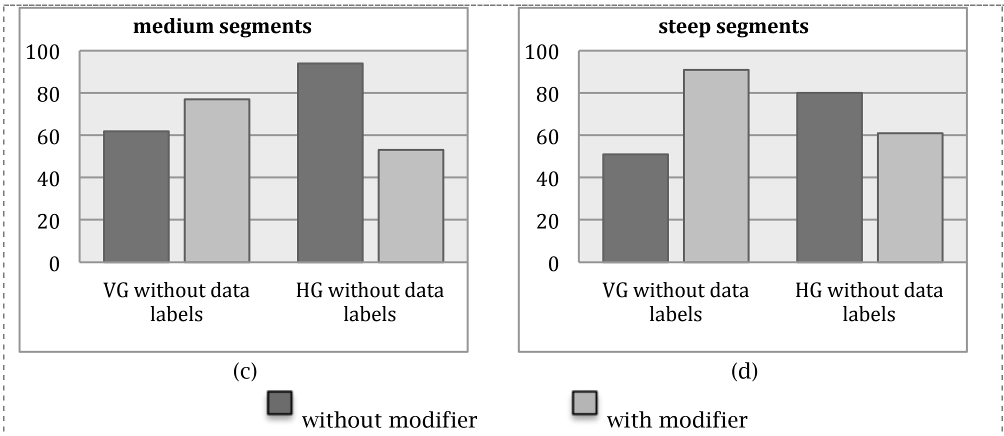


Figure 6-45 Effect of sensory modality

These results concluded that the use of modifiers for the separately mentioned segments was affected by the presence of data labels, by the sensory modality and the steepness of a line segment. Overall look at all three conditions showed that for the visual graphs without labels, the segments were referred in general with additional modifiers compared to the visual graph with labels and also compared to the haptic graphs. Furthermore, while all interactions between the modality and the slope-value was in line with the overall pattern in the modality-group, the medium and steep slope values seemed to have more effect for the label-group.

C. Errors In Referring Expressions

In addition to looking at how the segments were referred (with which attributes), whether they were referred correctly is also another crucial issue. The errors observed in the verbal descriptions can be classified into two main categories; (i) the errors regarding the direction of the line segments (the errors in the type-attribute, i.e. referring as a decrease while the line increases or stays stable) and (ii) the errors regarding the modifier of the state or action (the errors in the manner or in the size attribute). As previously mentioned, the segments may be referred as separate (individually) or they can be referred in a combination. To illustrate, following two verbal descriptions were used to describe the part of Graph-V depicted in Figure 6-46 and they exemplify two different types of errors.

Separately-mentioned:

There is a slight decrease (line-2/true) and then a very steep decrease (line-3/false)

In-combination:

There is a slight decrease (line-2 & line-3/false)

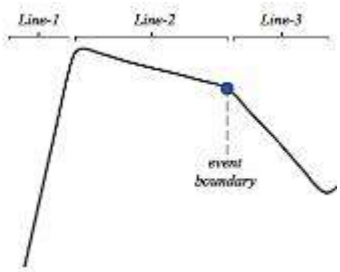


Figure 6-46 Separately-mentioned or the "in-combination" referring expressions

Table 6-36 gives the error rates (the number of wrongly-referred segment/the number of protocols that contain segment with the corresponding slope-value) for each slope-value for each condition.

Table 6-36 The error rates for each slope value and for each condition

	Steep	Medium	Slight	Horizontal
VG with labels	0.04	0.13	0.25	0.16
VG without labels	0.02	0.18	0.30	0.15
HG without labels	0.07	0.18	0.46	0.14

The results of Pearson's Chi Square tests showed that there was a significant association between the slope of the segment and the error rates, $\chi2$ (3)=130.68, $p<.001$. More errors for the slight segments were observed than that for the medium segments, $\chi2$ (1)=40.11, $p<.001$, than the horizontal segments $\chi2$ (1)=44.87, $p<.001$, N=940, and also than the steep segments $\chi2(1)=107.57$, $p<.001$. However, the participants made same amount of error for the medium and the horizontal segments, $p>.05$. The error rates for the medium and horizontal segments were significantly higher than that for the steep segments, ($\chi2$ (1)=30.24, $p<.001$; $\chi2$ (1)=24.89, $p<.001$ respectively), see Figure 6-47.

There was no significant difference among the conditions concerning the steep segments, $p >.05$, the medium segments, $p>.05$, and the horizontal segments $p>.05$. However there was significant difference for the slight segments, $\chi2$ (2)=17.53, $p<.001$. The errors rates for the haptic modality is higher than the visual graphs without labels $\chi2$ (1)=9.58, $p<.05$. However, there was no significant difference between the visual modalities.

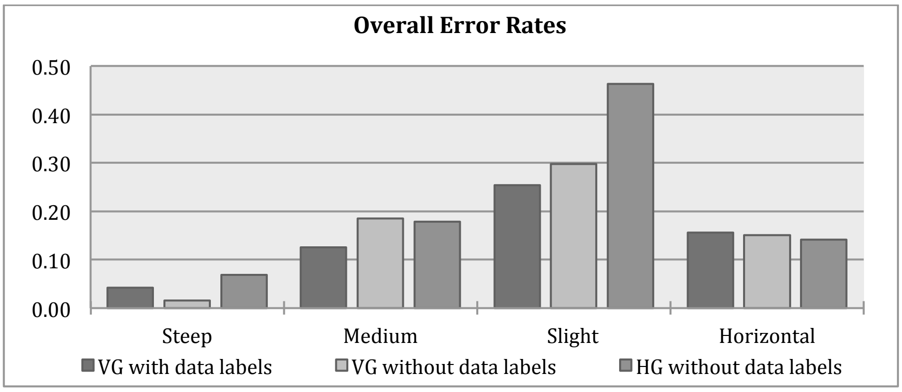


Figure 6-47 Overall Error Rates

C.1 Errors for steep segments. The participants made very few errors concerning steep segments. The errors for direction (the type-attribute) were only observed in the descriptions of Graph-7 (3 of 4 errors) and Graph 12 (1 of 4 errors) regarding the last steep segments. See Table 6-37 for the distribution of the errors across conditions.

Table 6-37 The number of the errors made regarding the shape segments

	Separate Units		Combined Units	
	Direction	Modifier	Direction	Modifier
VG with labels	2	3	0	0
VG without labels	0	2	0	0
HG without labels	2	3	0	3

Two participants in the visual condition with labels referred those regions in Graph-VI (Figure 6-48a) in the wrong order as "increase and decrease", though the sketch of one of the participants exhibited the decrease and increase in the correct order. However, the errors in the haptic condition point out more crucial issue, although the number of the error is also very small. Two different participants made these errors in different graphs regarding the changes of two consecutive steep segments. In the both cases, although the participants explored the second segment ("an increase" as exemplified in Figure 6-48c), they thought that they were doing haptically backward (right-to-left) motion (Figure

6-48d), therefore they conceptualized that part as just decreasing segment by ignoring the second part. The number of the errors regarding modifiers (the manner and the size of the action or the state) was also very low, and conducted by only the haptic explorers when they refer the steep segment in a referring expression in a combination with the other segments.

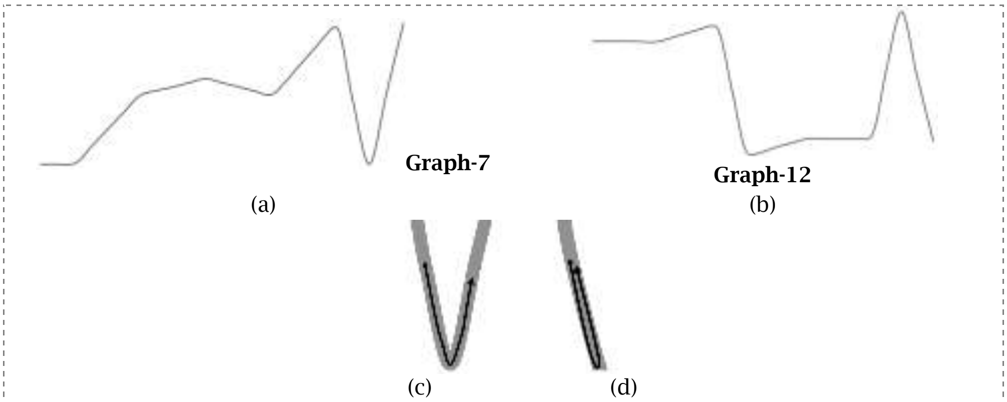


Figure 6-48 Common Errors in Graph-7

C.2 Errors for medium segments. The errors regarding medium segments were mostly made when the participants refer to them individually and were observed in the modifiers.

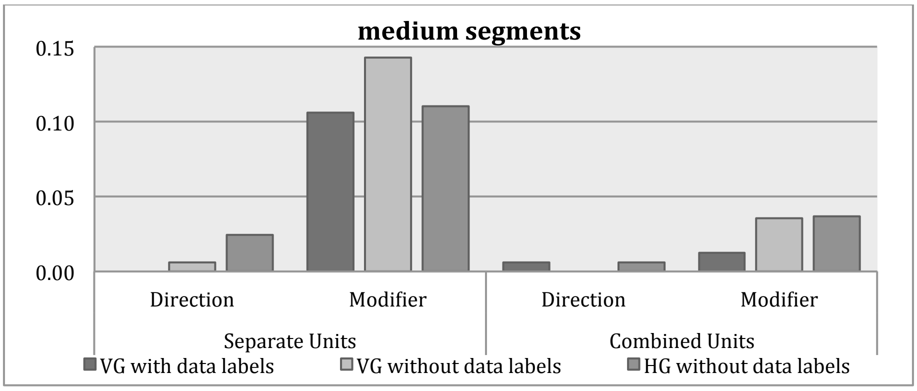


Figure 6-49 The error rates for the medium segments

C.3 Errors for slight segments. Unlike the errors regarding medium segments, the errors in regard to slight segments were mostly about the direction (the type-attribute). To exemplify, referring to a segment that represents a slight decrease as a horizontal line, or even as a slight increase is the most common mistake observed for this slope category.

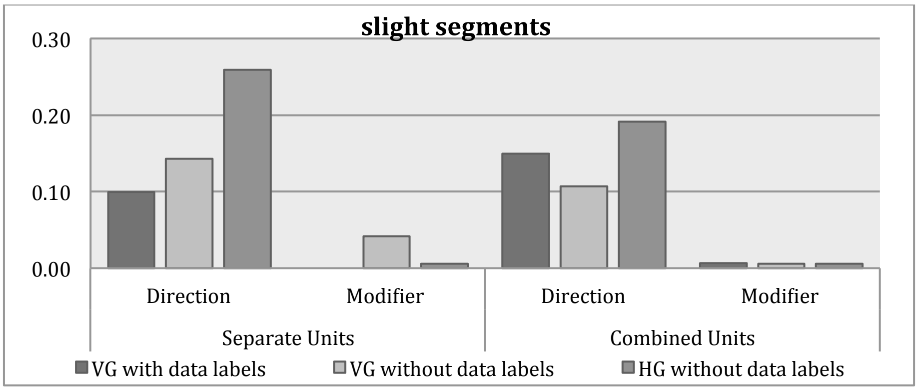


Figure 6-50 The error rates for the slight segments

C.4 Errors for horizontal segments. The errors for the horizontal segments exhibited same pattern with the errors for the slight segments, and they were made concerning the direction of the line segment. However, they mostly observed for the combined reference type by a majority indicating that they might be underestimated.

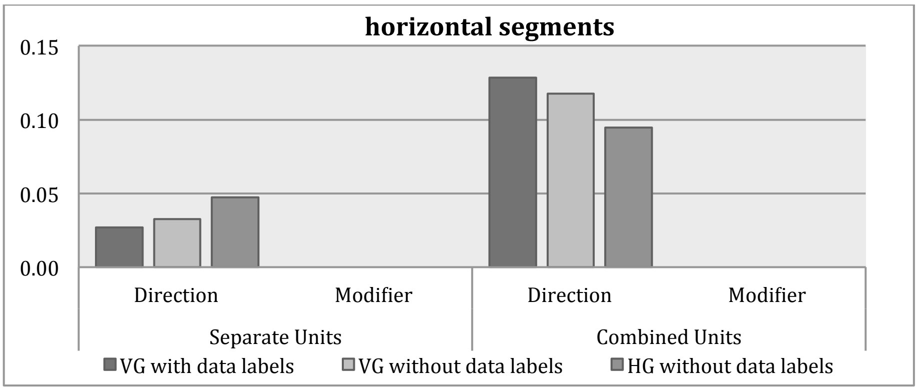
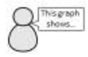


Figure 6-51 The error rates for the horizontal segments

To summarize, the steep segments were identified correctly in all conditions with a few exceptions. Although the direction of the medium segments is easy to grasp and less error were made about it, grasping the size or the manner of the segment can be problematic. On the other hand, the errors regarding the slight and horizontal segments were more basic and they were about the direction of the line.

6.4.7.2.Time And Event Denoting Expressions



The descriptions of abstract events depicted in the graphs involve referring expressions addressing both space and time dimensions. Lets take the sentence (1) as an example;

(1) The visit count increases rapidly from May to July.

As it has been introduced in 6.4.3, temporal information expressed in the verbal descriptions was evaluated under time-denoting expressions (for example, from May to July). The rest of the sentence can be labeled with event-denoting attributes. In this part, the differences in the use of temporal and event-denoting expressions were investigated across the conditions; see Table 6-38 for the basic attributes and an excerpt for each attribute. First, for each attribute in the annotation scheme, sub-events were coded with respect to whether it contains the attribute or not. Total number of instances of each attribute in the utterance was also count. To illustrate, a sub-event description such as “it increases from 10 to 70 in a steep and linear manner” can be identified as <attribute, value> pairs as follows;

<act, increase>, <value_from, 10>, <value_to_70>, <manner, steep>, <manner, linear>

Then the frequency for each attribute observed in the sub-event is calculated. For this example, the values are as follows

act#=1, value#=2, manner#=2

And all of the sub-event ascriptions were summed for each participant and divided to total sub-event count produced by him/her, the resulting mean value corresponds to the average value of each attribute per description. For overall analysis, the mean for all graphs were calculated. For graph based-analysis, only the mean value for each attribute was used. After that, the conditions were compared in a between-subject design. Appendix-G presents frequency tables for each attribute with respect to each graph in all conditions.

Table 6-38 Basic temporal and event denting expressions

Basic Time-denoting Attributes	Example
• Explicit reference to x-axis (i.e months)	in July, from April to June etc.
• Vague reference to x-axis (i.e seasons)	in winter
• Indexical reference to x-axis	after the third month,
• General	In general
• Location	After the point with the highest value
Basic Event-denoting Attributes	**Example**
• Type	
o State	There is a decrease
o Action	It decreases
• Manner	Slowly, curvy…
• Size	Steep, slight…
• Value	It reaches to 80.
• Shape	M-letter, volcano…
• Relation	It is higher than..
• No temporal tag	Type without any additional attributes

A. Effect of Data Labels

The effect of the data labels for each time-denoting and event-denoting attributes were compared by conducting a Mann-Whitney test (by comparing the visual graphs with labels and the visual graphs without labels), see Table 6-39.

A.1. Time-denoting Expressions. As expected, the participants in the "visual graph with labels" condition produced more time-denoting expression that contain "Months" compared to the participants in the "visual graph without labels" condition, U=51.50, z=-2.95, p<.05. On the other hand, the use of the time-denoting expressions that points to spatio-temporal information ("location"), was higher for the "visual graphs without labels" condition, U=77.50, z=-2.19, p<.05. Moreover, the number of sub-events that do not have any temporal expression was also higher for the visual graphs without labels, U=58.50, z=-2.63, p<.05.

Table 6-39 The descriptive statistics for the time-denoting attributes

	Explicit*	Vague	Indexical	General	Location*	No Tag*
VG with labels	0.81	0.13	0.16	0.14	0.03	0.28
VG without labels	0.32	0.06	0.37	0.15	0.12	0.56

In addition to the overall results, the graph-based differences were also calculated for each time-denoting expression. The use of explicit reference (such as months) showed differences for each graph. Besides, the number of the sub-event phrases without any time-denoting expressions in without-label condition was higher than that in with-label condition for 8 of12 graphs (from II to IX).

In addition to the overall comparison of each category across the conditions, the correctness of the content was also tested for the attributes that involve reference to the x-axis; these attributes are the explicit references, the vague references and the indexical references, see Table 6-38 for the excerpts that exemplify each category. Each reference was classified into three w.r.t its truth-value; true, partially true and false. The true category corresponds to the exact match between the data labels and the referred month. To illustrate (see Figure 6-52), for a utterance like "It reaches bottom in April", the temporal expression is classified as *true*. If it temporally refers to the neighbor months, March or May in this case, it is considered as *partially true*. The expressions other than those two are classified as *false*.

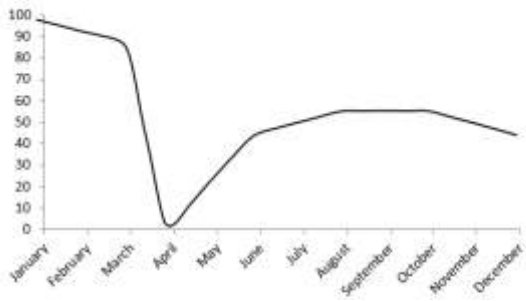


Figure 6-52 Sample Graph - revise caption

Table 6-40 True/False distributions for each temporal reference category in the label-Group

	Explicit Reference			Vague Reference			Indexical Reference		
	true	**partially true**	**false**	**true**	**partially true**	**false**	**true**	**partially true**	**false**
VG with labels	77.70	11.97	10.33	86.49	13.51	0.00	62.5	25	12.5
VG without labels	59.66	10.92	29.41	77.42	19.35	3.23	75.00	18.75	6.25

Table 6-40 presents the distribution of each expression category that contains a reference to the x-axis w.r.t their truth-value. Combining these results with the results presented in Table 6-39, it can be concluded that the users in the with-label condition prefer to use the explicit labels in their description, and they use this information correctly with the 77.7 % success. On the other hand, they were partially true or false for approx. 22 % of explicit references. The correct use of the explicit reference (almost 60%) was low for the condition without data-labels, and 30% of the explicit references contained wrong information. For the vague reference, more correct responses were observed, the vague references inherently have more uncertainty and the disparity compared to the explicit references, therefore higher truth-values were expected, but still the with-label group outperforms the without-labels group. The indexical references were preferred more by the participants in the without-label than the participants in with-label condition and similar to the vague references concerning the truth-values; the same pattern with very low false references was observed.

A.2. Event-denoting Expressions. For the use of event-denoting expressions, the overall analysis (the average values for all 12 graphs) indicated that the use of "size" attribute differed w.r.t the presence of the data labels, U=76.50, z=-1.95, p<05. Although the overall analysis did not point out the differences, a closer look with a graph-based analysis showed more variety regarding the differences in the use of other attributes, such as value (for graph VIII), and relation attributes (for graph V and VI).

Table 6-41 The descriptive statistics for the event-denoting attributes

	State	Action	Manner	Size*	Value	Shape	Relation	Additional
VG with labels	0.63	0.93	0.65	0.38	0.67	0.10	0.10	0.09
VG without labels	0.69	0.89	0.70	0.61	0.76	0.10	0.22	0.15

Similar to the truth-value analysis for time-denoting expressions regarding x-axis labels, the "value" attribute was also investigated in more detail w.r.t whether the reference truly describe the data represented in the y-axis of the graph. The "value" attribute can be also classified into two sub-categories: it can be numeric (such as it is 70) or a vague ascription (such as it is high). Table 6-42 presents the truth-value distributions for each category across the conditions. Interpreting these results with the results presented in the previous table, it can be concluded that although similar amount of reference was made in general in both conditions, the distribution shows variance, in regards to the numeric reference (but note that the differences are rather small). Having the data labels does not entail exact match between the reference and the referent, in other words the participants may not be exact (true) in their reference. However, when the participants in without–labels condition refer to the data labels in an explicit or vague manner, then they were mostly correct.

Table 6-42 The True/False Distributions for the "value" event-denoting attribute in the Label-Group

	Explicit(Numeric)			Vague Ascription		
	true	partially true	false	true	partially true	false
VG with labels	71.11	20.00	8.89	85.87	8.15	5.98
VG without labels	84.00	4.00	12.00	81.79	14.81	3.40

B. Effect of Sensory Modality.

Effect of sensory modality for each time-denoting and event-denoting attributes was compared by conducting a Mann-Whitney test (by comparing the visual graph without labels and the haptic graphs without labels), see Table 23.

B.1. Time-denoting Expressions. Time-denoting expressions that refer to seasons were preferred more in the haptic modality, U=69.50, z=-2.48, p<.05. There are no significant overall differences in the use of other time-denoting attributes between the two modalities. However, the graph-based analysis highlighted some additional individual differences due to global shape of the graphs. These differences concern explicit references (for graph I and II), indexical references (for graph V and XI) and the location attributes (for graph V).

Table 6-43 The descriptive statistics for the time-denoting attributes

	Explicit*	Vague	Indexical	General	Location	No Tag
VG without labels	0.32	0.06	0.37	0.15	0.12	0.56
HG without labels	0.30	0.24	0.22	0.14	0.18	0.47

Table 6-40 presents the distribution of each expression category that contains a reference to the x-axis w.r.t their truth-value. Combining these results with the results presented in Table 6-43, it can be concluded that the users in both condition exhibit a similar pattern in their use of explicit labels, although the graphs do not contain this information. The false references in haptic condition were lower than that in the visual condition, and the rates for the partially true references was higher in the haptic graph condition. Although the differences are not distinct to draw a conclusion, it can be hypothesized that when the participants assign virtual data labels to the graphs they are usually correct about them (totally or partially). For the vague and indexical references, more correct responses were observed; the pattern in the distribution across modalities was similar.

Table 6-44 The True/False Distributions for each temporal reference category in the Modality-Group

	Explicit Reference			Vague Reference			Indexical Reference		
	true	partially true	false	true	partially true	false	true	partially true	false
VG without labels	59.66	10.92	29.41	77.42	19.35	3.23	75.00	18.75	6.25
HG without labels	46.21	34.48	19.31	81.08	12.16	6.76	60.00	30.00	10.00

B.2. Event-denoting Expressions. For event-denoting expressions, only the use of the "value" attribute showed a significant difference, U=73.00, z=-2.08,p<.05 across two modalities. Additionally, the graph-based analysis showed more variety with respect to several attributes such as state (for graph VI), manner (for graph II and XI), and shape (for graph IV).

Table 6-45 The descriptive statistics for the event-denoting attributes

	State	Action	Manner	Size	Value*	Shape	Relation	Additional
VG without labels	0.69	0.89	0.70	0.61	0.76	0.10	0.22	0.15
HG without labels	0.74	0.93	0.62	0.62	0.53	0.04	0.30	0.12

Furthermore, the truth-value distributions of references regarding “value” attribute were presented in Table 6-46. Unlike the distribution patterns in which the visual graphs with and without labels were compared, clear differences in the pattern were observed. In this group, both the visual and haptic graphs do not contain data labels, but the visual reference frame around the graph line may help users in the visual condition to infer the numerical values. On the other hand, accessing the reference frame in haptic modality is more difficult, and the users can make such inferences by using only changes on the graph line in general. Still in the cases when they preferred to make an explicit numeric references, they were partially or exactly correct about it. On the other hand, it seems that they made more wrong estimations regarding the vague relational ascriptions (such as low or high).

Table 6-46 The True/False Distributions for the "value" event-denoting attribute in the Modality-Group

	Numeric			Relations		
	true	partially true	false	true	partially true	false
VG without labels	84.00	4.00	12.00	81.79	14.81	3.40
HG without labels	46.67	53.33	0.00	69.16	11.68	19.16

C. Interim Summary for the analyses on Time and Event Denoting Expressions

First of all, the overall analyses (the averaged data for all graphs) on the time and event denoting expressions do not seem to be conclusive and informative to pinpoint the differences. The graphical features appear to have more effect on the referring expressions. The detailed analyses on the content of the temporal expressions indicated that even in the graph with labels, the correct explicit reference rate was around 78%. In the absence of data labels, the reference rate for the explicit expressions was low, and the exact match rate was also low as could be expected. However, in general when we look at the sum of the exact match and the partial match rates, it can be concluded that when they prefer to use explicit reference they use it almost in a correct way. On the other hand, the vague descriptions were more preferred in the haptic condition; again their truth rate was quite high.

For visually perceived graphs, the participants were more successful in inferring the numeric values (of the y-axis) than in inferring the explicit labels of x-axis. For haptic explorers, referring to both axes explicitly seems to be problematic, on the other hand they are good at having a rough estimation about both temporal and value information. Considering low reference rates of values in general, one can speculate that if the user feels confident about the numeric values, then s/he adds this information to her/his verbal description. Most of the mistakes were observed in estimating value as vague ascriptions such as high or low (in relational terms), this indicates potential content for a verbal assistance.

6.4.7.3. Speech-Accompanying Gestures

Gestures were annotated w.r.t the annotation scheme provided in detail in Chapter 4.5.3. Shortly, each representational gesture was classified as *static* or *dynamic*. According to this classification, the hand movements conducted in a small space without having any directed trajectory were categorized as *static (non-directional)*[30], whereas the hand movements with the aimed trajectory on the air were classified as *dynamic (directional gestures).* To illustrate, the hand shape in a form of reverse L-letter was one of the frequently used *static* type of gesture to refer to peak value. On the other hand, the diagonal upward movement or drawing wave on the air are two examples of a *dynamic* gesture. The directional gestures (dynamic) were also classified into two categories; (i) *the single direction (*vertical/diagonal/horizontal), and (ii) *the multiple directions.* The gestures that contained movement in only one direction (such as upward) were classified under the "*single direction*" category, while category of the "*multiple directions*" covers the gestures formed with the combinations of the one-directional gestures in different directions or with the inflection points. Two coders analyzed and classified the data. The interrater reliability was calculated by Cohen's kappa. The results revealed a value of .72 that indicates substantial interrater agreement.

A. Static versus Dynamic Gestures

As a within-subject variable, the number of the dynamic (directional) and the static gestures produced by each participant in each condition were compared by performing a two-way ANOVA, see Table 6-47 for the overall descriptive statistics for each condition.

Table 6-47 The descriptive statistics for the speech-accompanying gestures for each condition

	Static	Dynamic	Total
VG with labels	46	227	273
VG without labels	90	465	555
HG without labels	88	503	591

The comparisons were based on the average gesture count for all graphs' descriptions. The findings indicates that the participants tended to produce more dynamical gestures than the static gestures, $F(1,37)=160.55$, $p<.001$, $\eta^2=.81$. There was no significant interaction between the condition and the gesture type. Additionally, no difference between the conditions in terms of the overall number of gesture was observed, as already presented in Table 6-26 in 6.4.7.2.

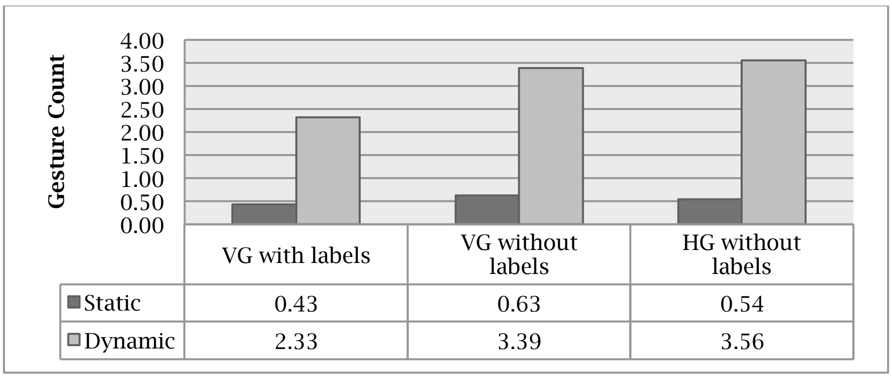


Figure 6-53 Average gesture count for each condition

[30] Reminder: It should be noted that a static gesture can be directional as well (i.e. a static diagonal hand posture) and such cases were annotated in line with the gesture annotation scheme that I employed. However, they were extremely low in this data set, therefore here the static gestures connotates non-directional gestures.

B. Directionality in Speech-accompanying Gestures

The results of the previous analysis showed that the participants tended to produce dynamic gestures during the communication over line graphs. Therefore a closer look on the dynamic gestures was taken regarding the differences in the directionality feature. The participants in all conditions tended to produce more one-directional gestures compared to multi-directional gestures $F(1,37)=38.54$, $p<.001$, $\eta^2=.51$. No significant difference in the use of one-directional gestures was found across the visual graph conditions. However, the results of a Mann Whitney test indicated that the use of multi-directional gestures exhibited a main effect across two modalities, $\chi 2=8\text{-}98$, $p=.011$, $U=49.500$, $z=-2.22$, $p<.05$). The number of the multi-directional gesture was significantly high for the haptic modality compared to the visual modality.

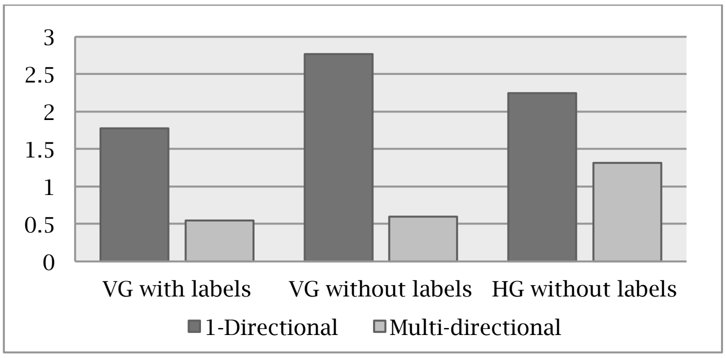


Figure 6-54 1-Directional versus Multi-Directional Gestures

C. Gesture and Referent Relation in terms of Steepness

In this part, I investigated the relation between the steepness of a line segment and whether its description is accompanied by a gesture or not. The descriptions for the line complexes were excluded since the individual effects of the slope values on gesture production were of interest. All the line segments which are referred separately and with accompanying gestures were categorized w.r.t their slope values. Then a 2-way ANOVA (condition x steepness) was conducted on the mean gesture count for the reference to each slope-value.

Effect of Data Labels. The results showed that there was a main effect of steepness on gesture production, $F(1.84, 55.15)=11.11$, $p<.001$, $\eta^2=27$, see Figure 6-55. Follow-up pairwise comparisons indicated that the gesture rate for the steep segments was higher than that for the medium segments ($F(1,30)=8.70$, $p<.01$, $\eta^2=22$), higher than for the slight segments ($F(1,30)=17.58$, $p<.001$, $\eta^2=37$) and also higher than for the horizontal segments ($F(1,30)=13.09$, $p<.001$, $\eta^2=30$). The gesture rate for the medium segment was also significantly higher than that for the slight segments ($F(1,30)=5.86$, $p<.05$, $\eta^2=16$). On the other hand, the difference between the medium segments and the horizontal segments and also the difference between the slight segments and the horizontal segments were not significant. No main effect of data labels was observed as well.

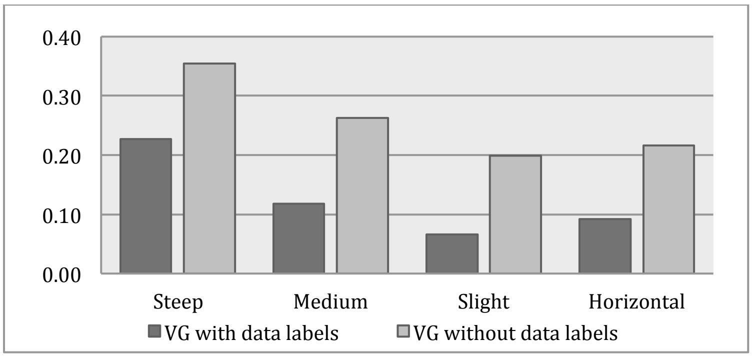


Figure 6-55 The gesture rate w.r.t slope values in the visual conditions

Effect of Modality. The results also showed that there was a main effect of steepness on gesture production, $F(1.99, 59.59)=19.81$, $p<.001$, $\eta^2=40$, see Figure 6-55. Follow-up pairwise comparisons indicated that the gesture rate for the steep segments was higher than that for the medium segments ($F(1,30)=14.71$, $p<.01$, $\eta^2=33$), higher than for the slight segments ($F(1,30)=22.43$, $p<.001$, $\eta^2=43$) and also higher than for the horizontal segments ($F(1,30)=34.07$, $p<.001$, $\eta^2=53$). The gesture rate for the medium segments was significant at marginal value ($p=.055$) than that for the slight segments but significantly higher than that for the horizontal segments, ($F(1,30)=13.81$, $p<.001$, $\eta^2=30$). Moreover, more gestures were produced for the slight segments compared to the horizontal segments, ($F(1,30)=4.34$, $p<.05$, $\eta^2=13$). There was no main effect of the modality.

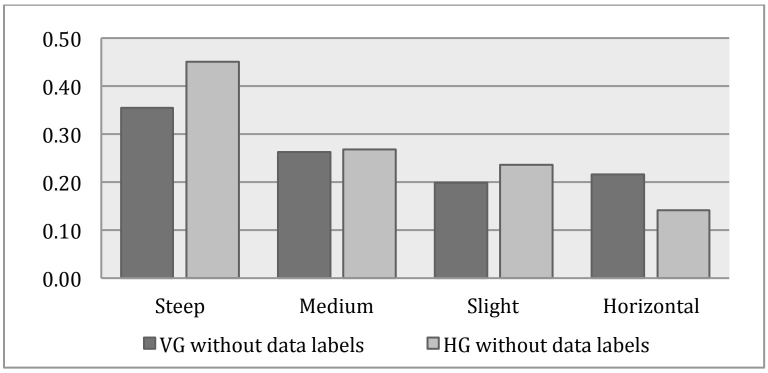


Figure 6-56 The gesture rate w.r.t slope value in the modality group

These results revealed that the graph comprehenders produced more gestures when they are describing the steep segments (it should be noted that in this analysis only separately/individually referred segments were included). With the graphs without labels (the visual and haptic graphs), the individual descriptions for medium slopes were accompanied with gestures more than the slight and horizontal segments, indicating that the steeper the referent slope is, the higher the possibility of the gesture production is.

D. Negated content and its reflection in co-verbal gestures

The results reported so far were obtained through systematically designed experimental conditions. The topic, which is investigated here, came into a prominence as a result of an investigation regarding the "negation" as an event-denoting attribute during the analyses of the verbal content. This analysis focused on the speech parts that involve the "negative statements", and also the "affirmative statements" that are followed or preceded by the negated statements that refer to the same segment; such as "it is not straight, it is curved" or in the other way around "it is curved, it is not straight". Since the participants were not forced or primed about neither the use of any linguistic structure nor the use of speech-accompanying gestures, the instances that exemplify this topic presented above were spontaneous and rare. Therefore the quantitative methods involving statistical tests were not applied to this analysis; instead I focus on the qualitative report enriched with frequencies and examples. However, even just the presence of such examples provides intriguing cases for a theoretical discussion.

Table 6-48 presents the number of gestures for each type of statements (the affirmative and the negated). First it is possible to produce a sentence that involves a negative statement without producing a gesture. Second, a gesture may accompany to the "affirmative statement" of the sentence, but not to the "negative statement". Third, a gesture may accompany only to the "negative statement". The fourth category involves the cases that the participants produced gestures for both affirmative and negative statements. The gestures were categorized semantically into two; the *Type-I* gestures are the gestures, which are congruent with the property of the referred graph segment, (*the congruent-with-the graph*). On the other hand, the *Type-II* gestures exhibit congruency with what modifier used in the verbal statement represents (*the congruent-with-the modifier).* To illustrate, let's take a verbal description such as "it is not sharp". According to this classification, a slightly diagonal (w.r.t horizontal axis) dynamic gesture that resembles to what the graphical entity is (i.e. a slight/not sharp movement) would be classified as *Type-I*, whereas a vertical or near vertical diagonal dynamic gesture that resembles to what the modifier represents (i.e. a sharp movement) would be classified as a *Type-II* gesture.

For the affirmative statements, mostly the congruent gestures were accompanied to the speech parts; i.e. the verbal description "it is sharp" accompanied by vertical or near vertical diagonal dynamic gesture. On the other hand, the gestures produced during the generation of the "negative statements" exhibited differences from the "affirmative statements". The Type-2 gestures display oppositeness; therefore they mostly accompany the negative statements. Additionally, both gesture types (I &II) were observed for the gestures that accompany to the negative statements.

5 of 16 participants in the visual condition with labels produced sentences that involve negated statements; this ratio was higher for the visual condition without labels (11 of 16 participants) and also in the haptic condition (12 of 16 participants). The overall number of the sentences that involve negated statements also showed similar pattern; 14 statements were produced in the visual condition with labels, on the other hand, 23 statements in the visual condition without labels B and 29 statements in the haptic condition were produced.

Table 6-48 The number of instances that involve negative statements and also the number of Type-I and Type-II gestures for each condition (st.:statement)

	Gesture types	VG with labels	VG without labels	HG without labels
Without co-verbal gesture				
Negative st.		6	10	16
With co-verbal gestures				
Affirmative st.	Type-I	8	4	3
Negative st.	*Type-1*	-	3	1
	Type-2	-	-	4
For both affirmative and negative st.	*Type-1*	-	5	1
	Type-2	-	1	4
TOTAL		14	23	29

In the visual condition with labels, the participants preferred not to produce gestures for the sentences that involve negative statements or they produced gestures for just the affirmative parts of the sentences.

On the other hand, the gesture production for the negative statements was nearly similar for the visual condition without labels and for the haptic condition. However, a striking difference in terms of the types of the gestures was observed between these two groups. While the participants in the visual-without label condition tended to produce Type-1 gestures (8 of 9 cases) for the negative statements, the participants in the haptic condition tended to produce Type-2 gesture (8 of 10 cases).

In the haptic condition, the previous exploration actions (mostly salient parts i.e. previous steep segment) seem to affect the online comprehension more. Beside the effect of the modality, the order of the negative and affirmative statements also seemed to have an effect on the gesture. The examples for three possible instances are presented below;

(1) Affirmative statement + negative statement

"it is curved, it is not straight"

(2) Negative statement + affirmative statement

"it is not straight, it is curved

(3) False affirmative statement and + negative of the previous false affirmative statement

"it is straight, no actually it is not straight"

When we look at the gestures produced in the haptic condition (see Table 6-49), we can see the effect of order more apparently. If the affirmative statement (that modifies the basic action) takes the first place and is accompanied by a gesture (by nature, they are Type-1) then again Type-1 gesture was observed for the negative statements (Excerpt#1). But if the affirmative statement in the first place was not accompanied by a gesture or the first part did not have any affirmative modifier (Excerpt#2) then the negative statements was followed by type-2 gestures (8 of 10 cases). In Excerpt#3 first, false affirmative sentence was produced with type-2 gesture (the hand showed a steep decrease but the participants described it as a "30° angle", the referred segments is a steep segment), then the participant updated and corrected the verbal description and continued with a previous concurrent (type-1) gesture.

Table 6-49 Three excerpts from the verbal descriptions and the gestures in the haptic condition

Excerpt No	Verbal Description
(1)	[In September, it starts to decrease], [that wiggling disappears (previously mentioned)] [Basic action (without modifier)]$_{\text{Type-I}}$ >> [Negative Statement]$_{\text{Type-II}}$ *In Turkish: Tekrar eylül ayında daha duşüşe geçip, o kıpırdanma yok olup…*
(2)	[It is never increasing straight], [it always increases in a sloping way] [Basic action (with modifier / negative statement)]$_{\text{Type-II}}$ >> [Affirmative statement]$_{\text{Type-I}}$ *In Turkish: Hiçbir zaman düz bir şekilde artmıyor, sürekli eğimli bir şekilde artıyor.*
(3)	[In a slight way, I mean, with approx. 30° angle it is going up.] [Actually, it is not 30° angle, it is like 70° angle, it goes toward up.] [Basic action (with modifier / affirmative statement)]$_{\text{Type-II}}$ >> [correction in the affirmative statement]$_{\text{Type-I}}$ *In Turkish: Hafif bir şekilde yani yatay eksenle yaklaşık 30 derecelik bir açı yaparak hafif yukarı çıkıyor, 30 derece de değil daha doğrusu 70 derece falan galiba, yukarı doğru çıkıyor.*

E. Interim Discussion for Speech-Accompanying Gestures

The analyses of the speech-accompanying gestures showed that in both haptic and visual conditions, more dynamic (directional) gestures compared to the static gestures were produced supporting the idea that line graphs emphasize trend conceptualization (Zacks and Tversky, 1999). The directionality was also examined under two categories; the one-directional and the multi-directional gestures. The results indicated that the one-directional gestures were produced more in all conditions; this is not surprising since the reference scope of the multi-directional gestures is tended to be broader. However for the graph perceived through haptic modality, more directional gestures were observed compared to the visual conditions indicating that the haptic graph comprehenders conceptualize graph or graph segments as a connected entity, although they still produce action-based fine-grained segmentations in their verbal descriptions. This also indicates that the segmentation of basic points could be suffered in the haptic modality and acquiring landmark information is challenging issue for the haptic comprehenders.

Additionally, a significant effect of perceptual saliency on gesture production was found, the steeper the segment is, higher the gesture production rate is. Combining this with the results of verbal description analyses, I can conclude that steep segments perceived through haptic modality are well comprehended; therefore the verbal assistance should be focused on other types of slope, which are less salient.

Finally, although the lack of statistical power, last analysis on the relation between the gesture production and the negated content showed that the sensory modality interferes with language and graph comprehension. This relation is not only interesting for graph comprehension research but also for language and gesture research. The results presented beforehand already demonstrated that the presence of data labels facilitate not only a segmentation but also a defragmentation. In other words, more coarse-grained semantically driven units were observed for the graphs that contain data labels. On the other hand, the graphs presented without labels (the visual or the haptic) leads to more action based fine-grained units. Previously perceived salient features that highlight relations can reflect themselves in the negative statements. However, in the visual condition (without labels) still the graphical features seem to be salient considering gesture production, which is congruent with the referent's graphical property. On the other hand, due to close relation between the motor actions performed during haptic exploration and the underlying mechanisms of the gesture production, the salient features, which were actively explored previously, might have more influence on gesture production for haptically perceived graphs. The presence of incongruence (mismatch) was affected by the order of affirmative and negative statements. Thus, this effect may not be simply due to effect of language on gesture production. Type-II errors might be originated from the active mental representation of the previous salient segment due to recent exploration.

Table 6-50 summarizes all the results (with significance and effect sizes) provided in this part. The overall discussion of the results reported in this part will be provided under the general discussion.

Table 6-50 Significance and effect sizes for all tests in "Event Description"

6.4.8.1 Verbal Descriptions			
	Label-Group		**Modality-Group**
A. The Choice on Reference Scope			
Main Effect	Sig. (η^2=57)		Sig. (η^2=59)
Choice x Condition	No. Sig.		No. Sig.
Most Preferred Type (Overall)	Separate & Combined.	*	Separate
A.1 Effect of Slope on Preferred Reference Scope			
Most Preferred Type (For Steep Seg.)	Separate		Separate
Choice X Condition (Steep)	No. Sig.		No. Sig.
Most Preferred Type (For Medium Seg.)	Separate & Combined.		Separate & Combined.
Choice X Condition (Medium)	No. Sig.		No. Sig.
Most Preferred Type (For Slight Seg..)	Combined	*	Combined and Separate
Choice X Condition (Slight)	Sig. (η^2=16)	*	No. Sig.
Most Preferred Type (For Horizontal Seg.)	ALL		ALL
Choice X Condition (Horizontal)	Sig. (η^2=17)		Sig. (η^2=20)

A.2 Effect of Angle on Preferred Reference Scope			
Choice (for A1-Value)	Sig. (η^2=49)		Sig. (η^2=48)
Most Preferred Type (A1)	Separate		Separate
Choice x Condition (A1)	No. Sig.		No. Sig.
Choice (for A2-Value)	Sig. (η^2=63)		Sig. (η^2=66)
Most Preferred Type (A2)	Separate		Separate
Choice x Condition (A2)	No. Sig.		No. Sig.
Choice (for A3-Value)	Sig. (η^2=46)		Sig. (η^2=48)
Most Preferred Type (A3)	Separate & Combined		Separate & Combined
Choice x Condition (A3)	No. Sig.		No. Sig.
Choice (for A4-Value)	Sig. (η^2=69)		Sig. (η^2=61)
Most Preferred Type (A4)	Combined		Combined
Choice x Condition (A4)	Sig. (η^2=12)	*	No. Sig.
A.3 Interaction Effect of Angle and Slope	See Figure 6.43		
B Use of Modifiers			
Main Effect	Sig. (χ2=36.53)		Sig. (χ2=33.87)
Main Effect (For Steep Seg.)	Sig. (χ2=15.26)		Sig. (χ2=12.39)
Main Effect (For Medium Seg.)	Sig. (χ2=15.77)		Sig. (χ2=10.78)
Main Effect (For Slight Seg.)	Sig. (χ2=4.27)		Sig. (χ2=10.70)
C Errors in Referring Expressions			
Main Effect	No. Sig.	*	Sig. (χ2=9.58)
Segment Type with more error rate	Slight		Slight
Most observed error type (steep)	NA		NA
Most observed error type (medium)	Modifier in Separate Units		Modifier in Separate Units
Most observed error type (slight)	Direction Errors		Direction Errors
Most observed error type (horizontal)	Direction Error in Combined Units		Direction Error in Combined Units
6.4.8.2 Time and Event Denoting Expressions			
Type of Attribute that shows difference in time-denoting expressions	Months & Location	*	Season
Type of Attribute that shows difference in event-denoting expressions	Size	*	Value
6.4.8.3 Speech-Accompanying Gestures			
Static vs. Dynamic	Dynamic		Dynamic
1-Directional vs. Multi-Directional	1-Directional		1-Directional
Condition (on the #of Multi-directional gestures)	No. Sig.		Sig. (z=-2.22)
Gesture & Referent Relation			
Main Effect of Slope	Sig. (η^2=27)		Sig. (η^2=40)
Most co-occur slope type	Steep		Steep
Condition	No. Sig.		No. Sig.
Gesture & Negated Content Relation			
Most preferred gesture type	Type-I		Type-II

6.5. General Discussion

In the design of the verbally assistance system that helps haptic graph comprehenders with hard to encode information, the segmentation of events and the descriptions to those events are two crucial topics that need to be addressed. As already suggested by Zacks and Swallow (2007), facilitating the event segmentation has important benefits. It has an influence on the online comprehension of the event and also on recalling the event for later use. These benefits may be especially crucial for some domain for example in education. Aiding haptic graph comprehension by providing verbal assistance would be also considered a domain where this line of research is beneficial. This chapter first provided existing literature on those topics, however the research conducted thus far is limited with the visual modality and with the investigation of simple animations of basic daily activities. In the graph domain, the concept of event differs in that respect. The events in graph domain manifests themselves in two layer, one is conceptual, namely the abstract event depicted by a graph and another is physical, namely the graph. The conceptual event is comprehended through objectified version of event. In the visual conditions, the concretization is conducted based on well-established Gestalt principles. Therefore in general, the perceptual saliency also corresponds to the conceptual saliency in the graph domain (if the appropriate graph type is chosen for the message conveyed). On the other hand, the haptic perception is sequential, and this difference in the sensory modality creates a gap in between the perceptual and conceptual layers. In order to provide a link between those layers, facilitating appropriate event segmentation in the partonomic level and providing distinct and right amount of information regarding the taxonomic level should be investigated.

This chapter was concentrated on the analyses of how the graph comprehenders segment the graphs (and the conceptual event) into sub-units, of which parameters govern this segmentation (concerning the partonomic and taxonomic relations) and of how they refer to those sub-units. Those questions were investigated through conducting detailed set of analyses. As a result, very rich multi-modal data sets that contain linguistic aspects of the verbal descriptions, speech accompanying gestures and post-exploration sketches were collected and analyzed for this purpose.

First of all, the expressivity evaluations of the verbal descriptions and the sketches provide valuable results. They indicate that the outcomes for the haptically perceived graphs lack of expressivity. Additionally, semantic and syntactic structure of the verbal descriptions with high expressivity scores would be a good candidate to be used in the design of a verbal assistance system. The gesture and drawing analyses also proved to be very helpful in understanding how the participants conceptualize the graph for the cases in which the verbal descriptions were not expressive enough (especially if they were vague about the event boundaries). They also provided additional information to resolve ambiguities occurred in the verbal descriptions to shed a light whether these problems occur in the conceptualization or just in the verbal outcome.

The overall results indicate that the effect of the graph shape was observed in nearly all parameters, indicating that the pattern of the graph is one of the most indicative factors that have an effect on the event segmentation and on the event referring. The main purpose of this experiment was to investigate the effect of the amodal geometric properties, therefore the effect of slope, angle, and polarity were subjected to the more detailed statistical tests. The haptic graph comprehenders do not have problem acquiring information concerning steep segments; therefore providing a verbal assistance for these regions might be backgrounded. The medium segments can be also considered as non-problematic regions as well, but the information about the steepness of the medium slope would be useful for haptic graph comprehenders. On the other hand, the results point out that the slight and horizontal segments are challenging regions for the haptic explorers. For these regions, the use of modifiers was very low and also the error rate was high. Particularly, detecting the presence of the change and its direction seems tricky issue for these regions.

Regardless of the sensory modality or the presence of data labels, the steep segments and the acute angles are usually treated as separate entities. However, for the rest of the features, segmentation exhibits differences among conditions regarding both sensory modality and the presence of data labels. As already discussed in the previously interim discussion, a successful segmentation requires dividing the content into the meaningful parts, some parts might refer to a small region (the fine-grained) and some parts might refer to a broader region (the coarse-grained) without taking small changes into account. The results showed that (also considering the high expressivity scores both for the descriptions and the sketches for the visual graphs with labels) the presence of data labels help participants to divide the conceptual event in a more effective way with more coarse-grained units. As existing literature suggested that, the coarse-grained descriptions are conceptually driven, the more object oriented and more precise about the object properties. On the other hand, the fine-grained units were perceptually driven, and focus on the actions. It seems that providing the landmark information in the haptic condition (not as haptic data labels but as perceivable points on the graph line) with a verbal assistance may be really useful in bridging this gap. These results also indicate that instead of providing information about all landmarks, adjusting the content w.r.t event segmentation considering temporal and taxonomic granularity into account would lead to successful conceptualization. More detailed analysis regarding the referring expression production during the collaborative activity will be elaborated in Chapter-9.

The analyses on the time and event denoting expressions showed that the amodal geometric properties exhibits more influence on how people refer to the sub-events rather than the effect of data-labels and the effect of the sensory modality although they exhibit main effects for some of the attributes (for "size", "month", "location" attributes in the Label-Set and for "Value", "Season" attributes in the Modality-Set)

Reference to the x-axis (*to the temporal domain*) and reference to the y-axis (*to the quantity*) are two different topics that need to be handled individually. The results showed that in the condition of the visual graphs with labels, the participants were equally good at referring to the both axes explicitly. However, in the condition of the visual graph without labels, the y-axis references for the explicit references were better recalled than the x-axis labels. For the haptic condition, both of them exhibited low accuracy in terms of the explicitness. Despite the low accuracy for the explicit references, the users were good at the vague description; it means that they are successful at acquiring rough idea about the quantity and temporal aspect. The verbal assistance system should also take this into account and explicit information for the selected content should be provided.

CHAPTER 7

General Remarks for Haptic Graph Comprehension

First chapter in this section focuses on the broader domain of haptic shape perception and recognition, whereas second chapter emphases on the more specific domain of event segmentation and description that is inherently dependent on the content of first chapter. The individual findings of the empirical studies were discussed in the interim discussions and in the chapter summaries. The topics presented in those two chapters are highly intertwined; to avoid over reporting, here I provide overview of the empirical investigations and only focus on the overall interpretation of the results and the general remarks concerning "*Haptic Graph Comprehension*" from both theoretical and applied-research perspectives.

Due to the perceptual differences between the visual and haptic modalities, the highlighted piece of information in the visual modality can be hidden when it is converted to the haptic modality. Additionally, the haptic condition requires the involvement of more memory constraints in order to integrate the explored local properties to the global shape during the exploration. Therefore, in the course of a development of fully automatic verbal assistance accompanied to haptic graph exploration, the differences and similarities in the two sensory modalities were investigated to detect and close the informational gap. For designing haptic graphs with assistance, it is necessary to determine the perceptually indistinct but conceptually important entities that may not be noticed or grasped without assistance.

In this section, three experimental studies by employing the single user paradigm (exploration of the graphs visually or haptically without having an assistant during the exploration session) are reported. The data were collected by means of various experimental methods, including the analysis of verbal protocols, referring expressions, speech accompanying gestures, sketches, and haptic exploration patterns. In the first two experiments presented in Chapter 5, the shape of the graph was investigated in a broader sense by not paying attention to the individual contribution of the various geometric features. The experiment reported in Chapter 6 was designed to focus on these individual effects (such as slope, angle, and curvature) and their contribution on the event segmentation and the event description as well. In the first experiment, blind-folded sighted participants explored the statistical graph and after the exploration is completed, they were asked to present a verbal description to a hypothetical audience and then produced sketch of the graph. The spontaneous speech accompanying gestures were also recorded. The graphs were presented in three conditions (the haptic graphs without labels, the visual graphs without labels and the visual graphs with labels). The focus of this study was to investigate the main effect of the sensory modality and the presence of data labels, with a secondary aim of taking glance at the effect of the amodal geometric properties on haptic graph comprehension. In the second experiment, visually impaired users participated to the study. The same haptic graph set was utilized. The aim of this study was to have preliminary insights about the use of the system and also to investigate conceptualization of the haptic graphs by the target group. Besides, the effect of graph-domain knowledge was also investigated by comparing two different user groups; the totally blind and partially blind users. In the third experiment, the same experimental paradigm that was employed in the first study was used. However, in order to have a systematic investigation of the effect of the amodal geometric properties on event (graph) segmentation and on event description, a different graph set was designed by controlling the parameters of slope and angle and used.

As aforementioned, Gestalt principles for visual perceptual organization play important role in visual graph comprehension (Pinker, 1990). Besides, the appropriate match between the form (*the graph*), and the content (*the abstract event*) facilitates the comprehension of the intended message, since in that case, the conceptually salient information is mapped into the perceptually salient features of that particular graph type (i.e the line of the line graph). However, gestalt principles may work differently for haptic perception (Frings and Spence, 2013) due to perceptual differences. To that end, it was hypothesized that this tight relation between the saliencies of the perceptual and the conceptual features might not be strong enough to enhance the same inference processes that facilitate reasoning. The empirical results reported beforehand were in favor of this hypothesis.

As previously stated, since the haptic space is not uniform, particularly at the far sides, the perception of orientation in haptic shapes was subjected to several illusions (Lederman and Klatzky, 2009). First of all, the oblique lines are considered as cognitively hard to process compared to the vertical and horizontal lines, as the explorer needs to take a reference with respect to the both axes. According to the findings, the oblique effect was observed only for the slight segments. Both steep and slight slope segments have 15˚ degree angle to their closest main axis. The steep segments are closer to the vertical dimension, whereas the slight segments are close to the horizontal dimension. However, the results (the number of the errors in the reference to those segments, in 6.4.7.1.C) showed that the participants did not have problem in comprehending the steep segments, but they do have for the slight segments. One can interpret this as the spatio-temporal reference frame hinders this negative effect for the steep segments. The diagonal's lines are claimed to be cognitively harder since taking reference to both vertical and horizontal axes are needed to estimate its relations to the other entities in the spatial layout (Gentaz and Hatwell, 1995). However, the vertical lines are not allowed in the statistical graph domain, and taking reference with respect to the x-axis only is sufficient and necessary to reach coherent interpretation of the graph line. Considering that steep segments have 75˚ angle to the x-axis, the difference is salient enough to infer that the slope is steep. On the other hand, the difference between the slight and the horizontal dimension is not salient and this may lead to make mistakes. The interpretation of the errors regarding the medium segments may shed a light on this. The medium segments have 45˚ angle to the both axes. Regarding medium segments, errors in the use of modifiers that refer to the size or the manner of the change were observed, but very few errors about the direction of the change were observed. On the other hand, for the slight segments, the participants exhibited more critical errors, which is about the direction (of the change) at the first place. This may indicate that the participants used the x-axis fundamentally to infer about the slope of the line. These results indicated that low slope values have higher chance of being misread or underestimated. Therefore the slight segments that carry conceptual importance should be definitely supported with verbal assistance.

As briefly introduced in Chapter 1.3, length is a special case for haptic graph comprehension. Similar to the case with the depth of 2D graph lines on the 3D virtual plane, length is also one of the irrelevant features. However, haptic exploration is performed through actions; hence the length seems to be inherently correlated with the exploration time (as supported by the results reported in *Chapter 6.4.5*). In more detail, within a same time interval in regards to the temporal information provided in the x-axis, i.e. from April to May, a steep line segment is represented with longer line-segment compared to less steep segments. If we assume that the user explores the graph at the (approx.) same speed, then the exploration time for the steep-segment will be longer than the slight segment. It has been also shown that time dominates space in spatial reasoning about layout and associations for many visually impaired people (Golledge, 1993). Combining these with the matters of haptic illusions, the length property might be considered as perceptually salient property although conceptually it is irrelevant. Therefore its interference with haptic graph comprehension should be investigated thoroughly. Further studies will address this issue.

The literature on visual shape perception states that the convex shapes are more discriminative than the concave shapes, thus they play a critical role in object segmentation (Cohen and Singh, 2007). On the other hand, the literature on haptic shape perception seems to be in favor of concave shapes although further studies are needed to reach compressive results, which -in particular- specific to the role of curvature on haptic shape perception explored via one finger or stylus. The empirical findings presented in Chapter 5 also indicated that the positive landmarks (concave points) were mentioned more. It should be noted that the stimuli set used in those experiments was not designed to check this parameter. On the other hand, the experiment presented in Chapter 6 employed a stimuli set suitable for this purpose. The findings regarding the perceptual saliency with respect to the acuteness of the turning angle did not provide any evidence supporting this hypothesis. However, a closer look at this issue revealed that the global minima points are mentioned more compared the global maxima points in the domain of the haptic graphs. This may indicate that the saliency of the minima/maxima points may not be only the perceptual outcome and but also under the influence of the conceptual factors too. However, this level of interpretation is highly speculative, and in order to reach conclusive understanding, further research should be performed.

The evidence obtained from the analyses of verbal descriptions, speech accompanying gestures and sketches in all experiments reported in this section demonstrated that the participants had difficulty in extraction of the metrical information from haptically perceived graphs. Combining this finding with the possible role of the spatial anisotropy on the perception of the haptic graph line, it might be suggested that metrical information is one type of content that need to be provided by the verbal assistance system. However, providing this information for all shape landmarks and segments would lead to a continuously speaking verbal assistance system, which would be impractical. Besides, haptic exploration is a rapid process, in order to catch up with the user's actions, refined but facilitator information regarding the conceptual aspects should be provided. The results also indicated that the haptic explorers are good at making rough estimation about both time and value dimensions, but they are having difficulties in relational reasoning. Thus, after segmenting the graph successfully into the meaningful parts, providing metrical information for only selected landmarks, which are critical for making relational judgments for that particular graph, may lead to a successful conceptualization.

Furthermore, the geometric approach for line shape representations concerning qualitative ascriptions that I employed in this dissertation (see Habel, Alaçam and Acartürk, under revision; Alaçam, Acartürk, Habel, 2014) proved itself as a useful tool. Besides, the qualitative ascriptions can be easily derived from the quantitative information of individual data values and their relation. They can be used in the decision of whether a verbal assistance is needed and also in the decision of its content.

As well as effective segmentation in the partonomic level, how to refer to a particular graphical entity (*segment*, *landmark*, or *their combination*) in the taxonomic level is also important for designing verbal assistance system. Referring to a segment as "there is a change", there is an increase" or "there is a slight increase" may cause different conceptualizations. Besides, *naming and labeling* have facilitator effects on shape perception and event segmentation as well, since they help readers transforming unfamiliar line combinations to familiar shapes. In that respect, recognition of haptic shapes and naming play a particular role in the design of verbal assistance system, due to their active role in linking the perceptual graphical entities to the conceptual graph-domain information.

Since the statistical graphs represent abstract events, I interchangeably used the terminology of *the graph schemata* and *the event schemata. Graph (event) schemata* guide a graph reader (the haptic or the visual) how to read a graph, how to to ignore the irrelevant features, and how to link the perceptual features to the conceptual graph-domain knowledge. Therefore, the facilitation of the appropriate graph (event) schemata is a key to a successful communication over graphs (e.g. Freedman and Shah (2002). Whether the correct schemata are activated or not, could be evaluated by looking at the errors committed by the participants. First, the experiment conducted with the visually impaired participants indicated that due to the lack of prior knowledge about the statistical graphs, the haptic exploration and comprehension were interfered with the use of irrelevant features (such as the depth).

The graph schemata have a central role in choosing appropriate reference frame when reading/exploring the graph. A reference frame is a coordinate system that is used to localize points in a standardized way. According to Lederman and Klatzky (2009, p.1449), multiple reference frames are simultaneously available to the haptic explorers, and a given task may require the use of a single frame or multiple frames. The use of reference frame is critical even for basic graph reading tasks such as localizing points, or more complex tasks such as making relational reasoning by using distances and directions. Statistical line graphs –visual or haptic- are left-to-right oriented graphs since they carry spatio-temporal information. This reference frame restricts the use of some spatial terminologies in this domain. For example, from an egocentric point of view, "backward" may point to back of the user, but from a spatio-temporal perspective, it corresponds to the leftward direction. In order to interpret the event depicted in the graph correctly, the spatio-temporal perspective should be adopted. This issue focusing on the misreadings during right-to-left exploration will be elaborated in *Chapter 9.3.1.* Besides, as mentioned before, the reference frame of a graph functions as a syntactic rule. According to the constraints of spatio-temporal perspective, the line of the statistical line graph cannot be vertical since there can be only one data point on the graph for each x-axis label. Therefore, such misreadings observed during a communication could be used as an indicator of the use of the wrong graph schemata and such cases highlight the need of a verbal assistance.

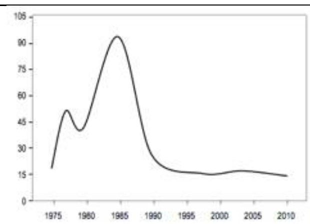
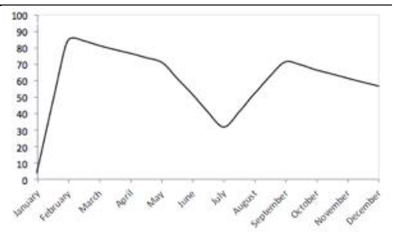
The vision impedance theory (Knauf, 2013) indicates that if the visual image contains details that are irrelevant for the inference, those features impair reasoning. This assumption seems to have a correspondence in the haptic modality. Due to frequent use of statistical line graphs in daily life that facilitates appropriate graph schemata, people are good at ignoring irrelevant features when reading visual graphs. However, the empirical findings indicated that the totally blind participants used the depth information when reading graphs and this interfered with their judgment until they updated their event schemata for the correct use. Although the effect of irrelevant features is beyond the scope of dissertation, the findings suggested that further research on that may provide fruitful environment for theoretical discussions and practical implications for designing verbally assisted haptic graph comprehension system.

Regarding every each of these research domains abovementioned such as visual graph comprehension, event segmentation & description as well as haptic object recognition, it has been suggested that bottom-up effects (such as the perceptual properties) and top-down conceptual factors (such as the event/graph schemata) have influences. The results presented here also exhibited that the conceptualization of the abstract events perceived through haptic graphs, namely haptic graph comprehension, is under the influence of both perceptual and conceptual features.

So far, I aimed to merge the results coming from the experiments reported in the previous two chapters without comparing them. A *rough* comparative evaluation based on the findings obtained through the analyses of speech accompanying gestures[31] and sketches appears to provide valuable insights. It indicated that the analyses of each of these experimental methods exhibited same pattern across the two experiments (5.3 and 6.4). Namely, in the both experiments, more directional dynamic gestures were produced than the static gestures, indicating that the line segments (rather than the shape landmarks) dominate the content of the communication over line graphs and also supporting the idea that line graphs emphasize trend conceptualization (Zacks & Tversky, 1999). Furthermore, the participants who explored the graphs haptically tended to produce more multi-directional gestures compared to the participants in the visual conditions. The results of the sketch analyses in the both experiments were in agreement about the superiority of the visual conditions over the haptic conditions in regards to what retains in the memory. The sketches drawn for the visual graphs exhibited more similarity to the original graphs than that for the haptic graphs.

However a *systematic* comparative analysis on linguistic modality between those experiments was not performed due to several reasons. One of the reasons was that the languages of the experiments were different (German vs. Turkish). As well as different languages, the granularity of the annotation, which was employed in the experiments, was different. The experiment presented in Chapter-6 was concentrated on how participants segment events and how they refer to them. The analysis of event segmentation requires more refined annotation scheme concerning the verbal descriptions with special focus on time denoting expressions. Therefore, the annotation scheme employed in Chapter 6 was broader than the former one used in Chapter 5. Furthermore, another difference between the graph sets is related to the conceptual type of the temporal axis variable, *x-axis*. Table 7-1 illustrates a sample from the each stimuli-set and sampling-related parameters. While the graphs in the Set-1 were in a yearly scale (as continuous scale), the graphs in the Set-2 were in a monthly scale (as categorical and ordinal).

Table 7-1 The differences between the two stimuli set in terms of the graphical features

	Stimuli Set-I (in Chapter 5)	**Stimuli Set-II (in Chapter 6)**
		
Value	Average yearly data	Average monthly data
Interval in X-axis label	Five years	A month
Sampling Rate/Label Units	5:1	1:1
The changes in amodal features	in more continuous scale than categorical	More categorical than in continuous scale

Besides, the graphs in the Set-1 consisted of the data samples collected for every year, but for the sake of the graph readability, the data labels were given within 5-years intervals. This inherently imposes that the changes between the two data labels (represented in the x-axis) do not need to exhibit a linear trend (or almost linear, the straight lines become curved due to smoothing procedure). Figure 7-1(a) illustrates an

[31] It should be noted that although gesture can be considered as a non-linguistic modality, they exhibit tight relation with language, as discussed in Chapter-3.

example of such a case; in this graph, the years between 2010 and 2015 were represented with data points on the graph, but not with labels on the x-axis. Here, the years 2010 and 2015 seem to be perceptually neighbor but not conceptually (see Freksa, 1991; 1992 for the details of conceptual neighborhood). Therefore, the change between these two years is not linear. On the other hand, if no measure was taken between 2010 and 2015, then the line only would provide connection between two points in a linear way. In this representation, 2010 and 2015 were both perceptually and conceptually neighbors. Monthly representation of x-axis (as in Table 7-1b) resembles to the latter condition.

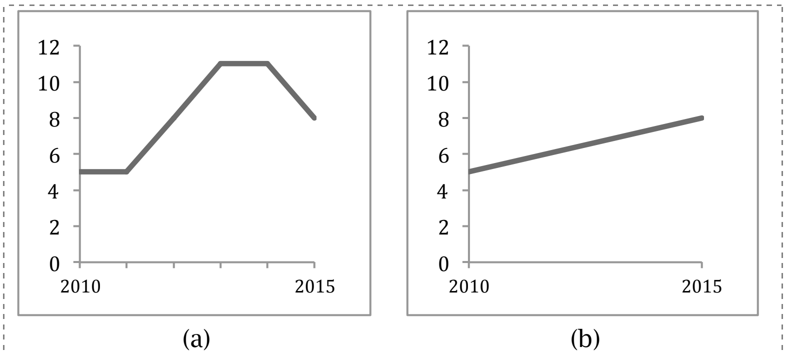


Figure 7-1 An illustration for the effect of minor data points

Furthermore, the changes in the y-values of the data points in the first stimuli-set were not systematically controlled (since they were retrieved from the original bird population database). On the other hand, in the second stimuli set, both the temporal interval and the values on the y-axis were controlled with respect to pre-defined categories (i.e. the steep segments are represented within one temporal interval, and they had 75° slope that corresponds to 80% of the maximum unit represented in the y-axis label). The changes on the graph lines in the first stimuli were on continuous scale, therefore the participants who explore those graphs may not have useful cues for discretizing them. Such linguistic categorization would also facilitate recognition. This kind of feature would be also considered important for both online graph comprehension and memory for later use. On the other hand, the scale on moths has 12 discrete categories and this could be used as a cue for the graphs in the second-stimuli.

In addition to the systematic differences between haptic and visual modalities discussed formerly, the findings also exhibited similarities between them regarding i.e. the location of the event boundaries, the gesture directionality and the use of time and event denoting expressions. These together with many, which were discussed in detail previous chapters (*5* and *6*), highlighted the main conclusion that regardless of the sensory modality, the graph shape is the most dominant factor that has an influence on the perception and recognition of the graphs (and of the abstract event depicted in the graph), and on the event segmentation and description.

Before going into detail on this, I would like to shortly refer back the layers of event representation and perception. The graphical communication can be investigated on three layers. First, there is a conceptual abstract event, which is depicted by a graph. Second, there is a physical representation, namely a visual or a haptic graph that represents the event. Last one corresponds to a human perceptual and cognitive system to comprehend the conceptual event represented through physical graph. Regardless of the modality, the first layer, the abstract event is the main source (for the perception) or the goal (for the comprehension) and the content of the information in this layer does not change with respect to sensory modality of graph reading. The psychical form also exhibits similarities, except the fact that textual information such as data labels is not suitable to be presented as haptically for the reasons explained before. This creates an informational inequivalence. Due to the differences elaborated on in detail earlier, the visual graphs may be perceived as objects, while the haptic graphs may be perceived as events through actions. This creates a computational (or functional) inequivalence. In brief, the sensory modality of exploration (i.e. visual vs. haptic exploration) exhibits different characteristics about perceptual information intake, thus leading to potential differences in conceptualization of the situations represented by the graphical entities.

During reading or exploring a graph, first the internal mental representation of the event depicted in the graphs is constructed. In spite of the perceptual differences among these modalities, the cross-modal studies showed that language, vision, hearing and touch can be (nearly) functionally equivalent depending on the adjustments in perceptual intake, since the mental representations exhibit amodal property, in other words, they are modality independent (Guidice, Betty and Loomis, 2011; Loomis and Klatzky, 2008; Loomis et. al, 2007, Avraamides et al, 2004). It has been suggested that the information coming from the spatial senses arrive at common region of the brain (the striate cortex and posterior parietal cortex, (Cohen and Anderson, 2004) and give rise to amodal spatial representations (Loomis and Klatzky, 2008) and involved in spatial processing. These amodal spatial representations are used in reasoning and communication over spatial layout (visit Chapter 1.5 for a reminder). A communication over a graph can be considered as a spatial reasoning and problem-solving task. Statistical line graphs are consisting of both modality dependent and independent features. Besides, they are spatio-temporal representations and they are based on spatial relations, therefore they employ wide variety of amodal geometric properties, such as local and global shape consisting of the curvature, the length, the orientation etc. And that modality-independent information can be accessed by vision and touch and well described by spatial language as well.

In this study, I mainly focused on the comprehension aspect rather than the perception; however they are highly intertwined as it has been discussed in Section-I and Section II in a distributed style. In a few words, due to perceptual differences between visual and haptic modalities, differences in the conceptualization of the graphs are expected. On the other hand, it is claimed that the reasoning and communication over graphs presented in different modalities rely on same amodal spatial representations, therefore similarities are also inevitable. So far little importance was given in the systematic comparative analyses concerning graph comprehension through these two modalities.

Concisely, the similarities observed in the communicational representations (the verbal descriptions and the speech accompanying gestures) w.r.t different representational modalities can be summarized with several conclusions from the results (see the interim discussions for more detailed information). First, the results indicated that the steep segments or the landmarks with the acute angles, in other words the graphical entities which are salient for both modalities were treated similar, in terms of their use as an event boundary. Furthermore, to how many parts the graph comprehenders segment the graphs (the visual or haptic), or the location of these event boundaries, and the general tendency to use more fine-grained partonomic levels also exhibited similarities across the modalities. Also the use of the event and time denoting expressions exhibited similar pattern for the use of many attributes in overall. Similarly for all conditions and all graphs more dynamic gestures were produced compared to the static gestures, and a closer look at the directionality of the gesture also revealed that more 1-directional gestures were compared to multi-directional gestures within each condition. Additionally, the preferred reference scope seems to be not affected by the sensory modality too. In accordance with the results of the experiments presented here, it can be concluded that in many cases the patterns for the visual and haptic modalities go hand-in-hand without exhibiting significant differences. The issues regarding event segmentation are highly dependent on the processing of amodal geometric properties and spatial relations. Altogether, the shape properties seem to exhibit more effect on how graph readers/explorers segment the graph and locate the event boundaries compared to the sensory modality of graph reading.

As Goel, Makal and Grafman (2004)'s study also proposed, relational reasoning is based on landmarks. The experiments indicated that the line segments were mentioned more compared to shape landmarks. Still, the landmarks did have role in event segmentation. For the graphs with labels, the segments were mostly mentioned with their corresponding x-axis labels (i.e. after July). For the graphs without labels, the salient landmarks were mentioned less explicitly (such as "after that decrease", "it *starts* to increase" or "after that *peak*"). The expressivity scores for both visual graph groups were higher, although they differ in their use of explicit or implicit (vague) reference for shape landmarks. On the other hand, the expressivity scores for haptically perceived graph were significantly lower than the visual counterpart. This highlights the fact that the use vague references may be still useful for establishing temporal relations as it is the case with *the visual graphs without labels*, but they should be generated thoroughly.

In this dissertation, I approached the domain of haptic graph comprehension from the perspective of human computer interaction therefore I do not go into detail of cognitive models regarding haptic graph comprehension. However, the experiments presented here also have potential to extend the existing graph comprehension theories, which address visual modality by shedding light into how it works in the haptic modality. In both Pinkers and Lohse's graph comprehension theories introduced in Chapter 1.3, the first step of graph reading is to explore the graph and construct a visual continuous array based on the perceptual properties available. During this process, the appropriate features should be selected and irrelevant ones should be ignored. The most available features of haptic shapes, which are explored through Phantom Omni, are friction, depth, shape, size, orientation etc.. The former two are irrelevant features and their use must be eliminated first, with the facilitation of the appropriate graph schemata. The results showed that the use of inappropriate use of features disappear after the graph reader updates the graph schemata. Then the inference process in accordance with the task at hand (i.e. comparing two points, deciding global max or general trend) follows. Although the exploration modality is different, it seems that haptic graph comprehension can fit to these general theories, but more detailed research is needed to understand the differences in the exploratory patterns for different tasks. In order to have conclusive results, one of the most preferred ways of investigating visual graph comprehension is to employ the eye tracking research paradigm. The counterpart of eye movement analyses in the haptic modality can be considered as the analyses of haptic exploration movements. Eye movements are central to the visual system. They are extremely fast, and metabolically cheap, they have a lower threshold for being triggered as compared to other motor movements. This makes eye tracking a very powerful and accurate tool to investigate cognition (Richardson et al., 2007). Although speed of eye movements and of hand metabolically are not same, the use of, for example mouse movements as an indicator of attended location is also one of the commonly used methods in HCI (Quek et al. 2002). Both of them provide sequence of attended locations (in haptic perception it is sequential this will be elaborated in Chapter 9), time of exploration for each region (are of interest), speed, back and forth movements between AOIs (*area-of-interests* or *region-of-interests*). Following the experimental paradigm of the study presented in Chapter 6, both eye movements of visual readers and haptic exploration movements of haptic explorers were recorded during online exploration. The comparative analyses would provide valuable insight about how much the models of graph comprehension explain haptic graph comprehension to what extent. This issue will be elaborated in the further studies by comparing eye movements and haptic exploration patterns.

Based on the empirical findings reported here, it might be concluded that this study extends the previous research on *event segmentation* and *graph comprehension,* which mostly address the visual modality, to the haptic modality. These empirical findings also provided valuable information to construct design guidelines (the heuristics) to build an effective and efficient verbally assisted system. Human computer interaction oriented interpretation of these experiments will be elaborated on in Chapter 10.

CHAPTER 8

The Contributions of Analysis Of Gesture-Language-Graph Relations

This chapter aims to summarize the findings concerning the analyses of speech-accompanying gestures focusing on the contribution of gesture analysis as a research paradigm for HCI Research. Although the investigation of how gestures are related with language is not the main purpose of this dissertation, the analyses appear to be promising and fruitful for this field as well.

The analysis of gestures provides insights about how a graph reader conceptualizes a graph in a way that a verbal data analysis is not able to provide. Gesture and speech are two complementary components of a single integrated system (McNeill, 1992), each modality has its own superiority in terms of conveying different aspects of the referent (Wagner et al. 2014; Hostetter and Alibali, 2008; Goldin-Meadow 2000; Cassell, 1998). Most studies of gesture-language research have only been carried out on the referents presented in visual or auditory modalities, giving little importance in haptic modality. As stated by Tversky (2011), graphs and gestures are visuo-spatial modalities and they share similar perceptual features to convey features like quantity, direction etc. This statement clearly addresses visual modality. However haptic exploration (as a sensory modality) and gesture production (as a communicational representation) also share common underlying mechanisms of motor movements (activation in the premotor and motor cortex concerning both actions). These associations and commonalities make gestures a valuable toolkit for the investigation of graph comprehension and for the human computer interaction research concerning haptic representations as well.

First of all, the analysis of speech-accompanying gestures appears to be a very useful complement for resolving ambiguities in the verbal descriptions. Gestures successfully carry shape and trajectory information and the boundaries of the referent and also the changes in the trends can be easily identified by looking at the gestures. As illustrated in Figure 8-1 (that is identical to Figure 6-15), despite of uttering same verbal descriptions like “it increases then goes stable”, the readers may have segmented the graph differently, which cannot be accessible from the analysis of verbal descriptions. For such cases, speech-accompanying gestures were used to clarify the situation and to identify the event boundaries (in the case if the participant produce gesture during his/her verbal description).

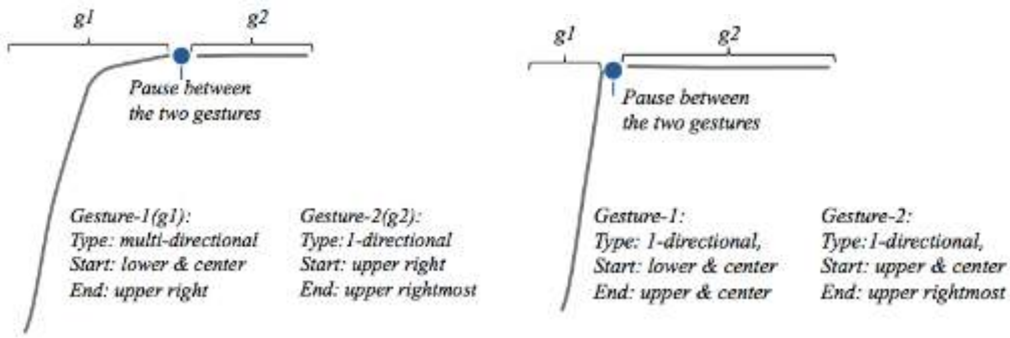


Figure 8-1 Gesture samples for the verbal descriptions presented in Figure 5

Furthermore, Hostetter and Alibali (2008) reported that the amount of action simulation increases as the stimuli moves from visual images through spatial images to motor images and as a result, more speech accompanying gestures are observed while describing those stimuli. (also see Hostetter and Alibali, 2007; Feyereisen and Havard, 1999). The results of the first experiment (presented in 5.3) indicated that the number of the participants who produced gestures in the haptic condition was higher than that in the visual conditions. In the experiments presented in 6.4, no significant difference between the modalities was observed for this parameter. The second graph-set consisted of the combinations of more categorical graph segments and landmarks, yielding differences in terms of amodal geometric properties. Speculatively, these differences may invite even visual graph readers to produce gestures. Further investigation of graph complexity may shed light into this issue.

Due to abovementioned commonalities in underlying mechanisms, the effect of the sensory modality on gesture production is expected, and this hypothesis is accepted as reported previously in the several empirical results. In brief, the findings indicated that in specific for the investigation of haptic graph comprehension and in general for the comparison of graph comprehension between visual and haptic modalities, the analysis of gestures play crucial role and gesture-language research might benefit from this association as well.

For example, the verbal descriptions focusing on how graph comprehenders segment the events depicted in the graphs into sub-events (see Chapter 6), did not exhibit overall differences between across two sensory modalities suggesting that people may segment haptic and visual graphs in a similar way. However, the complementary analysis of gestures provided evidence for event segmentation that occur in a different layer. Making a one-to-one relation between a word and a gesture is not a sound method since gestures are not discrete entities like words. They carry global meaning and this meaning represented in gestural movement is related with the meaning of speech part instead of pointing just one word (Cassell, 1998). According to the “Interface Hypothesis” (Kita and Özyurek, 2003), the preparation for language production requires organization of rich and comprehensive information into small packages that contain appropriate amount of informational complexity within a processing unit. This processing unit may correspond to a phrase for speech production, and the contents of a representational gesture are affected by the organization of these phrases. By taking this assumption into consideration, I focused on the relation between the gestures and the phrases for the sub-events. Briefly, I classified them into three categories; (i) *one-to-one* relation corresponds to production of only one gesture for one sub-event, (ii) *many-to-one* category involves the instances where more than one gesture were produced for one sub-event or (iii) *one-to-many* category corresponds to production of only one gesture for more than one sub-event. The result of this analysis indicated that while there is no significant difference across the modalities regarding the first two relations, third relation was only observed in the haptic graph condition. In other words, the haptic explorers tended to produce one combined big gesture that accompanies to several sub-events, which were mentioned with different phrases. This may indicate that event segmentation in linguistic level may be different than that in gesture production, and due to common underlying mechanism of haptic perception and gesture production (involvement of motor actions and motor imagery) the gestures might have been more influenced by haptic perception compared to visual perception of graphs. This result is in line with GSA framework that deliberately focused on the contribution of perceptual modality in gesture production. From the HCI point of view, this also supported the idea that the line of the haptic graph are perceived as a whole and highlighting the boundaries of the indistinct segments in respect to chosen partonomic and taxonomic level may have facilitating effect on event segmentation.

Furthermore, gesture can convey shape properties (such as global shape, direction, etc.) better than language can, especially in the cases the shape of the referent is complex and hard to verbalize. Several studies showed (Melinger and Kita (2007): Morsella and Krauss (2004) that "the rate of lexical gestures was higher when the line drawings to be described were less memorable, less describable and less verbally codable". In reference to events, in the visual conditions, people mostly produced more segmented and one directional gestures (straight). On the other hand, in the haptic condition the participants produced multi-directional gestures (that contain curvatures and/or multiple directional gestures in a smooth way) as well as one directional gestures. This may suggest that haptic perception facilitates as-a-whole interpretation compared to visual modality, and curvature points are regarded as a part of segment combinations in the partonomic level. But additionally, the results may also touch another issue regarding the effect of language and its interaction with sensory modality on gesture production. The results of a study conducted by McNeill (2000) on the effect of verb-framing on gesture production indicated that 'English speakers (satellite-framed language) break curvilinear paths into straight-line segments, Spanish speakers (verb-framed language) are likely to preserve the curvilinearity of the path, even when it is highly complex. Özyürek et al. (2005) also provided supporting evidence of that gesture of the same events differs across speakers of typologically different languages. Two experiments that I presented in this dissertation may be relevant in that respect. In one experiment (see Chapter 5), German speakers were participated to the study. German is a satellite-framed language, which means that path is encoded outside the verb. On the other hand, in the experiment presented in Chapter 6, Turkish speakers were employed, and Turkish language is considered as a verb framed language, in which path is encoded inside the verb and manner appears outside the verb. The results of the both experiments showed that the participants in the haptic condition produced more multi-directional gestures that preserve the curvilinearity than the participants in the visual condition without showing any difference at the pattern across two languages. Regarding curvilinearity of the produced gestures, it seems that not only language, but also the perceptual modality has an effect. Combining this with the abovementioned the "gesture-event relation" issue, the effect of haptic perception on the relations between the communicational modalities, namely gesture and language might be higher than the effect of visual perception due to high common neural activity in pre-motor and motor cortex regarding haptic perception and gesture production. Although this is highly speculative, it raises new questions for further studies.

A study presented by Trafton et al. (2006) suggested that people produce more gestures when describing spatial relations and locations (*geometric relations*) than when they refer to magnitude of spatial entities. As in line with previous research that indicates that gestures are not good at carrying exact metrical information, the empirical findings also supported this pattern. The absolute properties such as *direction* (i.e. increase or decrease), *shape* or *value* as a conceptual reference (i.e. "*the maximum*") were better represented in the gesture than the relative properties such as *size*, *value* (i.e. "*high value*").

Gesture analyses conducted in the scope of this dissertation also give useful hint about the interaction among gesture, language and graphs. According to McNeill (2005), the discourse context sets up a field of oppositions, and speakers are especially likely to gesture about information that contrasts with information already present in the field. How discourse context has an influence on gestures is one of the well investigated research fields (Hostetter and Alibali, 2008; McNeill, 2005). According to these studies, contrasting properties have high saliency, and this locates them to the focus of attention. This contrast presented in the discourse context is the key component in gesture production. As stated by Hostteter and Alibali (2008, p.506), "when speakers contrast two events, they are more likely to strongly simulate the contrasting elements; thus they are more likely to express these contrasting elements in gesture. However, a gesture that occurs in order to highlight a contextual differentiation will be particularly likely to take a form that captures the relevant difference". In Chapter 6, the effect of physical properties (particularly steepness) of a referent on gestural representation was investigated. The result indicated that the graph comprehenders produced more gestures when describing the steep segments (note that in this analysis, I only included separately/individually referred segments). For the graphs without labels (visual or haptic graphs), the verbal descriptions for the medium-slope segments were also accompanied by gesture more compared to the slight and horizontal segments. This points out that the steeper the referent slope is, the higher the possibility of gesture production is. In addition to the effect of sensory modality, the properties of a referent (i.e. steepness) seem to play a role on gesture production.

Gesture – speech mismatches is another interesting topic in the investigation of how gesture producer conceptualize a referent, in our case an event represented by the graph. The cognitive state that underlies mismatch involves having and activating two ideas on one task (Church and Goldin-Meadow, 1986). In order to activate a mental model for a negated concept, first the mental model of the concept itself needs to be activated. People may produce gesture that conveys different information from the information they convey in speech. These responses are labeled as "mismatches". It should be noted that the relation between a gesture and a speech, which contains negated content may or may not be labeled as mismatch. To illustrate, when referring to a steep segment, one can say that "it is steep" or "it is not slight". Steepness is one of the amodal geometric properties that the graphs possess. Therefore, being slight and steep can be considered as two extremities of one concept ("*steepness*") or they can be considered as two different concepts. Although it is worth mentioning, this distinction is not relevant for the current purposes. One thing that we can say for sure is that there is a contrasting property and the gesture-language relation differs for those contrasting cases with respect to sensory modality of the graph exploration. In this investigation of gesture, verbal statement in a negated form and referent (*the graph segment*), I used slightly different terminology than "*mismatches*"; the gestures which are (i) "congruent with graph (segment)" or (ii) "congruent with the part-of-speech." The reason of why the negated modifier is preferred seems to be due to possible recent exposure to the contrasting property in the earlier stages of the exploration and if the contrast is salient enough, it may affect the conceptualization of the current segment. In brief, while "congruent with referent" type of gestures was observed in the description of visual graphs (without labels), for the description of the haptic graphs, gestures which are congruent with negated modifier were observed. To exemplify, when referring to a slight segment as "it is not steep", a slight diagonal gestures was observed in visual condition (without labels). On the other hand, in the haptic condition, a steep diagonal gesture was observed. Additionally, the gesture type accompanied to the negated context also seems to be affected by the order of the statement. If the first sentence is affirmative then the gesture of the following negated sentence was also congruent with the referent. But in the opposite order, in other words, if the negated sentence was uttered first, a congruency with the negated modifier (the contrasting element with previously explored region) was observed. Because of the small sample size, the comprehensive statistical results cannot be reached as a result of this specific analysis. Still, looking at the data in a qualitative way, and having such contradictory examples brings up a really crucial point. The relation between the negated content and the produced gesture can uniquely provide valuable information to understand the effect of language, effect of the sensory modality and also their relation.

Although the data does not give evidence about the underlying mechanism of the interaction between the sensory modality and language, they provided evidence that there is an interaction.

Before elaborating the empirical results from theoretical perspective, it should be reminded that these frameworks predominantly address the relation between the visual or verbal representational modalities and the gestures (readers are encouraged to revisit Chapter 3 for more detailed review on the frameworks). From the gesture-language research perspective, the findings presented here provide counter evidence for the frameworks that claim that language and gesture are two different cognitive systems. On the contrary, the results suggested that both language and spatial representations play role in gesture production as in line with the argumentation of several other frameworks that approaches this issue from different perspective. Interface Model focusses on the contribution of language. GSA framework highlights the contribution of sensory modality. According to the recent cognitive modeling approach (Bergmann, Kahl and Kopp, 2013), the dynamic activation of visuospatial and propositional representations gives rise to speech accompanying gestures. Previous research already showed that the brain regions that are attributed to the processing of visuospatial representation are also responsible for constructing and processing of spatial models perceived through other sensory modalities (e.g. Knauf 2013, pg85; Fiehler et.al, 2008; Sathian et al 1997, Blake et al 2004; Amedi et al 2001; Prather et al, 2004; Zhang et al 2004). Therefore in that sense the empirical results reported within the scope of this dissertation are in line with this framework. Further, it may be considered as an extension that provides examples for integrating spatial representations, which is constructed through haptic sensory modality.

Moreover, verbally assistance system requires cooperation of two agents; explorer and verbal assistant (i.e. realized by an automated system). Haptic ostensive actions, which are elaborated in the upcoming chapter, can be also considered as exploratory gestures (Quek et al, 2002) and they provide deictic reference to the graph being explored to catch the attention of the verbal assistant to a referred region. From the HCI point of view, detecting and at interpreting such gestures based on the analysis of haptic exploration patterns would be very useful.

To summarize, gesture and speech have different affordances and can carry different aspects of the content in a different way. Therefore analysis of gesture adds another layer, which is already there but not usually investigated. This richer multimodal data may help to resolve ambiguities in the verbal data. The results obtained through the analysis of speech-accompanying gestures in relation to verbal content and the referent point out that gesture analyses provides new insights in the understanding of how graph comprehender conceptualizes the event represented by the graph. This may help designer to build more user-friendly and adoptable assistance systems. The results discussed above might be also useful to shed light into these triple relations from a theoretical perspective. Providing informationally or functionally isomorphic referent regarding the two different modalities can be tricky issue, however statistical graphs provide very structured information in language like manner and they are formalized visualizations by language-like conventions (Schmidt-Weigand, 2006) since they have syntactic, semantic and pragmatic levels like language (Kosslyn, 1989; Schnotz, 2002). One of the main property of the graphs is they concretize abstract concepts (such as bird population and temperature), this concrete representation can be presented visually and haptically in a nearly informationally isomorphic way. Investigation haptic perception is important to understand the sensory modality on gesture production and there is lack of research that address differences in gesture production in a systematic way comparing these two modalities by utilizing nearly isomorphic representations.

SECTION – III

Verbally Assisted Haptic Graph Comprehension System

CHAPTER 9

Dynamics of the Task Oriented Joint Activity

The fundamentals of the verbal assistance system for haptic graph exploration were already introduced in Chapter 2.3. This chapter aims to provide a more comprehensive overview of task-oriented joint activity dynamics from the perspective of assistive system design with several empirical investigations concentrating on the issues of "what to say", "how to say" it and "when to say" it. More specifically, these investigations address (i) the choice of the most appropriate perspective during the haptic exploration with respect to the communicative goal, (ii) the role of taking initiative in a dialogue, (iii) the semantic method for modeling the verbal assistance content with, and also (iv) the issues regarding the timing of the verbal assistance.

9.1. Speech Acts and Dialogue as a Joint Activity

A dialogue can be defined as an interaction that involves two or more conversational partners. All partners contribute to the conversation by listening and responding. Therefore they can be considered as participants of a joint and collaborative activity (Clark 1996; Clark and Wilkes-Gibbs, 1986). According to Lewis's definition (1969), a dialogue is a game of cooperation, in which both conversational partners "win" only if both understand the dialogue. The comprehension and production processes occur at the same time during a dialogue (Garrod and Pickering, 2004). More specifically, a dialogue involves the on-the-fly production of responses, the comprehension of elliptical utterances and turn-taking decisions. According to Traum and Hinkelman (*the Conversation-Acts theory,* 1992), successful dialogues require the existence of several conservation acts that can be grouped under four aspects: turn-taking, grounding, core speech acts (i.e. informing, yes/no questions, instruct, suggest, reject etc.) and also argumentation acts (i.e. elaborate, summarize, question-answer etc.).

The definition of a joint activity shows varieties. Sebanz, Bekkering and Knoblich (2006, p.70) describe joint activity as ''two or more individuals coordinate their actions in space and time to bring about a change in the environment''. They argue that such joint actions require merging the action plans of both interlocutors based on the shared representations. Predicting the interlocutor's intentions and actions and integrating them with one's own intentions and actions are the core processes in a joint action. On the other hand, observing a change in the environment is not obligatory according to Pickering and Garrod (2004). They define a dialogue as a joint action with or without having any explicit consequences of the dialogue on the environment, and the alignment of the internal representations occurs as a result of a joint action, regardless of the interchange of conscious information. This use of terminology for a joint action is also in line with Clark's definition (1996). In short, a dialogue is a form of joint action in which interlocutors have the goal of aligning their understanding of the situation.

As introduced before, verbally assisted haptic graph exploration is a *task-oriented* joint activity between two partners, a (visually impaired) explorer (*E*) of a haptic graph and an observing assistant (*A*) providing verbal assistance (Figure 9-1). The joint activities that I focus on here foreground the changes in *E*'s mental representations. In order to reach this goal, *E* and *A* have to establish a common "understanding of what they are talking about" (Garrod and Pickering, 2004).

Verbally assisted haptic exploration has a dialogue-like character:

(1.a) *A* has to synchronize language production with *E*'s hand-movements in a turn-taking manner.

(1.b) The quality of the verbal assistance depends on whether the appropriate referential and co-referential links have been established.

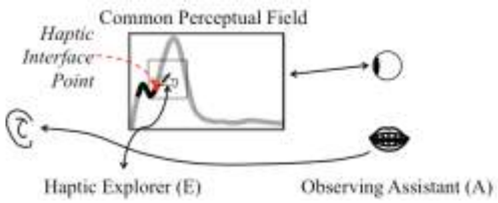


Figure 9-1 Assisted haptic graph exploration, a joint activity

A and *E* share a common field of perception, namely the haptic graph, but their perception and comprehension processes are very different. As well as having those differences, they also have different activity roles; *a request maker* and *an assistance provider*. Using similar linguistic structures or sharing the same physical perspective may help them adjusting their perspective with respect to each other.

The success of a joint activity in general, and also the success of *A*'s verbal assistance in particular, depends on joint attention (Sebanz et al., 2006) and on the alignment of the interlocutor's internal models (Garrod and Pickering, 2004). *E*'s internal model of the activity space, i.e. the haptic graph and *E*'s explorations, is perceived via haptic and motor sensation, whereas *A*'s internal model of the same space is based on visual perception. Therefore similarities and differences in their conceptualization play the central role in aligning at the level of the situation-model[32].

To be assistive, *A* should verbally provide *E* with content (i.e., conceptual representations), that is difficult to acquire haptically. Haptic explorers' contributions to the dialogue and their exploration movements occur in a concurrent manner. Thus, for the observing assistant, the referring expressions produced are accompanied with the explorer's current exploration location on the graph. In other words, each one of *E*'s exploration movement evokes a referential link deictically —analogue to Foster, Bard, Hill, Guhe, Oberlander and Knoll's (2008) haptic ostensive reference. And thus, a common ground is established and the given-new contract between *E* and *A* is fulfilled (Clark and Haviland, 1977; Clark and Brennan, 1991). When it is eventually *A*'s turn, s/he is expected to provide the most helpful and relevant information for *E* at that particular moment. To maintain the common ground, *A* has to synchronize his/her language production with *E*'s hand-movements in a turn-taking manner. That is because the quality of the verbal assistance depends on establishing the appropriate referential links.

Furthermore, in a dialogue, the coupling between the production and comprehension processes is required since the utterances the speaker produces becomes part of what

[32] According to Van Dijk, and Kintsch (1983), situation models are the mental representations of events (actions, states and situations) evoked in perceptual representations (see Zwaan and Radvansky, 1198 for a general overview).

the listener has to comprehend (Linnell, 1998). The listener's role in the context of a joint activity is not just hearing what the speaker has uttered; s/he also needs to predict the speaker's actions and intentions in order to respond accordingly (Garrod and Pickering, 2009).

Besides, the meaning which is being communicated in a dialogue is not just dictionary meaning. Its successful interpretation depends on the agreement between the interlocutors. A number of studies emphasized that temporarily created expressions constitute a norm in dialogue and are used extensively (Brennan and Clark, 1996; Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987). For example, in a dialogue about a graph, the use of onomatopoeic words (such as "*tick tick*" or "*hop hop*") by the explorers would be an intriguing case: Their meaning is highly dependent on the context. Hence, there needs to be an agreement between the interlocutors on their meaning if they are to successfully use such non-dictionary words when communicating with each other (This is discussed in more detail in 9.2).

In the conceptual design of verbally assisted systems, an explorer (*E*) produces a question and expects an answer of a particular type; and accordingly, an assistant (*A*) hears the question and has to produce an answer of the expected type by either giving short statements, such as affirmations or full informative responses. To illustrate this, the question asked by the explorer in (1) is a request for navigational assistance. The assistant should understand that the explorer is not using the ego-centric frame of reference and should provide a response accordingly. The second example is about reference resolution, which is also another crucial component in an assistance system.

(1) (E) Can I go backward here? (*Frame of Reference*)
(A) Yes, you can, it continues to the left.

(2) (E) Is this lower than the other one? (*Reference Resolution*)
(A) Yes, it is slightly higher than the first one.

The help request and the assistance form an *asymmetric* joint activity: the partners have different activity roles (explorer vs. assistant). As a part of experimental procedure, the partners were told that *E* should initiate the help request and *A* should provide help based on the explorer's need. Although the dialogues accompanying the haptic explorations are—in principle—mixed-initiative dialogues, explorer-initiatives are the standard case in such an assistive design.

9.2. (Implicit) Common Ground, Situation Models and Alignment

Research on how language is used in dialogues (Watson, Pickering and Branigan, 2004; Branigan, Pickering and Cleland, 2000; Clark and Wilkes- Gibbs, 1986; Garrod and Anderson, 1987) has shown that interlocutors align representations in the course of a dialogue at several levels of representation including the conceptual (Garrod and Anderson, 1987), lexical (Clark and Wilkes-Gibbs, 1986), syntactic (Branigan et al., 2000) and articulation levels (Giles, Coupland and Coupland, 1991). The results of these empirical studies indicated that participants usually achieve alignment without realizing what they are aligning. Based on the huge amount of empirical findings, it can be assumed that this alignment step forms the basis for a successful dialogue and misunderstanding occurs when the interlocutors fail to align with each other.

Common ground is used to refer to the shared knowledge about a situation between communicative partners (Clark and Marshall, 1981) and sharing a common ground is a key to establishing successful communication (Clark and Wilkes-Gibbs, 1986). According to the common ground theory, the "establishment of common ground involves a good deal of modeling of one's interlocutor's mental state", that is, interlocutors can assume what their partners know, think or intend based on the evidence (linguistic or non-linguistic) at hand.

According to Clark and his colleagues (Clark, Schreuder and Buttrick, 1983; Clark and Marshall, 1981; Gibbs, 1981; Clark, 1979), the establishment of a common ground in a dialogue is dependent on several sources of information: Firstly, the contribution of perceptual salience (as also called perceptual evidence) is highly crucial. Furthermore, speaker assertions are also important as these assertions become a part of the common ground, and also a part of the interlocutor's knowledge. Similarly, the speaker's presuppositions, goals and intentions are regarded as crucial since they are important cues in the interpretation of the speaker's utterance by the listener.

As discussed before, the common ground theory (Clark, 1996) proposes that interlocutors intentionally convey meaning regarding each of these layers mentioned above such as conceptual, lexical, syntactic etc. In addition, the concepts of shared knowledge and belief are foregrounded. The difference between the "common ground" and "implicit common ground" theories is originated from whether the alignment process is intentional or not as well as from the nature of the common ground knowledge. According to the "implicit common ground" theory, the alignment of the situation models is of very high importance to establishing a common understanding. Ity proposes that the "alignment of situation models follows from lower-level alignment, and is therefore a much more automatic process" (Pickering and Garrod, 2004, p. 178). The alignment can, for example, occur on linguistic levels in a resource-free (automatic) way, since hearing an utterance highlights or contributes to a particular aspect of a situation model and in return hearer's response to that particular utterance more likely conveys that highlighted aspect.

A mechanistic model of language processing for a dialogue, called the interactive alignment model, proposed by Pickering and Garrod (2004) assumes that as the dialogue proceeds, the interlocutors come to align their linguistic representations at many levels, i.e phonological, syntactic and semantic levels. Figure 9-2 shows the schematic of the "Interactive Alignment Model". As can be seen in the depiction, production and comprehension are tightly coupled and alignment occurs at different levels automatically. In particular, the alignment at one level facilitates the alignment at other levels, promoting mutual understanding between the communicative partners. If both interlocutors' situation models are aligned, the communication can be stable. This means that a successful dialogue may occur when A's message is consistent with what B's comprehension of that message is (Garrod and Clark 1993). Therefore, the use of aligned linguistic structures (that is, 'local' alignment) facilitates the alignment at the higher-level semantic and pragmatic representations (that is, 'global' alignment) (see Pickering and Garrod, 2004 for a detailed discussion).

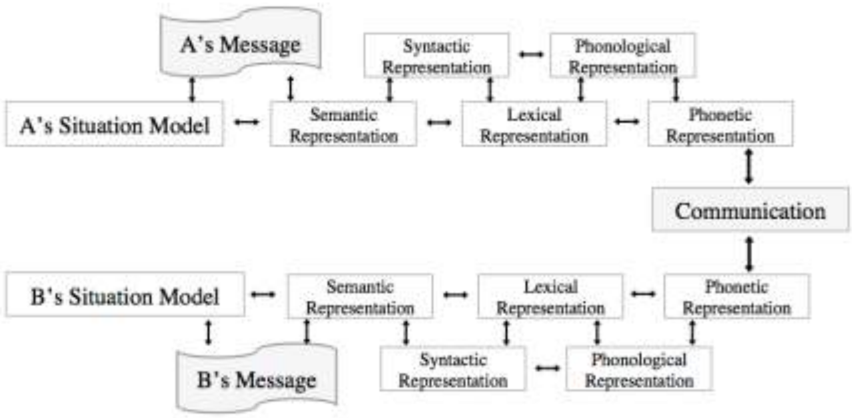


Figure 9-2 The schematic representation of the "Interactive Alignment Model" (retrieved from Pickering and Garrod, 2004 and redrawn based on the original)

This automatic and largely unconscious alignment process depends on simple priming mechanisms that operate on the different representational levels. Priming and mutual understanding is central to this theory. Furthermore, the information that is used to construct a situation model remains a part of the common ground for both the production and comprehension processes. The interlocutors have access to the same foregrounded information, and the communication proceeds based on this interaction; one's contribution becomes parts of the other's knowledge and vice versa. As stated in Pickering and Garrod (2004), this does not mean that the interlocutors' situation models are exactly the same; more shared information leads to more aligned situation models. This account claims that developing and sustaining a full common ground intentionally is a costly process and not obligatory for a joint activity. On the other hand, the alignment through implicit and automatic mechanisms is unconscious resource-free and may lead to a full (conscious) common ground when it is necessary, i.e. in the case of apparent misalignment.

9.2.1. Alignment at Lower-Levels

The alignment of lexical processing in a dialogue was specifically investigated by Garrod and Anderson (1987), and by Clark and his colleagues (Brennan and Clark, 1996; Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992). These studies showed that, when referring to objects in the focus of the communication, the interlocutors tend to develop similar set of referring expressions. Moreover, the expressions become shorter and more similar when repeating with the same interlocutor. This alignment may be on a syntactic level, such as repeating phrases, or on a lexical level, like using the same nominal or prepositions etc.

Clear evidence for syntactic alignment was presented by Branigan et al. (2000). Their findings indicated that the representations are activated through priming and not just by explicit information shared during the language production and comprehension cycle. That is, alignment and syntactic priming are closely related and the alignment is dependent on the syntactic priming in general. Moreover, interlocutors align at the level of articulation as well. Besides, Giles et al. (1991) indicated that the alignment might also occur when it comes to the accent and the speech rate.

As a result of the alignment process another useful toolkit for an effective communication, the routinization, seems to emerge: If an interlocutor uses an expression in a particular way, this use of the expression may become a routine for the purposes of that conversation. Repeated expressions convey aligned information at many linguistic levels. It has been suggested that interlocutors use same the expressions to refer to the same things if their situation models are similar on the semantic and syntactic levels (Pickering and Garrod, 2006; Garrod and Pickering, 2004). For the design of task-oriented assistive system, this theoretical issue is particularly beneficial since it foregrounds the importance of repetition, of consistency, and of the introduction of a topic or terms that facilitate the construction of an appropriate situation model.

9.2.2. Alignment at Higher-Levels

Watson, Pickering and Branigan (2004) showed that alignment extends beyond the language faculty and that the situation models can be aligned at different levels of conceptual representation. For instance, interlocutors need to align their frames of reference to describe object locations in a scene.

A reference frame is an axial co-ordinate system. Speakers may employ different types of reference frames when referring to a referent; these are the absolute, relative, and intrinsic frame of references. The absolute reference frame locates a referent in an environment, i.e. by using the cardinal directions. The relative reference frame locates a referent with respect to the observer's viewpoint. The intrinsic reference frame locates a referent with respect to its directional features (see Levinson, 2003 for a more detailed overview regarding spatial reference frames). Coming to an alignment at this level is particularly important in order to understand the spatial layout of the referents. The results of a study conducted by Carlson-Radvansky and Jiang (1998) demonstrated that the choice of reference frames is influenced by low-level priming. In particular their results indicated that it can be influenced even by activating only one axis of the chosen reference frame. This suggests that reference frames constitute a holistic representation. Furthermore, the results of Schober (1993, 1995) illustrated that interlocutors are also influenced by each other's choice of reference frames and that they are prone to align at this level, too, as the dialogue proceeds.

Boroditsky (2000) reported that the spatial reference frame also has an influence on the use of the temporal reference frame. According to his findings, the reference frame preferred when describing the spatial layout is also preferred to describe the temporal structure, hence highlighting the shared representation of spatial and temporal aspects in situation models. This result demonstrated that the abstract aspects of situation models also have an influence on coming into alignment.

As elaborated on in 9.3.1, the explorer and also the assistant may choose and switch among more than one type of reference frames during the haptic exploration of a graph. In addition, the efficiency of the chosen reference frame depends on the task at hand. Therefore, this information forms a basis for providing mechanisms that make switching between frames of reference easier and prime the appropriate frame of reference.

The choice of reference frame is dependent on the user's choice of perspective to a referent. The "Spatial Grounding Hypothesis" proposed by Beveridge and Pickering (2013) investigates two forms of perspective-taking (*the action perspective* and *the spatial perspective)* and the choice of one over another in action language comprehension. This theory claims that action simulations are grounded in spatial context. If a relative frame of reference is adopted, one can use an egocentric spatial-perspective (from one's own point of view) or an allocentric spatial-perspective (from another's point of view). More specifically, some embodied cognition accounts claim that when people read i.e. descriptions about an action in a scene, they adopt an embodied agent's perspective, namely the perspective of the person or object that performs that particular action (Zwaan and Taylor, 2006; Barsalou, 2009). Furthermore, according to the Spatial-Grounding Hypothesis, the alignment of situation models premises the interlocutors' aligning with respect to the spatial perspective and at the choice of reference frame. This hypothesis also addresses the effect of having a partner on the possibility of shifting spatial perspective. As this account and also the empirical research suggested, speakers usually adopt an allocentric spatial perspective in the presence of a (potential) agent who acts on the spatial layout that is being communicated about (Mazzarella, Hamilton, Trojano, Mastromauro and Conson, 2012; Tversky and Hard, 2009). This is discussed in more detail in 9.3.1 with empirical evidence obtained from the studies regarding the graph domain.

The situation models contain integrated knowledge about events, whereas mental simulations are more about online action-perspective taking regarding a particular action (Beveridge and Pickering, 2013). Rather than the form of action language (modality-specific representation), these mental representations (the situation models, the mental simulations etc.) are retained in memory and are used for on- and offline comprehension (e.g. Johnson-Laird and Stevenson, 1970). Beveridge and Pickering (2013, p.7) proposed that "this 'nesting' of action simulations within situation models is what links spatial- and action-perspective taking in language."

To sum up, the "Interactive Alignment Theory" assumes that interlocutors operate on common representations and that their situation models come to align as the conversation proceeds (Pickering and Garrod, 2004). Therefore, the production and comprehension cycle is a key concept for this theory. Speakers may adopt their partner's perspective or align their utterances on the syntactic or semantic level on the fly.

This model was developed to explain the tightly coupled interaction in spontaneous dyadic conversation between interlocutors with similar activity roles and with short contributions. Such conversations are regarded as basic and less complex conversational settings (Linnell, 1998; Clark, 1996). As it has been discussed by Pickering and Garrod (2004), the process of alignment becomes less automatic for more complex conversational settings such as video-mediated conversations, multi-party discussions and tutorials. In such cases, it has been claimed that instead of relying on implicit and automatic alignment, speakers intentionally construct situation models by taking their interlocutor's situation models into consideration. A verbally-assisted haptic graph exploration task is also another example of a "non-ideal" conversational setting that requires conscious contribution in order to accomplish the task and stay aligned with respect to changing communicative goals that inherently employs different situation models.

The verbally assisted haptic comprehension system has two main endeavors: helping the users in being able to navigate on the physical graph (*the technical aspect*) and in being able to grasp the abstract information (*the conceptual aspect*). The main goal of the system is to assist graph readers in enhancing their abstract graph-domain comprehension. Therefore, the dialogues exhibit asymmetric roles in this context; the graph explorers make a request when they need assistance by asking a question that requires additional information or by making a statement that requires affirmation or clarification. Making a request and providing information are two different activity roles (Duran, Dale and Kreuz, 2011). The results of a study presented by Yoon et al. (2012) indicated that speakers who request something from their partner in such a communication task mostly prefer using the allocentric perspective (from his/her partner's perspective) and exhibit a different pattern when they are asked to give information. By doing this, partners minimize both their own and their partner's effort. This kind of behavior follows the principle of *least collaborative effort* by Clark and Wilkes-Gibbs (1986). Schober and Brennan's study (2003) indicated that this effort mostly occurs when partners have asymmetric activity roles, as it is the case in verbally assistive systems.

Furthermore, the linguistic characteristics of the utterances produced by the explorer may have an influence on the verbal assistant and sub-consciously prime him/her to use the same linguistic structure. There are studies that indicated that computer systems, which align their sentences with their interlocutors, are accepted more in terms of user-friendliness and effectiveness (e.g. Branigan, Pickering, Pearson and MacLean, 2010).

9.3. Empirical Investigations

A graphical representation itself (without the presence of an assistant) is considered as a form of communication and thus it can also be discussed under the aspect of alignment and common ground. For a successful communication, the message that is intended to be conveyed with the graph (by the producer or the publisher of the graph) needs to be aligned with the comprehended message (by the graph reader). If chosen appropriately, the graph itself carries syntactic, semantic and pragmatic aspects of the message (see Chapter 1.1 for the revisit of form-content match). However, the verbally assisted system adds another user to this communication: In the proposed design, the assistant becomes a messenger between the producer and the graph reader and helps the reader successfully understand the abstract concept depicted by the graph.

As discussed densely in the previous parts of this section, establishing a common ground is a key to successful collaborative activity. However, due to the assigned roles of this particular task-oriented activity, the haptic explorer is the one with incomplete mental representation whereas the verbal assistant has complete knowledge about the abstract event. This asymmetry in the activity roles highlights the necessity of that the assistant needs to react with respect to explorer's request. Based on the existing literature, the common ground can be achieved by employing the abovementioned implicit and explicit mechanisms.

The empirical investigations presented in the following explore online haptic graph comprehension in the presence of verbal assistance. This line of research contributes to our understanding of the dynamics of this particular task-oriented joint activity between haptic explorer and verbal assistant. These investigations are presented under the four titles; (i) aligning and switching reference frames, (ii) the role of taking initiative, (ii) referring to shape representations and a semantic annotation method for modeling the verbal assistance content, and also (iv) interpreting haptic exploration patterns to infer the right timing for assistance. Looking at each of these issues contributes our understanding of both the theoretical issues regarding joint activity dynamics and perspective-taking. It also helps to understand the practical issues that need to be considered in order to design an effective and efficient system that smooths the communication and enhances/ facilitates haptic graph comprehension for visually impaired people. The results presented here were partially reported in the several publications (Alaçam, Habel and Acartürk, 2015; Alaçam, Acartürk, Habel, 2015; Alaçam, Habel and Acartürk, 2014; Alaçam, Acartürk and Habel, 2014).

9.3.1. Aligning and Switching Reference Frames

This empirical investigation[33] focuses on the participants' use of reference frames and spatial perspectives. In the experimental setting described in Figure 9-1, *E* and *A* have access to the same absolute graph frame via different perceptual modalities. The comprehension of statistical line graphs requires a canonical interpretation of directionality, i.e. the interpretation of the line segments as left-to-right directed lines as well as the integration of temporal and spatial aspects into an abstract spatiotemporal representation (*the spatiotemporal-perspective*). The haptic explorers' spatial reference frames, however, are usually induced by the exploration movements, leading to an *action-perspective*. For instance, right-to-left exploration (*backward)* may result in *unusual* (*misinterpreted*) referring expressions, e.g., an *increase of the graph line* may be expressed as a *decrease* from the backward perspective (see Figure 9-3: "sp2-sp1" line segment) due to *misuse of action-perspective*. From the viewpoint of the assistant (*A*), *E*'s movement-induced frame is visually accessible. However, *A*'s simultaneous visual exposure to the graph may favor a spatiotemporal perspective that yields graph-domain descriptions (instead of descriptions tailored to the actions performed by E during haptic exploration). The *action-perspective* and the *spatiotemporal perspective*, which are clearly distinct in right-to-left exploration, are barely distinguishable in left-to-right exploration. For example, an instruction like "*you are at the left-end"* or "*this is the start point"* is compatible with both perspectives. Hence they do not give a hint about the adopted perspective or whether a perspective switch has occurred.

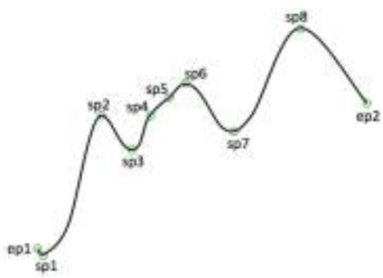


Figure 9-3 Shape landmarks for a sample graph

In their very compressive review, Beveridge and Pickering (2013) claim that the ego-centric or intrinsic reference frames are underspecified as terms and that they are not accurate enough to understand the speaker's preference, especially when the action-perspective is taken. Therefore, they proposed a more fine-grained terminology to better explain the participants' perspective within the *embodied cognition approach* (e.g. Brunyé, Ditman, Mahoney, Augustyn and Taylor, 2009; Bergen and Chang, 2005). They distinguish between *embodied agents*, *embodied observers* and *non-embodied observers*. Before going

[33] This analysis was published in Alaçam, Habel and Acartürk, 2015.

into the details of the embodiment-oriented terminology for a verbally assisted system, the use of pronouns as an indicator of the taken perspective is elaborated on here. The choice and the interpretation of *potentially self-referential* ("*you"* and "*I"*) vs. *non-self-referential* ("*it"*) pronouns seem to be determined by the perspectives the participants take. The existing empirical research indicated that if the sentence involves a self-referential pronoun ("*you*", "*I*") in the agent role, people tend to adopt the embodied agent's perspective (Willems, Hagoort and Casasanto, 2010; Pulvermüller, 2005; Pulvermüller, Shtyrov and Ilmoniemi, 2005; Hauk, Johnsrude and Pulvermüller, 2004). For the cases in which the sentence contains a non-self-referential (*third-person*) pronoun, the results of the studies are not conclusive: Some findings (Tomasino, Werner, Weiss and Fink, 2007; Buccino, Riggio, Melli, Binkofski, Gallese and Rizzolatti, 2005) indicated that people adopt the embodied agent's perspective, whereas other showed that people adopt an embodied observer's perspective (Papeo, Corradi-Dell'Acqua and Rumiati,2011; Brunyé et al., 2009). On the other hand, if the sentence contains self-referential pronoun in a thematic role, people tend to adopt the perspective of the thematic role assigned to that pronoun, not the perspective of the agent (Glenberg and Kaschak, 2002). Furthermore, a dialogue with potentially self-referential pronouns ("I" and "You" in a request/response cycle) is claimed to be complex since the partners need to maintain a consistent perspective for each pronoun (Pickering, McLean and Gambi, 2012). However, this complexity does not lead to problematic case in the design of the verbally assistant system for haptic exploration. The haptic explorer is always in the position of the embodied agent/observer since s/he is the one who act on the graph line, and the verbal assistant is always in the position of the observer. As for graph comprehension, visual graphs may be conceived as *non-embodied representations* that trigger the adoption of the *non-embodied observer* perspective. In contrast, due to active exploration, haptic graphs are forms of *embodied representations* that may bring about the adoption of both the *embodied-agent* perspective (e.g. I go up) and the *embodied-observer (*e.g. the population increases) perspective by the explorer. In the experimental settings employed in this investigation, the verbal assistant did not only have access to the visual graphs (that trigger the non-embodied observer perspective) but they also had access to the explorer's actions on the graph, which may lead to a switch from the *non-embodied-observer* perspective to the *embodied-observer* perspective. In the case of joint haptic exploration, the assistant's preference for the subject pronoun may lead the explorer to using a proper reference frame.

9.3.1.1.Experimental Setup

In this investigation, the data gathered from the Experiment IV was used. Here, I only provide the overview of the experimental setup, see Chapter 4.4.4 for more detailed information regarding the experimental setup and the data annotation approach.

Participants: Thirteen subject pairs (in each pair, an observing assistant A and a blind-folded haptic explorer E, Mage=25.3, SD=3.27) from Middle East Technical University collaborated in the exploration of haptic line graphs. All of the participants were native Turkish speakers.

Procedure and Materials: *A* and *E* were located in separate rooms such that they could not have visual contact. Their only way to communicate was via speakers. The pairs were presented, in random order, with two haptic line graphs in the familiarization session and with five graphs in the experiment session. The participants were informed about the graphs representing bird-population data of various species (Graph-Set I, see 4.3.1). *E* explored the graphs haptically. *A* was able to observe the graphs and the current location of *E*'s exploration—depicted by an animated point marker on the visual graph—on the screen in front of him/her. However, haptic pointing was possible only for *E*.

The pairs explored informationally equivalent graphs in different modalities of presentation (haptic and visual). Finally, *E* was instructed to explore the graph and ask for verbal assistance when needed by turning the microphone on, whereas *A* was instructed to provide verbal assistance shortly and plainly, when asked by *E*. No time limitation was imposed. After the experiment session, both participants (*E* and *A*) were asked to present single-sentence verbal descriptions of the graphs to a hypothetical audience independently. In this analysis, I focus on the dialogues between *E* and *A* during the joint activity without considering the post-exploration verbal descriptions.

9.3.1.2. Results

The available body of empirical evidence has convincingly demonstrated that both bottom-up perceptual and top-down conceptual factors (i.e. communicative goal, prior knowledge etc.) play a crucial role in graph comprehension (see Chapter 1). The choice of perspective is also subject to this dichotomy (as mentioned throughout this chapter). In this section, I first report on the cases that exhibit *misuse of the action-perspective.* After that, the findings concerning the linguistic alignment and the effect of the communicative goal on the choice of reference frame are presented.

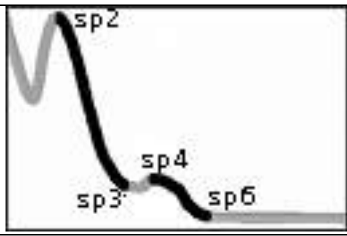
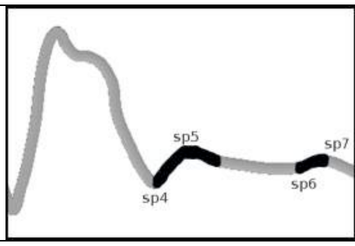
A. Misuse of the Action-Perspective

The utterances produced during the joint activities (1588 individual utterances in total) were classified into three categories: (i) request-response pairs (63.9%), (ii) alerts initiated by *A* (5.9%,) and (iii) think-aloud sentences (30.2%). Six of the thirteen haptic explorers produced utterances that were clear cases of the above-mentioned issue of *misinterpreting descriptions* during right-to-left reading. These utterances were mostly think-aloud sentences uttered during the exploration of the first two graphs (cf. the familiarization phase). It should be noted that self-talking was up to the participants' preference, therefore the participants, who did not think aloud, still might have experienced such a misconception, however there is no verbal evidence. After one or two graphs, all haptic explorers who exhibited this *unusual* reading pattern gained sensitivity to overcome the movement-induced action-perspective (this term is adopted by Beveridge and Pickering, 2013). They used graph-domain spatiotemporal perspective by focusing on left-to-right directed graph reading, and they switched between the perspectives when necessary.

Table 9-1 contains two think-aloud excerpts produced by two different haptic explorers. The utterances show how the participants used (in this case, misused) the *action-perspective* for conceptualizing the graph-domain information during right-to-left reading of the graph. For example, in Excerpt-1, the user misread two decreasing segments of the graph line and referred to them as increases during self-talking. This misuse, exhibited explicitly by six of the thirteen participants in total, indicates that in the early stages of haptic graph exploration, the action-perspective may dominate over the spatiotemporal-perspective, which is necessary to interpret the graph-domain information accurately.

Qualitative analysis of the dialogues pointed out that the participants developed some strategies in order to avoid the possibility of a misreading when they performed backward (right-to-left) graph reading. Some participants exhibited a tendency to use terms without direction or temporal information (e.g. by referring to the segments by *fluctuations* or by counting the salient landmarks). The more cautious strategy adopted by some of the participants was switching to silent mode during the backward-motion until they reach to the start point of the graph or the segment where they can begin left-to-right reading again.

Table 9-1 Two examples of "unusual" descriptions

Location	Exploration Direction	Utterance	Explorer's Location on the Graph
		Excerpt-1	
sp6–sp4	Right-to-Left	"There is an elevation here"	
sp3-sp2	Right-to-Left	"Then again, it increases"	
		Exercpt-2	
sp6 - sp7	Left-to-Right	"Here, in the last part, there is an elevation."	
sp6 - sp5	Right-to-Left	"We have another one here"	
sp5 - sp4	Right-to-Left	"A descent"	

B. Linguistic Alignment

The request-response pairs (N=400, but 6 were excluded since they had to be repeated due to a problem with the audio) were analyzed in more detail in terms of the alignment between interlocutors. 189 of 394 requests were answered with Yes/No statements. Those affirmative responses were excluded from the analyses because they do not explicitly exhibit the assistant's choice of perspective. For 171 of the remaining 205 utterances, the assistants used the same referential form (self or non-self) as the haptic explorers, indicating that the observing assistants showed a tendency to lessen the explorers' burden of aligning by using the explorers' perspective.

C. Communicative Goal

The communicative goal of the utterances is another factor that characterizes the alignment between the interlocutors. Since the focus is on the communicative utterances in this section, I narrowed down the analysis to the request-response pairs and alerts. The affirmative responses by A were also excluded for the same reason mentioned above. The utterances, produced by both E and A, were split into two main categories based on the communicative intention, namely whether it is instructive and descriptive. The instructive utterances category is comprised of those utterances (requests or responses) that focus on the line to be explored, such as navigational help like "You should go up" and positioning questions like "Am I at the beginning?". On the other hand, the descriptive utterances mainly convey content about a particular domain, e.g. 'bird population'.

In total, 400 utterances from *Es* and 465 utterances from *A*s were evaluated (Figure 9-4). For this analysis, a finer-grained terminology, namely the embodied perspective, and focused on the subject pronoun preferences for the reasons stated above was preferred. I used "the potentially self-referential pronouns" and "1st and 2nd person pronouns" interchangeably, and same for "the self-referential pronoun" and "3rd person pronoun". The participants' utterances were examined on whether there is an association between the communicative goal and the choice of subject pronoun by conducting Pearson's Chi-square. The details about the analyses for each activity roles are presented separately in what follows:

Explorers' utterances: The results showed that the type of communicative goal had a significant effect on the choice of the person pronoun ($\chi 2$ (1) = 66.24, p<.001). An additional analysis using the standardized residuals[34] was conducted to break down the significant Chi-squared test, see Figure 9-4a. When *E* had an instructive communicative goal, the standardized residual was significant for the use of the "1st person pronoun" (46.9%, N=183, z=2.9), and for the use of the 3rd person pronoun (26.9%, N= 105, z=-3.0). Moreover, when *E* had a descriptive communicative goal, the standardized residuals were significant for those who chose the "2nd person pronoun" (4.4%, N=17, z=-2.7) and for those who chose the "3rd subject pronoun" (21.8%, N=85, z=5.0 indicating that the association between the type of communicative goal and the choice of person pronoun is driven by both descriptive and instructive communicative content.

Assistants' utterances: The assistants' utterances also echoed a similar pattern in overall ($\chi 2$ (1) = 30.03, p<.001). Affirmative responses by the assistant that involve repeating the verb or noun used by the explorers are not informative enough to identify the verbal assistants' perspective. Therefore, only those utterances that involved additional information were tested separately, see Figure 9-4b. When *A* had the instructive communicative goal, the standardized residuals were not significant at all for the use of the "2nd person pronoun" (60.9%, N=78, z=1.5) and barely significant for the use of the "3rd person pronoun" (16.4%, N= 21, z=-2.1). On the other hand, when *A* had the descriptive communicative goal, the standardized residuals were significant for those who chose the "2nd person pronoun" (5.5%, N=7, z=-2.8) and those who chose the "3rd person pronoun" (17.2%, N=22, z=3.9). That means, indicating that the association between the type of communicative goal and the choice of person pronoun is mainly driven by descriptive content.

The results indicated that for instructive expressions (such as "you are going up"), the embodied-perspective, which inherently requires *1st* or *2nd* person pronoun, was preferred. In order to deliver the descriptive content about the graph domain, the non-embodied-perspective, which inherently requires *3rd* person pronoun (such as "it increases"), was preferred.

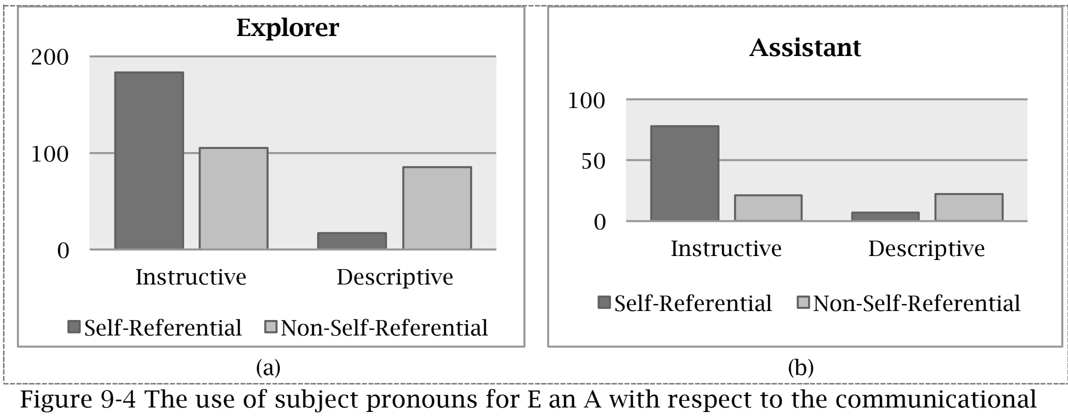


Figure 9-4 The use of subject pronouns for E an A with respect to the communicational goal

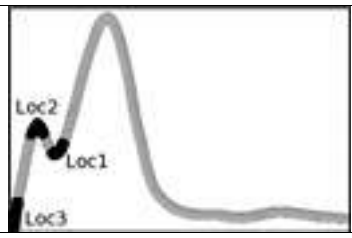
Table 9-2 presents an excerpt that illustrates a dialogue with instructive content. The verbal assistant provided navigational instructions by using the embodied-observer perspective during right-to-left exploration and he used the 2nd person pronoun instead of

[34] A z-score value bigger than 1.96 (ignoring the minus sign) is an indication of a significant standardized residual.

saying "*the graph line continues up and left*". It seems that *A* responded to *E*'s question by answering *E*'s intention of reaching to the left-end with a navigational instruction, since the abstract spatiotemporal statements would be irrelevant in that case.

Table 9-2 Instructive Assistance

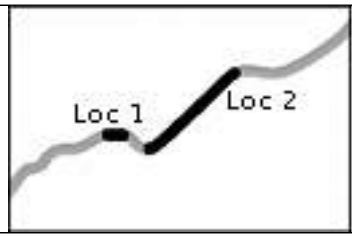
Location	**Utterance**
Loc1	E: *Is this left-end?*
Loc1	A: *No, go upwards to left*
Loc2	A: *go downwards there*
Loc3	A: *Now, you are at the left-end*



Another excerpt with descriptive content is shown in Table 9-3. Graph-domain descriptive information was given in alignment with the spatiotemporal perspective and by using the 3rd person pronoun.

Table 9-3 Descriptive Assistance

Location	**Utterance**
Loc1	E: *Is here flat?*
Loc1	A: *Yes*
Loc2	E: *then it increases?*
Loc2	A: *it increases but before it increases, there is a slight fall.*



9.3.1.3. Interim Discussion

In addition to having different sensory abilities, the participants had different activity roles (an explorer vs. an assistant) in this joint activity. The findings suggest various implications for the design of a verbal assistance system for haptic graph comprehension. Firstly, communicating about how to navigate on the line, while considering the current position of the explorer on the graph frame at the same time, should be taken into consideration by the verbal assistance system. The haptic explorers' first preference is to focus on the basic spatial properties of the graphs by employing an action-perspective. However, when they realize that it may not be sufficient to explain the changes in the graph domain or that it might conflict with the graph reading, they switch to a spatiotemporal perspective. In addition, the results showed that a given action is being simulated from the exploring agent's perspective (action-perspective).

The results are in line with Beveridge and Pickering's (2013) *Spatial Grounding Hypothesis* stating that action simulations are grounded in spatial context; however, the factors that affect the perspective-taking highlight one specific issue. In our experimental setting, the spatial context is always available. But the abstract spatiotemporal perspective is relatively hard to adopt since the exploration is performed with an action that inherently facilitates the embodied (*action-)* perspective for the haptic explorer. Moreover, the assistants are also visually exposed to *E*'s actions; this seems to facilitate adopting the *embodied-observer* perspective instead of the *non-embodied-observer* (*spatio-temporal*) one.

For the sake of communication, the verbal assistance system should align itself with the explorer's current perspectives; however, the explorer's perspective is not always the most efficient perspective for conceptualizing the graph, but the assistant's use of the *non-embodied-observer perspective* may help to activate the abstract spatiotemporal representations. Furthermore, the observed association between the choice of subject pronoun (hence the adopted perspective) and the communicative goal makes the choice of subject pronoun very useful tool for priming appropriate perspective w.r.t communicative goal.

These findings are highly promising for designing a natural language generation system that produces adequate referring expressions for the verbal assistance, for the alignment of the interlocutors' internal models and for building a common ground in the course of verbally assisted graph exploration.

9.3.2. The Role of Taking Initiative

In this empirical analysis[35], I focused on two factors, which are highly relevant for designing computational verbal assistance for haptic exploration of line graphs: (i) the role of the haptic explorer as the dialogue-initiator and as a consequence the assistants' competence to interact cooperatively, and (ii) the role of the type and richness of content that the verbal assistant provides.

9.3.2.1. Experimental Setup

Participants: In order to investigate the contribution of verbal assistance to haptic graph comprehension, I performed a comparative analysis of the results obtained through Experiment-I (*single-user paradigm*) and Experiment-IV (*dual-user paradigm*, see 4.3.1 and 4.3.4 for the details of the experiments), see Figure 9-5 for the illustration of the experimental conditions.

> **Condition 1:** The first group (nine university students, four females, Mean age = 25.0, SD = 6.3) examined haptic exploration of line graphs while being blindfolded and in the absence of verbal assistance (Experiment I-Condition 1).
>
> **Condition 2:** In the second group (Experiment-IV), participant pairs (a blindfolded haptic explorer and a verbal assistant who was able to observe the haptic exploration, 13 pairs, 11 females, M=25.3, SD=3.27) collaborated in exploring the same haptic line graphs. The haptic explorers were told that the goal of the study was to design efficient and effective verbal assistance systems and that the haptic explorer would initiate the help request. The task of the verbal assistant was to provide the necessary information in a short description whenever requested by the haptic explorer.

In brief, after a warm-up and instruction sessions, the participants in both conditions were presented with the stimuli. After the exploration of each graph, the participants produced single-sentence verbal descriptions of that graph to a hypothetical audience. After the verbal description, they produced a sketch of the graph using paper and pencil. In this analysis, the post-exploration sketches produced by the haptic explorers in both conditions were compared.

Group 1 (from Exp-I)	Group 2 (from Exp-IV)

[35] This analysis was published in Alaçam, Habel and Acartürk, 2014.

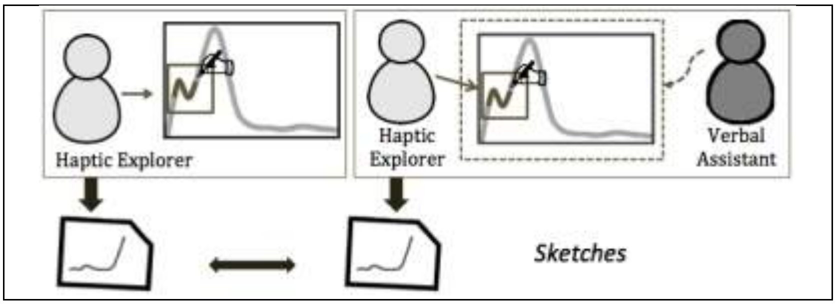


Figure 9-5 An illustration of the experimental conditions in the focus of this analysis and the resulting outputs

Providing information in response to an explorer's question or statement has the potential to enhance *E*'s comprehension of the graph at that particular moment of verbal assistance. In other words, *E* can use the information provided by *A*'s utterance and form a potentially more correct representation of the graph or of a graph segment by combining it with his/her exploration. In the course of the dialogue between the haptic explorer and the verbal assistant, each of the explorer's utterances can affect the explorer's mental representation of the graph.

In this analysis, I focus on the communicative goal of the dialogues. The assistant's contributions to the dialogue can be classified as follows: (1) instructive (i.e. navigational, such as "go downward from there"), or (2) descriptive. Descriptive utterances provide (2a) confirmative assistance (specifying exploration events or graph entities without using modifiers - such as"there is a decrease"),and (2b) additional assistance (specifying properties of exploration events or graph entities using modifiers, such as "there is a steep decrease"). Based on this scheme (see Figure 9-6), I classified the dialogues into two major groups. Firstly, I identified *weak content* dialogues, which were less informative. These were the dialogues that contained assistance focused on (or restricted to) '*basic spatial properties*' of the currently-explored region (i.e. the location or polarity of the graph segments). Secondly, I identified *rich content* dialogues, in which the verbal assistant also provided additional properties about the region that were being explored (e.g. information about the steepness or length of the graph segments). Two coders performed the classification task. The interrater reliability between both coders was calculated by Cohen's kappa. The results revealed a value of 80, which indicates a substantial interrater agreement.

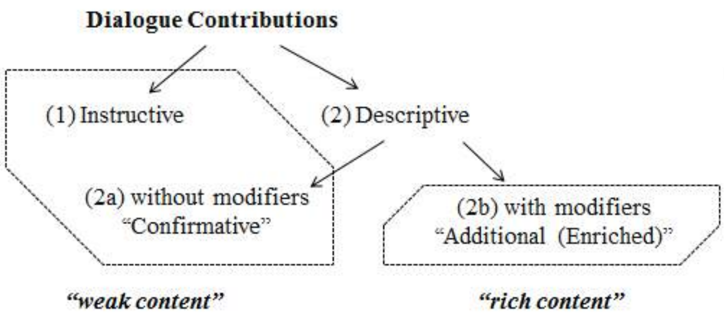


Figure 9-6 Classification scheme for the dialogue contributions

9.3.2.2.Results

The participants in the condition-1 followed a single-user experimental paradigm (9 haptic explorers x 5 graphs resulting 45 protocols in total). On the other hand, in the condition-2, a dual user experimental paradigm was employed. The results of the condition-2 showed that out of 65 experimental protocols (13 haptic explorers x 5 graphs) of graph stimuli, 28 protocols involved at least one request from the verbal assistant. In other words, the haptic explorers were the dialogue initiators in 28 of 65 experiment protocols of the graph stimuli. Since this corresponds to almost half of the pairs in Condition 2, a further analysis of the data was performed after dividing the protocols of Condition 2 into two groups: (i) Dialogue-initiated protocols (henceforth referred to as dialogue-initiator protocols) and (ii) The protocols that involved no dialogue initiative, thus no verbal assistance was given (henceforth referred to as no-initiator protocols).

In this comparative analysis, the focus was on the effect of receiving verbal assistance (after having initiated a request) and on the content of the dialogue. The latter was done by taking the similarity of the participants' sketches to the stimulus graphs as a performance measure. Two raters scored the sketches (all raters were blind to the goals of the study) with respect to their similarity to the stimulus graphs by using a 1 (least similar) to 5 (most similar) Likert Scale. For the analysis of the sketches, inter-rater reliability between the two raters was assessed using a two-way mixed, consistency average-measures ICC (Intra-class correlation). The resulting ICC=.69 can be classified as "good reliability" (Cicchetti, 1994; p. 286).

Analysis of Post-Exploration Sketches: I analyzed the sketches drawn by the participants in terms of their similarity to the stimulus graphs. A statistical analysis using the Kruskal-Wallis test revealed a significant difference between the ratings ($\chi2(2, N=108)=23.3$, $p<.01$) among single-user protocols, no-initiator protocols, and dialogue-initiator protocols. Post-hoc testing of contrast using Mann-Whitney with Bonferroni correction (so all effects are reported at a .0167 level of significance) showed that the sketches in the no-initiator protocols ($M=1.93$, $SD=0.90$) received lower similarity scores compared to both the sketches in the single-user protocols ($M=2.81$, $SD=1.16$), $U=410.0$, $p<.01$ and the dialogue-initiator protocols ($M=3.17$, $SD=1.14$) $U=170.5$, $p<.01$, without a significant difference between the latter two. Another Mann-Whitney test (with Bonferroni correction) was conducted for investigating the effect of the information content of the utterances. The utterances with rich content resulted in higher similarity scores for the sketches ($M=3.47$, $SD=.72$) in the dialogue-initiator protocols than the sketches in the single-user protocols ($M=2.81$, $SD=1.16$), $U=236.50$, $p<.05$ and the other conditions, see Figure 9-7 .

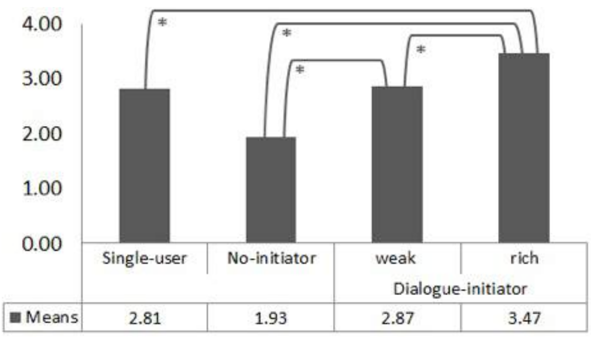


Figure 9-7 Ratings for sketches in five point Likert Scale (1: least similar and 5: most similar)

This indicates that the dialogues containing more specific information (such as slight increase, biggest curve etc.) resulted in a better sketch production, which is an indicator of more complete conceptualization of the event, see Figure 9-8Figure 9-8 for sketch samples.

(a) (b) (c) (d)

Figure 9-8 Sketches after the protocols (a) without verbal assistance, (b) with weak content verbal assistance (c) with rich content verbal assistance and (d) the original haptic graph-stimuli

9.3.2.3. Interim Discussion

In this analysis, I investigated the role of the haptic explorer as dialogue-initiator (or no-initiator) and the role of the verbal assistance content in a collaborative activity. Although the haptic explorer and the visual assistant share a common field of perception, their perception and comprehension processes differ significantly. In order to generate automatic verbal assistance with adequate content, it is necessary to identify the types and roles of the individual utterances as well as the structure of the dialogue content by using empirical studies. The experiment was conducted under two conditions to investigate the contribution of verbal assistance in haptic graph exploration. The sketches produced by the participants of the first condition (single–user protocols without verbal assistance) and those of the second condition in two types of protocols (no-initiator protocols without assistance and dialogue-initiator protocols with assistance) were compared with respect to their similarity to the original graphs. High similarity ratings of the sketches were correlated with the richness of the content provided by the verbal assistance. The sketches from the dialogue-initiator protocols were significantly more similar to the stimulus-graphs compared to the single-user and no-initiator protocols. The results also demonstrated a considerable effect of verbal assistance content, meaning that the dialogues that contained modifiers (cf. *rich content*) were helpful to the explorer. The presence of modifiers made the assistance more elaborate; it helped the participant to notice the features of the event, which were currently explored (e.g. steepness of the curve and length, relation with another curve).

9.3.3. Referring to Shape Entities: Semantic Representations

As has been shown in the previous empirical investigations, the referring expressions produced by the haptic explorers and verbal assistants during a collaborative activity give insights into how graph readers comprehend graphs, which elements are mentioned the most, and how they are referred to. This analysis[36] particularly focuses on the linguistic analysis of the utterances in a dialogue and it also deals with the semantic representation method as well as the production of haptic ostensive expressions during the joint activity.

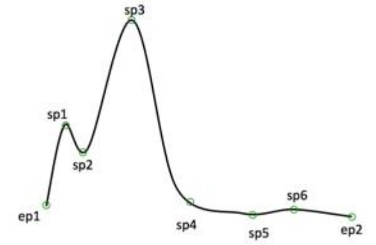
The investigation of multimodal interactions (namely the interaction by means of language, gesture and graph) requires both a systematic qualitative analysis and a quantitative analysis. I followed the widely accepted method developed by Dale and Reiter (1995), which addresses the generation of referring expressions. This method was applied as a tool for making systematic mapping between the semantic properties of the graphical features and participants' referring expressions, and for characterizing the semantic properties of the graphical segments and the referring expressions produced during the collaborative activity. According to Dale (1992), Gricean-like conversational maxims should be sustained by a system that generates referring expressions. These maxims address with adequacy, efficiency and sensitivity. The maxim of adequacy corresponds to containing enough information to allow the hearer to identify the referent. The maxim of efficiency entails that the expression should not contain any unnecessary information. And according to the maxim of sensitivity, the referring expression should be sensitive to the needs and the abilities of the hearer.

In order to generate referring expressions that sustain these maxims, Dale and Reiter propose and implement a cost function, which is based on empirical research. This function assumes that people first and usually prefer to refer to type properties (zero cost), then to refer to absolute properties. Relative properties and relations (the highest

[36] This analysis was published in Alaçam, Acartürk and Habel, 2014.

cost) follow right after that. By following this method, ⟨attribute, value⟩ pair representation to characterize the qualitative representations of graph shapes and landmarks was employed. To illustrate that, the attribute set, which is available for the "*ep1-sp1*" shape segment (see Table 9-4) possesses the following properties: ⟨*type, curved*⟩, ⟨*manner, steep*⟩, and ⟨*direction, up*⟩.

Table 9-4 Qualitative ascriptions for shape landmarks and segment: "*ep1-sp1*"



Shape landmarks		
	Landmark characteristics	**Global properties**
ep1	left end pt., local min.	higher than sp4, ep2
sp1	smooth pt., local max.	higher than ep1, sp2, sp4, sp5, ep2
Shape segments		
	Shape characteristics	**Vertical orientation**
ep1–sp1	curved	steeply upward

To achieve a systematic data analysis, the verbal data produced in a joint activity was also characterized by using this method since it successfully foregrounds the common properties of multimodal data; see Table 9-5 for the semantic attribute scheme employed in the annotation of the verbal data. For instance, an utterance like "is there an increase here?" produced by the haptic explorer can be modeled as <type, increase>. If the verbal assistant responds to it by providing additional information which is not modeled in E's attribute set, i.e. "yes, it is a moderate increase", then an update of attribute set occurs resulting in additional <attribute, value> pairs like <type, increase> and <manner, moderate>.

Table 9-5 Semantic attribute scheme

Type Properties:
- Terms
  - ⟨term, peak⟩, ⟨term, something⟩
- Location
  - Frame of Reference Terms ("start point")
  - Haptic Ostensive Expressions

Absolute Properties:
- ⟨value, 0⟩ for "it is 0"
- ⟨count, 3 peaks⟩

Relative Properties:
- ⟨size, small⟩ , ⟨manner, slowly⟩
- ⟨direction, up⟩

Relations:
- ⟨temporal relations, after the fall⟩
- ⟨spatial relations, higher⟩

Others:
- Interjections (hmm, ah…)
- Affirmations/Negations

In addition to the attributes stated by Dale and Reiter (1995), we identified haptic ostensive expressions (*HOE*s). The haptic explorers produced *HOE*s that referred to the pointed locations, which are also accompanied by assistance request from the verbal assistant. Foster and colleagues (2008) define the *HOE* as a reference, which involves a deictic reference to the referred object by haptically manipulating it. Since the haptic explorer's location is visible to the verbal assistant during joint activity, haptic actions are useful to create joint attention between *E* and *A*. The following excerpt illustrates the use of HOEs in a dialogue.

E: "Is <u>this</u> the left end?"

The question asked by the haptic explorer contains a demonstrative pronoun "*this*", and the explorer also performed simultaneously small hand movements with the stylus causing the pointer on the graph which the assistant can observe move in order to catch the assistant's attention.

9.3.3.1.Experimental Setup

In this investigation, again, the data collected from Experiment-IV was used. See 9.3.1 for details about the experimental setup.

9.3.3.2.Results

The participants produced 65 dialogues (5 stimuli x 13 pairs). The average length of a dialogue was 103 seconds (SD=62 sec.). Table 9-6 contains the mean completion times for each graph. Each utterance occurring in the dialogues was transcribed and time-coded. The transcriptions were then annotated according to the semantic attribute scheme presented in Table 9-5. The term "utterance" refers to speech parts produced coherently and individually by each participant. As mentioned before (9.3.1), the utterances were classified into three categories; (i) request-response Pairs, (ii) alerts initiated by *A*s (but not requiring any response from *E*s) and (iii) think-aloud sentences. Utterances that were initiated by the *A*s without any request coming from *E*s were mostly utterances that alerted *E* when s/he reached a start point or an end point. Although the *E*s were not instructed to use the think-aloud protocol, self-talking during haptic exploration was observed with 10 out of 13 haptic explorers. In this experimental design, the explorers were asked to turn the microphone on to open a communicational channel, during self-talking none of the participants turned the microphone on; therefore, these sentences were produced without a communicative goal shared with the partner. Because I focused on the communicative utterances in this analysis, I focused on request-response pairs and alerts excluding all think-aloud sentences.

Help Request Content: The results pointed out that the most frequently observed assistive content held information about the positioning, such as being on a start point or on an end point, on the frame, or being inside or outside of the line. 72.4% of the utterances (341 utterances in total, of which 46 were initiated by *A*s) addressed this type of information. *E*s showed a tendency to request assistance by directing "Yes/No Questions or Statements" at the *A*s (N=418) instead of using open-ended questions (N=7).

Assistance Content: *A*s' contributions to the dialogue can also be classified as follows: (1) *instructive*, N=69 (i.e. navigational, such as"go downward from there"), or (2) *descriptive utterances*, N=386 (see Figure 9-6 for the classification scheme). *Descriptive utterances* included, (2a) *confirmative assistance*, *N*= 342 (confirming the information the haptic explorer has already accessed), and (2b) *additional assistance*, *N*=44 (introducing a new property or updating the value of properties already stated). Below we present sample request-response pairs, which introduced new information or updated the value of the attributes already introduced.

Table 9-6 Mean completion time for each graph

Graph No	Mean Completion Time (Sec.)
A	76.4 (SD=41.3)
B	118.8 (SD=90.1)
C	93.7 (SD=56.3)
D	116.6 (SD=53.7)
E	110.5 (SD=62.2)
Overall	103.2

A. Attribute Update

The non-parametric correlation analysis using Kendall's tau discovered a positive correlation between the *existence* of attribute updates in the dialogues and higher sketching scores (N=62, τ=.46, p=<.01). Also, the *number* of attribute updates is positively correlated with higher sketching scores (N=62, τ=.45, p=<.01). The following is one of the dialogues between *E* and *A* and illustrates an attribute update: *E* asked a question ("*Is this going perpendicularly?*") to *A* by pointing at the "ep1-sp1" segment of the graph shown in Table 9-4. As stated in Table 9-7, this shape segment can be labeled with the ⟨type, curved⟩, ⟨manner, steep⟩, ⟨direction, up⟩ attributes. With his question, *E* addressed both manner and direction attributes. However, the word for "perpendicular" in Turkish can be used to refer to both being perpendicular and steep. Here *A*'s response ("*There is a slight slope*") updates *E*'s information and it also clarifies possible misunderstandings, since perpendicular lines are not allowed in statistical graphs in the time domain. The resulting request-response pair covers all attribute pairs for that particular graph shape (the region which *E* needs assistance for) and the sketch was rated with a 4.5 in average (on a 1-to-5 Likert Scale). The parameters (Dale and Reiter, 1995) (i) the number of attributes that are available for use in a referring expression and (ii) the number of attributes mentioned in entire dialogue seem to be a useful indicator to evaluate the success of communication.

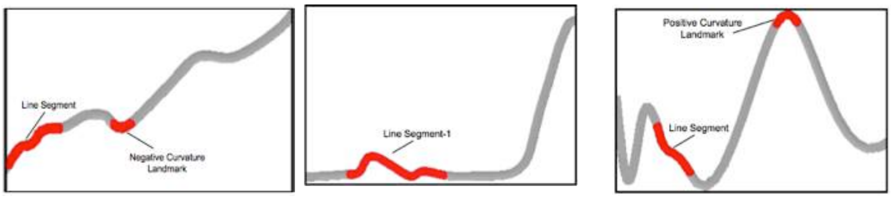
Table 9-7 Two examples that illustrates attribute updates

Utterances	Semantic Representations
Example-1	
E: "Is this the start point?"	⟨type, start point⟩
A: "Yes, it is also the origin."	⟨type, start point⟩ & ⟨type, origin⟩
A updates the ⟨type, start point⟩ attribute pair as ⟨type, origin⟩ that emphasizes 2D frame of reference, which implicitly carries over the value for the starting point	
Example-2	
E: *No request.*	
A: "You are at the first curve."	⟨type, curve⟩, ⟨relation, order, first⟩
Both type and relation attributes were introduced to the dialogue	

B. Content of the Request and Assistance

A further qualitative analysis was conducted focusing on the "type" attribute, which corresponds to highlighting a particular point or a region on the graph, showed that these expressions might also be split into two sub-categories: definite expressions (such as curve, maximum, bottom etc.) and indefinite expressions (such as "there is *something*"), see Figure 9-9. By way of illustration, the line segment in Figure 9-9a was referred to as "like an M letter", and the line segment in Figure 9-9b was described as showing a "stair" and "steps" Meanwhile, the line segment highlighted in Figure 9-9c was referred to as "something. Verbal assistants' referring expressions that contain a "type" attribute referring to a point or a region on the graph were all definite type attributes (such as "curve", modified by expressions such as "biggest", "smallest", "first" etc.). On the other hand, haptic explorers showed a tendency to use simpler expressions such as "peak", "hill", "elevation". Finally, all onomatopoeic words ("tick tick" or "hop hop") were introduced and used by the haptic explorers only. The use of such words created on the fly indicates that the haptic explorers had difficulties in naming the entities. It can be also considered as indicator of the problems in switching to spatio-temporal perspective, since those words seems to be triggered by the use of action perspective.

Mentioning these line segments and landmarks using indefinite type attributes in the haptic modality is nontrivial because they point out the difference between the two modalities. These regions were not mentioned using definite type attributes by the participants who explored the visual graphs in our previous study (see Chapter 5.3). In the visual modality, these regions were mainly described by means of descriptions of fluctuations or small variations.



(a) Graph-II (b) Graph-III (c) Graph-IV

Figure 9-9 Example graphs of curvature landmarks and line segments

C. Haptic Ostensive Actions and Expressions (HOEs)

The haptic ostensive actions and expressions performed to catch the attention of the assistant do not directly contribute to the conceptualization of the graph shape. Still, their communicative role in dialogues is rather important. 20.4% (*N*=247) of all the communicative utterances contained *HOE* enhancing reference resolution. As such, shorter descriptions could be produced instead of long ones. The analysis of verbal data revealed two major subcategories of *HOE*s: (i) demonstrative pronouns (*DP*s) such as “This/Here” or “like this”, and (ii) temporal pointings (*TP*s) such as “Now”. Table 9-8 presents the frequency of occurrence for each *HOE* category. Non-parametric Wilcoxon signed-rank tests were conducted to investigate the use of the different *HOE* types. The results showed that the haptic explorers produced more *DP*s (z=-4.88, p=<.001) and *TP*s (z=-3.75, p=<.001) than the assistants. While there was no significant difference in the number of *DP*s and *TP*s produced by *E*s (z=-.50, p=>.05), the *assistant*s preferred to use *TP*s rather than *DP*s. Only a few instances (*N*=5) of *DP*s uttered by an *E* were responded by the *A*’s use of *DP*s. Instances in which an *A* responded to the *E* by using a different *HOE* category than the one used by *E* ,were not observed at all.

Table 9-8 The number of HOEs for each category

	Only by *E*	**Only by *A***	**Both *E* & *A***
Demonstrative Pronoun-*DP*	99	6	5
Temporal Pointing-*TP*	67	27	19

A further analysis was performed on salient graph parts by focusing on the areas of the graph the participants preferred to use one of the two *HOE* categories (demonstratives and temporal pointing) in for referring. For this, the accompanying content (the location being referred to) were classified into three groups, namely (i) the references to start and end points, (ii) the references to intermediate points or regions on the graph and (iii) the references to the frame (such as being on the frame, or being outside of the line). The results of the analysis showed a significant association between the referred location and the *HOE* type preference, $X^2(2)$=38.2, p<.001. The results (the standard residuals for each combination) further indicated that when the participants referred to a start/end point of the graph line, they used *DP*s (*N*=48, z=-.6) and *TP*s (*N*=48, z=-.7). However, for referring to any particular point or any region on the graph, they preferred using *DP*s (*N*=59, z=2.8) rather than using *TP*s (*N*=16, z=-3.1). Moreover, when they mentioned about the events related to the reference frame, they preferred *TP*s (*N*=29, z=3.3) rather than DPs (*N*=6, z=-3, *all p values were smaller than .05*). But no main association was found between the *HOE* types (*DP*s or *TP*s) and whether the referred region is a point or an area. This indicated that specific points (i.e. landmarks) as well as broader regions (i.e. line segments) haptically highlighted by *E* were accompanied by any of the *HOE* types. The position of the point or region on the graph (i.e. at the beginning or at some intermediate region on the line), however, had an effect on which *HOE* type is preferred.

9.3.3.3. Interim Discussion

Following Dale and Reiter (1995), the graph shapes (segments/landmarks) and the verbal data were categorized as attribute pairs such as ⟨type, maximum⟩. When *E* needed assistance for a segment, or a global shape, her/his question was modeled as a specification of the choices of some of the attributes. As a response to the request for assistance, the description of *E* may be complete, lacking or partially or completely inaccurate. In order to have a successful communication, the verbal assistant should provide necessary information or correct the incorrect interpretation to complete the coverage of attributes in the "target set" of attributes. Hence, within this framework, I assume that successful communication is achieved when *E* requests assistance (initiated by him/her based on his/her needs to avoid receiving over-assistance) and *A* updates the existing attribute pairs or introduces new ones. Moreover, since *E* already has access to basic spatial properties, a useful solution would be to provide information with domain specific terms, relative terms (since absolute terms are difficult to implement) and relational terms which emphasize size and manner gradually (considering the haptic explorer's needs and current knowledge). Moreover, the verbal assistants introduced more graph domain oriented concepts to dialogues, while haptic explorers tended to use simpler daily terms or even onomatopoeic words. This information is important when forming attribute set for graph shapes.

The focus of this investigation was to investigate the content that needs additional assistance, but the results also pointed out the information that can be provided more effectively through a different modality than the verbal one. The research by Moll and Sallnäs (2009) and Huang, Moll, Sallnäs and Sundblad (2012) suggest audio-haptic guidance for visually impaired people to enhance the navigational guidance in virtual environments. Hereby, the participants can focus on communication at a higher level. Their results indicated that "by using haptic guiding one can communicate information about direction that does not need to be verbalized" (Moll and Sallnäs, 2009, p.9) and "sound provides information that otherwise has to be conveyed through verbal guidance and communication" (Huang et al., 2012, p.265). Considering that 72.4% of the utterances in our experiment contained information about positioning (being at the start point, or on the line etc.), providing this information to the explorer seems crucial for the system to be assistive. However, delivering this information verbally would yield continuously speaking assistance system; therefore sonification could be a good way of carrying this message. Additionally, haptic exploration allows haptic ostensive actions that highlight the visited location. The location attribute has different characteristics compared to other attribute pairs. It grounds joint attention between partners by pointing where the assistance is needed, after which other attributes provide additional information about what the graph shape means. As for *HOE*s, the type of referring expressions (demonstrative pronouns or temporal pointing) seems affected by the location referred to (start/end points, intermediate regions or the graph frame). The results also indicated that the explorers produce significantly more *HOE*s during joint activity than the verbal assistants. In the collaborative activity settings that allow both users (*the human explorer* and *the human or robot assistant*) to manipulate the environment haptically, the assistants' haptic ostensive actions exhibits salient communicative functional role (Foster et al., 2008; Moll and Sallnäs, 2009). However, in our assistance setting, only haptic explorers have an active role in the haptic exploration. Even after requesting assistance from *A* regarding a specific point or region indicated by means of an *HOE*, *E* may still continue to explore. Therefore verbal assistants tend to omit uttering an *HOE*. When necessary, they use temporal indicators to relate a previously mentioned expression to the currently explored region. This preference shown by the verbal assistants may be for preventing that the explorers perform reference resolution incorrectly.

Finally, in addition to the attribute-set approach by Dale and Reiter (1995), a more context sensitive version that implemented salience weights was proposed by Krahmer and Theune (2002). The comparative study between visual and haptic perception of graphs indicated that haptic readers tend to overestimate small variations on the graph shape due to the haptic salience induced by haptic friction. They tend to underestimate smooth regions that can be useful for segmentation as well (Habel Alaçam and Acartürk, 2013). Choosing appropriate attribute values enhanced with salience weights for these kind of haptically problematic regions might overcome this problem.

Taking the Gricean maxims into account when generating referring expressions, (careful selecting the information provided in the "attribute pairs", updating attributes gradually and being sure that at the end of the communication, the target attribute set is covered) seems useful for enhancing the conversational success of the communication (Grice, 1975; Dale, 1992; Dale and Reiter, 1995).

9.3.4. Interpreting Haptic Exploration Patterns

The results elaborated on in the previous investigations revealed valuable insights into how the comprehension is affected by the provided language content. So far, I have mainly focused on the information content of the verbal assistance that is to be delivered to the user of the haptic interface. In other words, this information has been about "*what to say*" and "*how to say*" to the user in order to improve his/her comprehension in the course of haptic exploration. The findings have shown that appropriate information content leads to a successful conceptualization of the events that are represented by the graph lines. The assistants do not only have to decide what to say but in particular when to say it. Therefore, an equally important aspect in designing haptic graphs is to identify "*when to say*". Combining this information with the haptic exploration patterns before/during/after the explorer's help request will provide a concrete base for the automatic detection of the verbal assistance need. The detection of what a graph reader wants to know at a particular time during exploration based on a analysis of his/her current position of exploration, previous exploration movements, and referring utterances (the referred locations and how these regions were referred) would yield a more effective design of the (learning) environment for the graph reader compared to presenting all possible information to him/her at once.

This investigation[37] focuses on the design of a feature set for the purpose of describing patterns of haptic exploration that is able to detect and predict whether the user needs verbal assistance by analyzing his/her haptic exploration patterns in the course of haptic exploration.

The design of this model is also based on the data collected in Experiment-IV (see 4.4.4 for details). The setting allowed recording the information content of verbal assistances as well as the explorers' haptic exploration patterns before, during and after a help-request. During *E*'s exploration, *A* analyzes the ongoing exploration event. In the long run, *A* has to take into account the history of exploration events and the history of the verbal assistance into account to be able to produce appropriate verbal assistance.

An automatic verbal assistance system, which is able to predict when to produce an utterance, should aim at modeling the alignment between the interlocutors and at making the communication between *E* and *A* efficient and effective. In the first prototype, called the OBSERVINGASSISTANT, we employed a rule-based approach. The system analyzes the users' exploration patterns and triggers canned text in a reactive manner, as realized by the MARY text-to-speech system (Kerzel, Alaçam, Habel and Acartürk, 2014).

In this investigation, I took a machine-learning and classification perspective by analyzing the current position of exploration, the history of exploration movements, and the qualitative ascriptions for shape segments and shape landmarks while taking their perceptual and conceptual saliency into account. The following section provides the technical background for the model by introducing the basic shape concepts in graph-line descriptions.

9.3.4.1. Data Collection

A total of 13 pairs of sighted and blindfolded university students participated in the experiment. Each pair was composed of a haptic explorer (*E*) and a verbal assistant (*A*).

[37] This analysis has been published partially in Alaçam, Acartürk and Habel (2015).

Not all haptic exploration protocols had dialogue or request for verbal assistance therefore 26 protocols from 7 pairs were chosen from the empirical data for the feature set construction mentioned in the next section.

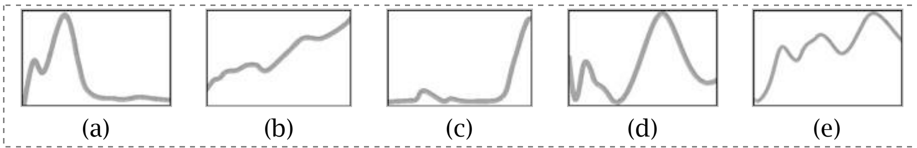


Figure 9-10 Five different haptic graphs

9.3.4.2.Feature Construction and Transformation

The raw data set consisted of the recorded locations of the haptic device stylus on the 2D horizontal plane (horizontal and vertical coordinates). In this section, the feature construction and the transformation from 2D raw data into a 14D feature set, the cleansing and chunking procedures and the classification of the feature sets are presented.

As a first step in the model design, the raw data points (i.e. the coordinates of the stylus) were segmented according to the landmarks of the graph lines they lay on (seeTable 9-4). The segments were then labeled with their corresponding landmarks and line segments (Feature #3: Temporal Order in Table 9-9). Figure 9-11 exemplifies a set of basic geometric shape concepts for describing graph lines. For example the data point *(15,7)* was labeled *"SP3-SP4"* and the data point *(23, 2)* was labeled *"SP4"* based on the landmarks that the data points belonged to.

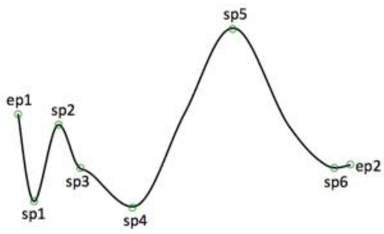


Figure 9-11 Qualitative shape landmark ascription for a sample graph

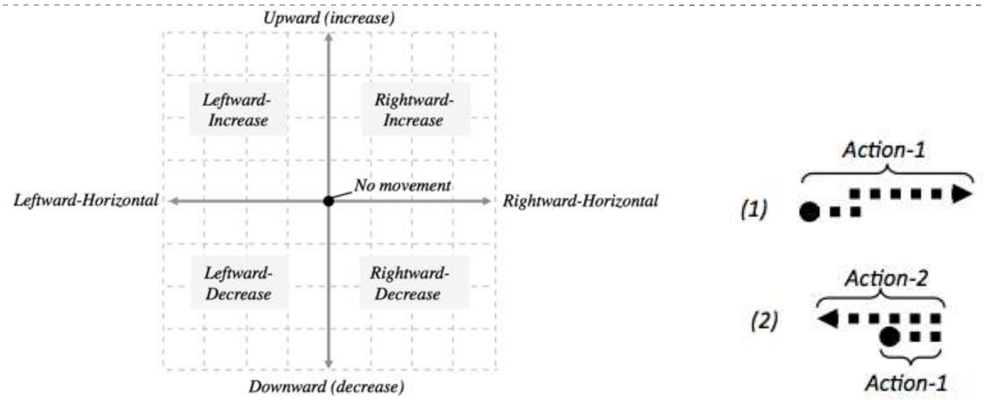
After the labeling step, the features were identified within the range from *basic (low-level)* features to *complex (high-level)* features. The first feature set (*SET I*) in Table 9-9 contains a set of basic features that can be derived from haptic explorations. The second set (*SET II*) involves higher-level features, such as the change of direction while exploring the same unit and the visit history (time sensitivity). Two low-level graphical features, namely the Graph ID and the size of the graphical unit, were added in the third set (*SET III*). The last set of features (*Set IV*) contains conceptual features.

Low-Level User Action Features: After the labeling of each data point, the Euclidean distance (Feature#4 in Table 9-9) between each consecutive data point was calculated. Time and speed were also fundamental features included in this data set. Afterwards, the direction of the action between each consecutive data point was labeled as shown in Figure 9-10a. The direction feature was not put directly into the final feature sets; instead it was used to calculate the number of direction changes in the same unit during the same action, as explained below. A data cleansing procedure was applied subsequently. As Grunwald, Muniyandi, Kim, Kim, Krause, Müller and Srinivasan (2014) stated, "Human haptic perception is accompanied by movement pauses of the exploring fingers and hands in healthy humans". The raw data contained instances of "*no-movement*" between the continuous actions as well, such as a *Left-Left-NoMovement-Left-Left-Left* pattern. Therefore the data points corresponding to small pauses, which were preceded and followed by an action in the same direction, were labeled with their temporal neighbors' direction.

Table 9-9 The list of features

SET I	*Basic Features*	*Type*
	1. Participant ID	Nominal-unique ID
	2. Temporal Order	Numerical
	3. Unit Label	Nominal
	Low-Level User Action Features	
	4. Total Distance	Numerical
	5. Speed	Numerical
	6. Duration	Numerical
SET II	*High-Level User Action Features*	
	7. Direction Change Count	Numerical
	8. Count of Visits in Last 10 Segments (or Last 6, Last 20)	Numerical
SET III	*Low-Level Graphical Features (Perceptual)*	
	9. Graph ID	*Nominal- Unique ID*
	10. Unit Size	Numerical
SET IV	*High-Level Graphical Features (Conceptual)*	
	11. End Point or Intermediate	*Nominal*
	12. Unit Type	*Nominal*
	13. Qualitative Ascriptions	*Nominal*
	Target Feature	
	14. The Availability of Verbal Assistance	*Nominal (VA+ / VA-)*

High-Level User Action Features: The direction change (Feature#7) between two consecutive data points with different directions was also labeled with respect to the two axes. In other words, the action with a Rightward-Increase, when followed by leftward actions or when followed by downward actions, were assumed to exhibit a direction change (see Figure 9-12a). One should note that both *NoMovement* instances and actions with a change in the main direction need special attention. Because the width of the haptic graph line is two pixels, the data needs to be corrected in order to eliminate the miscalculation of direction changes. Consider the following example to clarify the correction procedure: Let's assume that the dots in Figure 9-12 correspond to pixels. When we look at just two consecutive actions, we can label the action presented in Figure 9-12b as a *…Right-Up-Right-Right-Right…* action. However, the upward action in that case does not have a meaning in terms of the user's intended action, which is "to move right". Therefore the actions with Left, Right, Up or Down as main directions were checked with respect to their previous and next directions. If the previous action coincided with next one, then the value of this data point was updated. On the other hand, if they were different (Figure 9-12b), then they were marked as being part of a direction change (Feature#7).



(a) (b)

Figure 9-12 (a) Labels for direction mapping and (b) simple user actions on a haptic line with a width of 2 pixels

It is common practice to handle data by *chunking* it into units (segments or landmarks). Therefore, the next step in the analysis was to merge the data points that belonged to the same graph unit by taking their temporal order into account. Figure 9-13 is an example of the step-by-step chunking procedure. As a result of this procedure, consecutive data points that are on the same region were merged, and their features (i.e. distance, time, direction count) were also merged.

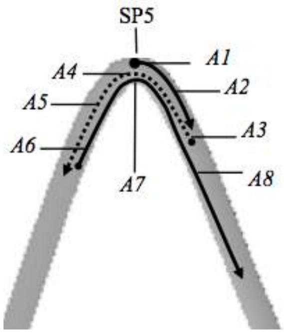
Time Order	Actions	Features
1	EP1-SP1	*Rightward-decrease (d1,t1)*
2	EP1-SP1	*Rightward-decrease (d2,t2)*
3	EP1-SP1	*Leftward-increase (d3,t3)*
4	EP1-SP1	*Leftward-increase (d4,t4)*
5	EP1-SP1	*Leftward-increase (d5,t5)*

EP1-SP1	*Rightward-decrease (d1+d2, t1+t2)*
EP1-SP1	*Leftward-increase (d3+d4+d5, t3+t4+t5)*

EP1-SP1	*Direction_Change = 1* *Distance = d1+d2+d3+d4+d5,* *Duration = t1+t2+t3+t4+t5,*

Figure 9-13 A sample result of the chunking procedure (d=distance, t=duration for each action)

Statistical line graphs are spatiotemporal representations when they represent time-series data. The features that are presented so far address spatial properties of the graphs. To keep track of the frequency of the visit in a time-sensitive way, a history of visit count was addressed by Feature#8 in Table 9-9. Since haptic exploration is sequential, the haptic explorer has to pass all the units on the way to that particular unit, in order to reach to an attended/intended location at a certain time. Therefore instead of counting the number of visits during the whole exploration, looking for the back-and-forth movement among the units in a pre-defined temporal window seems more efficient. In the present study, two versions of the visit history were calculated: The visit count of a particular data point (i) in the last 6 steps or (ii) in the last 10 steps. Figure 9-14 shows how the calculation is done.



Chunks	Actions(A)	Last Visit Count
1	1. @SP5	Visit#=1
2	2. @SP5-SP6	Visit#=1
	3. @SP5-SP6	Visit#=1
3	4. @SP5	Visit#=2
4	5. @SP4-SP5	Visit#=1
	6. @SP4-SP5	Visit#=1
5	7. @SP5	Visit#=3
6	8. @SP5-SP6	Visit#=2

Figure 9-14 A simple example of the visit count calculation

Low-Level Graphical Features (Perceptual): In the analysis of the features presented above, only haptic user actions were taken into account. On the other hand, empirical investigations presented in Section-II indicated that perceptual characteristics of the graphs play a crucial role in haptic exploration. Those characteristics were addressed in two features: The first feature (Feature#9) identifies the graph with a unique graph ID. The actual size of the graphical elements (the length of the units in pixels) is represented by the second feature (Feature#10). The second feature highlights the perceptual saliency of the units (i.e. a long segment vs. a short segment).

High-Level Graphical Features (Conceptual): In the last feature set (SET IV), I used a set of conceptual features that were derived from the raw coordinate data. First, the units were split into two categories: intermediate units and end points (Feature#11). Empirical results indicated that the explorers might request assistance for descriptive instructions (i.e. the graph domain content) or navigational instructions concerning the current positioning (being at the start point, or being on the line, etc.). The navigational instructions were mostly requested for end points of the graph lines. Therefore the data points were also classified with respect to their location on the graph. After then, each data point was categorized into two categories with respect to whether they belonged to *a segment* or a *landmark* (Feature#12), e.g. "*SP1*": landmark and "*SP1-SP2*": segment. The data points were then categorized based on their qualitative ascriptions, as described in

Section 1.2 and Table 2 (i.e. "*Ep1-SP1*": steep and "*SP5*": global max). Finally, the last feature represents a target feature, which is nominal. It addresses the binary classification problem itself.

As a result of the feature construction and chunking procedures described above, 69,574 instances (data points with horizontal and vertical coordinates) were reduced to 8015 instances (units), comprised of the units without verbal assistance (6953 VA-) and the units with verbal assistance (1062 VA+).

9.3.4.3.Results

The goal of the present study has been to develop a feature set useful for detecting the assistance need during haptic exploration (not to make a comparative analysis of classification algorithms). The features were tested by using the J48 decision tree algorithm. This algorithm was selected, because it is fast, simple and it can be easily converted to a set of production rules. Also it does not have any priori assumptions about the nature of the data. The Weka (Waikato Environment for Knowledge Analysis) environment was used to apply the classification algorithm to the dataset.

I performed a 10-fold cross-validation. Table 3 presents the results for each combination of the feature sets. In the first trial, I tested the feature set with low-level user actions (*Set I*). The overall accuracy (*ACC*, the rate of the correctly classified items) was %91.79. Then, I added the features of *SET II*. To reveal the contribution of the history window (Feature#8 in Table 9-9), I used three versions of the feature: the count of visits (1) in the last 6 chunks, (2) in the last 10 chunks and (3) in the last 20 chunks. The results showed that adding the high-level user actions slightly improved the performance. According to overall accuracy results, all feature sets have approx. same value, but when we look at the precision and recall values, first version (with the last 6-chunk window) seems to have the worst result. In the next run, the low-level graphical (perceptual) features were added to *SET II*. It improved the overall accuracy of the model by approximately 2%. In the last test, high-level graphical (conceptual) features were added to the model (SET IV), and this also slightly improved the overall accuracy. In SET III and IV, the history window (Feature#8) was calculated for last 10 chunks based on SET I+II (v2).

Table 9-10 The results of the J48 algorithm for the feature sets (VA: Verbal Assistance, TP: True Positive, FP: False Positive)

	Overall Accuracy (ACC, %)	**Kappa**	**VA**	**TP Rate**	**FP Rate**	**Precision**	**Recall**
SET I (7)	91.79	.59	VA-	.97	.45	.93	.97
			VA+	.54	.02	.77	.54
SET I+II (v1) (9)	91.86	.59	VA-	.98	.46	.93	.98
			VA+	.54	.02	.78	.54
SET I+II (v2) (9)	92.50	.63	VA-	.98	.41	.94	.98
			VA+	.59	.02	.79	.59
SET I+II (v3) (9)	92.56	.62	VA-	.98	.43	.93	.98
			VA+	.57	.02	.81	.57
SET I+II+III (11)	94.46	.73	VA-	.98	.32	.95	.98
			VA+	.68	.02	.87	.68
SET I+II+III+IV (13)	94.59	.75	VA-	.98	.29	.96	.98
			VA+	.70	.02	.86	.71

Although I obtained high accuracy results with *SET I* alone, it should be noted that it is more important for the verbal assistance system to provide verbal assistance when it is needed. Consequently, the correct classification of VA+ (*specificity*) cases is of higher importance. As expected, the results are lower for VA+ cases than for VA- cases due to the unbalanced distribution of the data set. Figure 9-15 illustrates the precision and recall values (presented in Table 9-10) for the VA+ class. The results indicated that J48 performed better with the inclusion of the graphical features, resulting in rise in the precision rates (from 77% to 86%) and in the recall rates (from 54% to 71%) for the VA+ cases.

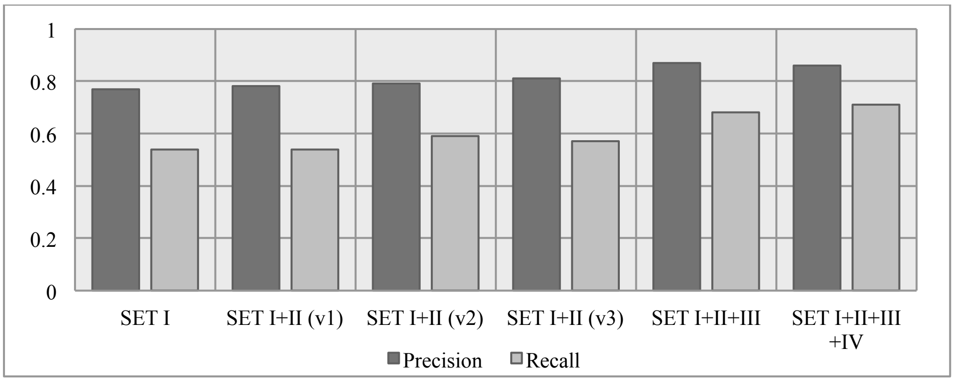


Figure 9-15 Precision and Recall values for VA+ Class

9.3.4.4.Interim Discussion

In this investigation, I introduced a classifier, which predicts whether verbal assistance is needed or not, based on the haptic exploration patterns. The motivation for this investigation was that in addition to deciding *"what to say"* and *"how to say"* it, providing this content at the appropriate time *"when to say"* is also crucial for the implementation of an automated verbal assistance system. The feature sets were derived from the raw data (the coordinates of the stylus on the haptic device). This data was obtained from a previous empirical study (Experiment-IV) on the production of referring expressions in a joint activity between a haptic explorer and a verbal assistant.

I employed a (supervised) classification algorithm, namely the J48 decision tree. All feature set combinations (starting from the low level user actions) achieved a high accuracy level (>91.5%). The best results in terms of sensitivity (*true positive rate*) and specificity (*true negative rate)* were achieved by adding the low-level graphical features (to the feature sets). Meanwhile, adding the high-level graphical features only slightly improved the classification. The results are in line with the experimental study, which showed that haptic explorers might have difficulty grasping these conceptual features due to the low bandwidth and inherent sequentiality of haptic exploration movements.

In the experimental study, the haptic explorers requested assistance when needed and, as a response to the request, verbal assistance was provided by the assistant. Both the content and the timing of the assistance were up to the haptic explorer. This resulted in an unbalanced distribution of the data points in our data set: The instances without verbal assistance outnumbered the instances with verbal assistance. Therefore in our results, I focus more on the specificity over sensitivity.

Finally, the empirical data showed that 72.4% of the verbal assistance was requested for the start and end points. As a consequence, the majority of the data points, where verbal assistance was requested, belonged to a limited region. In future work, I plan to provide sonification to alert the haptic explorer about the start point and the end point of the graph line so that s/he may focus on communicating about the intermediate regions and the changes in the trend without having to worry about navigational complexities. That would provide a suitable testing environment to assess the applicability of the feature set proposed here on the new data-set and, more specifically, to evaluate the contribution of the conceptual features in deciding the verbal assistance need.

9.3.5. General Discussion

This chapter focused on the dynamics of the collaborative activity and on the several key concepts that play a crucial role in the design of verbal assistance systems, concerning "*what to say*", "*how to say*" and "*when to say*". In an experimental setting, which employed a joint-activity framework, pairs of participants (haptic explorers and verbal assistants) explored the graphs and they exchanged verbal information when necessary. The results of the experiment showed that the assistants' role in the explorers' comprehension is critical and the haptic graph readers benefit from the verbal assistance to achieve more successful conceptualization of the events that are represented by graph lines. First, the assistants have more complete mental representations of the graphs starting from the onset of haptic exploration, since they have spontaneous access both global and local information on the graph. Their guidance through the salient points using additional attributes or their alignment of the instructions with the haptic explorers' current understanding of the graph enhances their comprehension. The results also revealed valuable insights into how the comprehension is affected by the provided language content. The information content and the information need is another crucial topic addressed in this investigation. To sum up, taking initiative in requesting help and having adequate verbal assistance enriched by modifiers, rather than just simple confirmation of the basic spatial properties in a response, seems a superb combination for a successful joint activity that inherently requires asymmetric dialogues between two users with different roles. Furthermore, the expressions that the verbal assistants used to refer to a point or a region on the graph, corresponding to "type" property, were mostly graph-domain terms (such as "curve", "peak" etc.). On the other hand, haptic explorers showed a tendency to use simpler expressions such as "something", "hill", "elevation". This indicated that haptic explorers have difficulty accessing graph-domain vocabulary to name the regions or the shape, so that they choose alternative ways to name it (including use of onomatopoeic words such as "hop hop"). Following Dale and Reiter's approach, I represented graph shapes and verbal data as attribute pairs. The empirical results revealed that a more successful communication was observed when the attributes used by the haptic explorers were enriched using the specific, graph-domain terminology. Accordingly, building up a multimodal system based on this approach looks promising. Finally, finding effective haptic exploration features and combining them with the (perceptual and conceptual) properties of the representation, which is being explored appears to be a promising solution in the detection of assistance need.

CHAPTER 10

Evidence-Based Design Heuristics

10.1 Introduction

The conceptual design of the proposed system is based on the interaction between two agents; a human explorer who perceives the graph haptically through active exploration and a verbal assistance agent that provides helps to the explorer with the requested help automatically. In order to design such human-computer interaction system, empirical research that focus on the task (*haptic graph comprehension*), and the affordances of the environments (*dynamics of a joint activity*) were investigated through single-user experimental paradigms as well as through human-human joint activity settings focusing on various research questions.

The results of the empirical studies presented in this dissertation contributed in preparation of the design heuristics. This chapter focuses on the issues that need to be taken into consideration in the design phase of an efficient and effective human computer interface. In this phase, different approaches are applicable for this purpose. The most standard and frequently used methods are *empirical user testing*, *cognitive (user) modeling tools* (e.g. Kieras, 1997) and *design heuristics* (*design guidelines*). As discussed in Chapter 2.4, in the method of empirical user testing, also called as evidence-based user research, users are observed interacting with a system, and the performance measures are collected through various methodologies (i.e. verbal descriptions, think aloud procedures, eye movement analyses etc.) and analyzed thoroughly. However, this method is highly costly and time-consuming. Therefore, low-cost methods such as (cognitive) user modeling tools and heuristic evaluation (e.g. Nielsen, 1994) are utilized as alternatives to this method. The cognitive modeling tools such as *the GOMS-family* of HCI tools (Goals, Operators, Methods and Selection, see Card, Moran, and Newell, 1983) and *CogTool* (developed at Carnegie-Mellon University, see John, Prevas, Salvucci and Koedinger, 2004) provide a descriptive model of how a user performs a task on a system, in other words it predicts user exploration preferences by taking the interface' affordances and the task at hand into account. The development of such models using these tools also requires domain knowledge and can benefit from the use of evidence-based user studies. However, as discussed previously, the underlying mechanisms of haptic graph comprehension has been scarcely investigated from theoretical point of view. Therefore, the empirical results presented in this dissertation might provide valuable input towards a design of such a cognitive architecture, however still further work needs to be done to reach GOMS style modeling.

Another frequently used HCI evaluation method is the *design heuristics*. They are usually constructed by the experts based on the expertise and the proficiency in the field or based on the evidence-based user testing, which is mentioned above. The design guidelines contain accumulated knowledge in the domain and serve to the purpose of making the reuse of good design solutions easier. Together with their cost and time effectiveness, they are one of the valuable tools in the HCI design. However, considering the rapid improvements on the emergence of new technologies, finding a guideline that target specific topic is a challenging issue. In general the haptic exploration interfaces and in specific haptic graph comprehension are two fields, which are lacking of this kind of research. In this chapter, I aim to provide design guidelines regarding verbally assistant system that allows user to explore the graph haptically and provides verbal assistance by taking user's actions into account as well as the graph's perceptual and conceptual properties. It should be noted that the proposed design involves perception, comprehension and communication of multi-modal information; therefore it contains different and intertwined layers. The guidelines coming from studies which do not incorporate *active haptic exploration*, or *verbal assistance* tailored to users' needs, are not directly applicable to such multi-modal interactive systems. To my knowledge, this study is the first one that addresses "haptic exploration-verbal assistance" constellations, to be more precise, that addresses a HCI system for verbally assisted haptic graph comprehension.

The topics investigated in this dissertation are in the focus of several mainstream interdisciplinary research fields; human computer interaction, cognitive science and computational (psycho-) linguistic. Besides its interdisciplinary nature in theoretical aspect, applied approach for designing such an assistive and instructive multi-modal system requires the involvement of many dimensions from the perspective of HCI. As can be seen in the schema illustrated in Figure 10-1 (that is identical to Figure 2-6), the proposed assistance system combines different dimensions of human-computer interaction systems. Therefore, the design guidelines, which target specific dimensions, need to be taken into consideration and be converged for this purpose. Each item was discussed in Chapter 2.4 by giving the reasons of why they are considered crucial to be regarded in the design of this particular system (the reader is encouraged to revisit Chapter 2.3 for a reminder).

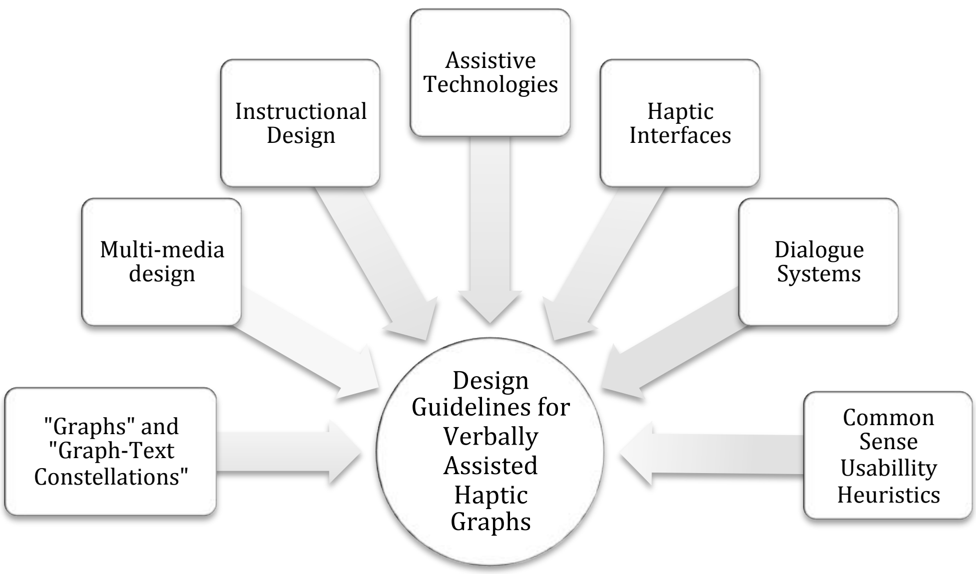


Figure 10-1 Different aspects of the proposed assistive system

As introduced and elaborated in Chapter 2.4, the contribution of the existing design guidelines in this study was two-fold. First the collaborative experimental setting was designed by following the relevant advices mentioned in those guidelines, and that part was already presented in corresponding chapter (2.4). In brief, the system

- utilizes sensory substitution to close functional and informational gap between visual and haptic modalities
- allows graph readers to explore graph actively and haptically
- assigns asymmetric activity roles to an explorer and an assistant
- aims to provide relevant information and to avoid redundancy

Although the technical aspects concerning haptic graph design are outside the scope of this dissertation, some relevant issues were also touched upon from this perspective.

The empirical studies were conducted by employing this task-oriented joint activity design. Afterwards, the results of the empirical studies were evaluated and merged with those abovementioned guidelines. This integrated knowledge was used as an input to construct design guidelines for a verbally assisted haptic graph comprehension. This chapter concentrates on the principles and guidelines that are constituted based on the empirical findings.

10.2 Design Heuristics

For successful graph comprehension, the conceptual event and the aspects of the event should be successfully carried by the physical properties of the graphs. This also holds for verbally assisted haptic graph comprehension. For this purpose, three aspects of communication for developing a successful interactive system were addressed; these topics are "*what to say*", "*how to say*" and "*when to say*"[38]. Concisely, the fundamental principle of the verbally assisted haptic graph comprehension can be described as follows;

The fundamental principle of the system: "provide relevant information in a relevant way in a relevant time".

In contrast to providing all likely information to the graph reader all together, the detection of what s/he wants to know at a particular time would yield a more effective design of the (learning) environment for the graph reader. This information could be extracted from the explorer's current position, previous haptic exploration movements and utterances (the referred locations and how these regions were referred).

All the principles and guidelines presented in following can be considered as a sub-category of this main principle. This general principle also entails the matching of Gricean Maxims addressing adequacy, efficiency and sensitivity criteria.

Furthermore, it should be also noted that the findings (the expressivity ratings of the verbal descriptions and the similarity scores of the sketches in Chapter 6.4.6) suggested that the visual graph readers have more complete mental representation of the event depicted in the graph. Therefore the taxonomic and partonomic patterns in event segmentation as well as the referring expressions produced by the visual graph readers during post-exploration descriptions or during the collaborative activities were employed as a guide in construction of the design guidelines.

As well as providing relevancy, another issue that needs to be touched upon is to provide natural communication environment. The human likeness of the system plays an important role in the user's performance and user's willingness to use the system (Davis, 2006; Venkatesh, 2000; Dillon and Morris, 1996). Concerning the human-likeness of the communication, several issues that need to be touched upon can be listed shortly as follows;

[38] From the HCI perspective these items are formulated as "saying the right thing at the right time in the right way" (Fischer, 2001).

- reduced information as the dialogue proceeds
- adjusting the granularity of the verbal content w.r.t user's prior knowledge and also w.r.t the prior utterances in the dialog
- appropriate timing (without exhibiting delay)
- using familiar terminology i.e. shapes (such as looks like a staircase)
- aligning the perspectives

These items are also taken into consideration in the preparation of the guidelines. In the following, the guidelines derived from the experimental findings with an relevant support from existing design guidelines were categorized into several sub-domains[39].

- Guidelines on the verbal content
  - Users and Individualization
  - Providing Prior (context-dependent) Knowledge
  - Event Schemata and Mental Models
  - Highlighting Context
  - Plain and Simple Verbal Content
    - General Remarks
    - Amodal Geometric Properties
    - Conceptual Properties
    - Metrical Information and Data Labels
  - Temporal Congruity
    - Sonification
    - Haptic Ostensive Expressions
    - Taxonomic Granularity
    - Spatio-Temporal Congruity
    - Partonomic Granularity
- Guidelines on *the dialogue dynamic*s
  - Perspective Taking
    - Alignment
    - Communicative Goal
  - Initiative Taking
- General Guidelines
  - The Division of Labor among the modalities
  - Limiting Complexity

10.2.1 Users and Individualization

Target group analysis is one of the fundamental steps that need to be taken care of, especially for a HCI design that targets visually impaired user with a purpose of being assistive. As such, understanding the abilities of the users in the target group and also the needs of the user concerning the tasks and goals is the fundamental part of usable HCI design that addresses *effectiveness*, *efficiency* and *satisfaction* criteria. The user aspect is one of the issues, which were touched at the ISO 9241-10 "Dialogue Principles". In a design phase, attention span, short-term memory constraints, learning strategies and degree of experience (i.e. having prior knowledge, or frequency of use of the system) should be taken into account.

However, *a typical user* for such a system does not exist as highlighted in the experiment performed with the visually impaired participants (see Chapter 5.4). Depending on the degree of impairment (i.e. whether the user is congenital blind or not), different approaches in information presentation (i.e. highlighting information in different

[39] It should be noted that due to intertwined nature of the content, some issues were elaborated more than once.

granularities or providing context-dependent knowledge prior to graph exploration) might be required. For example, for a congenitally and severely blind people, the depth is prominent feature in exploring the 3D objects. Therefore due to these predominant perceptual habits, these kinds of haptic properties might be highly salient for visually impaired users, as a result of this dominance, ignoring such property may not be easy as sighted participants.

Thus, *individualization* aspect that concerns users' ability, the degree of impairment and also the proficiency in the use of haptic device are important design issues, for an assistive system. ISO 9241 Guideline addresses these issues under different sub-titles such as "the intentional individualization" (Guideline 3.2 in ISO 9241-920) and "the suitability for individualization" (3.6 in ISO 9241-10). Furthermore, the "Personalization principle" proposed by Mayer (2009) states "people learn better from multimedia lessons when words are in conversational style rather than formal style". This topic also caught attention from the perspective of the "Universal Design for Instruction (UDI)" under the principles of "flexibility in use" and "equitable use" (Burgstahler, 2009).

Based on these principles and guidelines, the following recommendations can be suggested for the proposed system.

- Take the user's prior knowledge about graph domain into account. For the participant with less domain knowledge, make the graph-domain terminology and particularly the necessity of spatiotemporal perspective use explicit.
- Provide appropriate referring expressions regarding the shape of the graph by taking user's degree of impairment into account. For instance, when providing assistance for participants with congenital and high degree impairment, do not give metaphors or do not refer to objects and events that can be only visually accessible/understandable.
- Haptic actions (i.e. speed) are also subjected to individual differences. Therefore being able to adjust with respect to each user's haptic patterns would help to increase successful detection of the user's needs.

10.2.2 Providing Prior (context-dependent) Knowledge

According to the "pre-training principle" proposed by Mayer (2009), "people learn better from a multimedia lesson when they know the names and characteristics of the main concepts. ISO 9241-920 guideline refers this issue as "identifying information values (i.e. providing symbolic legend, or set of reference values)". Furthermore, this issue was also addressed in another item in ISO 9241-920 (Guideline 3.1.3) and also in Sjöström (2002) under the title of providing contextual information. Shah and Hoeffner (2002) touched upon this topic concerning the use of graphs as well.

Based on those principles, which were widely addressed, plain and simple introductory information about the abstract event depicted in the graph such as what the axes represent, what the labels and unit intervals are should be provided. For example, providing maximum value in the y-axis on the reference frame would be beneficial since people can use this information to make reasoning over proportions that make the remembering easier. Providing prior context-dependent information would also help to prime the use of spatio-temporal perspective.

Guideline. Provide introductory information such as what the graph represents, what the data labels carry.

10.2.3 Event Schemata and Mental Models

Mental models are considered as a key concept in the development of instructional technologies, tutorials, and other forms of user assistance. In classical HCI interface design that takes user's mental model into consideration, the mental models are more attributed to user's understanding about how the system works (Gentner and Stevens, 1983). On the contrary, the assistive system design requires more elaborated approach, namely, the system should be sensitive to align itself into user's perspective by adaptive mechanism and then help the user forming the correct mental representation of the task at hand. Therefore, the system should be able to adjust itself to the users' need, and provide help to the user in forming more complete mental representations. The empirical findings (5.4) also indicated that the haptic graph readers may not have sufficient knowledge in their graph schemata to establish appropriate mapping between the graphical features and the conceptual event depicted in the graph. This issue was also one of the addressed topics in Shah and Hoeffner's instructions (2002).

Guideline. help users to form correct mapping between the relevant haptic graph features (shape, size, orientation etc.) and the conceptual aspects of the event. In other words the system should activate the correct graph (event) schemata or update the existing one appropriately.

10.2.4 Highlighting Context

Another important issue was addressed in the ISO 9241-920 under the title of "presenting realistic experiences". As densely discussed, there is no need to provide all information to the participants; the verbal assistance system should provide most relevant information regarding the event depicted in the graph. From another perspective, there is also no need to highlight the information, which is also ignored by the participants in the visual graph condition (since their data is taken as basis in the decision of the content).

Guidelines. simplify the verbal assistance content by focusing on important features

Guideline. highlight the conceptually salient perceptually indistinct entities.

Guideline. do not provide information about perceptually salient conceptually indistinct entities if it is not asked by the participants

10.2.5 Plain and Simple Verbal Content

Many design guidelines originated from different HCI fields draw attention to the principle of coherence. According to Mayer's coherence principle (2001, 2009), "people learn better when extraneous words, pictures, and sounds are excluded rather than included." This principle highlights the importance of providing relevant information and of avoiding redundancies as in line with the previous item. From the instructional point of view, the "perceptible information guideline" of UDI suggests "the design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities". Therefore, the detection of the conceptually salient but perceptually indistinct regions was in the focus of the empirical investigations.

The decision of the verbal assistance content is about finding optimal balance between being *plain and simple* and being *informative and facilitator*. The verbal assistance should be plain and simple due to two main reasons. First, providing all the information concerning the event depicted in the graph would be ineffective. The results of the experiments provided here indicated that explorers are good that acquiring some of the features, i.e. at detecting salient graph shapes, and at making vague estimations about time and value. Therefore this kind of information can be avoided or it can be used to facilitate the acquisition of hard to encode information. This principle of being plain and simple also has support from the graph-text constellation domain, namely, one of the guidelines proposed in Acartürk (2010) stated that "transform complex graph-related sentences into simple ones".

Guideline. provide plain and simple verbal assistance rather than complex structure.

Applying this principle would also help to avoid continuously speaking assistance system and to reduce the communicational load. Besides, haptic exploration is a rapid process, therefore until the full description concerning the attended location is formed and uttered[40], users may start to explore another section. For these reasons, optimal balance between the qualitative and quantitative aspects of the content need to be sustained. Thus, avoiding easy-to-access or easy-to-infer content in verbal assistance, and instead, bringing perceptually indistinct (or underestimated) but conceptually informative concepts into front should be prioritized.

The results reported in Chapter 9.3.2 and 9.3.3 demonstrated considerable effect of verbal assistance content on the haptic explorer's comprehension. The dialogues that contained modifiers (cf. *rich content*) were helpful to the explorer. Modifier presence made the assistance more elaborate; it helped the participant to notice the features of the event, which were currently explored (e.g. steepness of the curve and length, relation with another curve). Besides, the protocols where the explorer initiated a dialogue by asking help resulted in more successful conceptualization of the haptic graphs.

Principle. Taking initiative for requesting help and having adequate verbal assistance enriched by modifiers, rather than just confirmation of the basic spatial properties in response, seems a superb combination for a successful joint activity that inherently requires asymmetric dialogues between two users with different roles; haptic explorer and verbal assistant.

Modeling the graphical features, and verbal utterances (both explorer's request and system's assistance) with respect to <attribute, value> representation seems to be very effective and systematic method, which suits the proposed design. Furthermore, taking the Gricean Maxims into account in the generation of referring expressions (careful selection of the information provided in "attribute pairs", updating attributes gradually and being sure that at the end of the communication target attribute set is covered) seems useful in enhancing the conversational success of the communication (Grice, 1975; Dale, 1992; Dale & Reiter, 1995).

Guideline. provide lacking information or correct the incorrect interpretation to complete the coverage of attributes in "target set" of attributes.

Guideline. provide content enriched with graph-domain terms, relative terms and relational terms (which emphasize size and manner gradually) (since haptic explorers are able to comprehend basic spatial properties through haptic exploration)

10.2.5.1 Amodal Geometric Properties

As widely discussed throughout this dissertation, the shape of the graph line has high importance in the comprehension of the line graph. Therefore the global and local properties of the graph shape should be taken into consideration by the design guidelines. The facilitation of encoding processes by highlighting object shape was one of the issues cited in the ISO 9241-920. The findings of the empirical investigations also pointed out the importance of using shape in referring expressions and providing further details giving reference to a familiar shape.

Guideline. make unfamiliar shape entities familiar by naming, if it is possible.

However, while doing this, the familiar shape concepts should be chosen carefully. For example, for severely and congenital blind users, referring to, for instance, Latin alphabet such as "it has a Letter-M like shape" would lead to misconceptualization since they usually use Braille alphabet, which represent letters different than the Latin alphabet does. This may also cause negative effect on the acceptance of the system by the user,

[40] One of the most effective state-of-art method way of providing content for such dynamic haptic environments is to use incremental natural language generation algorithm (see Lohmann, Kerzel and Habel, 2010).

Guideline. build shape vocabulary which will be employed in such system in cooperation with real target group and sub-groups (see Chapter 5.4)

Besides, the consistency between the modalities of the multi-modal system should be also taken into consideration. Geometric properties, such as size, orientation, shape etc., are mostly amodal properties that can be extracted from vision, haptics and language in a similar way. Therefore maintaining the coherence between these modalities is critical for a multimodal system (ISO 9241-920, 3.1.8 and 4.2). Shah and Hoeffner's (2002) also instruct to make graphs and text consistent with the aim of making graph readers less dependent on prior knowledge.

Guideline. maintain coherence in the representation of amodal properties among all modalities.

Guideline. provide coherence among the scalar entities in all modalities (visual, haptic and verbal). To illustrate, the coherence between the scalar entities provided with verbal assistance and their referred graphical counterparts on haptic graph line or on data labels should be sustained.

This coherence principle is also important for keeping the irrelevant features less salient. For example, the depth of the graph line or the friction provided by the device should be kept homogeneous regarding the all parts of the graph as much as possible. This is a tricky issue considering 3D virtual line graph design.

Guideline. keep the coherence in the representation of irrelevant features as well.

Additional specific design recommendations based on the empirical results regarding the role of specific amodal properties on event segmentation are presented under the title of partonomic granularities.

10.2.5.2 Conceptual Properties

Qualitative ascriptions are easily derived from quantitative information and the use of those ascriptions seems to be very useful in deciding on the perceptual saliency (namely *the distinguishability*) of the conceptually salient entities (Chapter 5.2).

Guideline. calculate *distinguishability* of each conceptually salient entities and highlight the perceptually indistinct ones explicitly.

Beside the basic calculations from raw frequency data, the empirical investigations also provide some insight on this topic. For instance, the findings indicated the global minimum landmarks were mentioned more than the global maximum landmarks. Therefore, special attention would be given to global maxima depending on their distinguishability.

Guideline. If the global maximum is perceptually indistinct from the local maxima, then bring this conceptually important entity into front.

Some additional design guidelines concerning conceptual properties of shape segments and landmarks are also partially elaborated on in upcoming title of "Partonomic Granularity".

10.2.5.3 Data Labels

As the empirical results pointed out, in the absence of data labels, the production of course-grained descriptions that is more conceptually driven and that contain object-oriented properties than action-oriented ones seems to be impaired.

Guideline. facilitate grouping in haptic modality via verbal assistance.

10.2.5.4 Metrical Information

The empirical results indicated that the participants had difficulty in the extraction of metrical information through haptic modality. Providing metrical information is one type of content that needs to be addressed by the verbal assistance system. The findings also indicated that haptic explorers were good at making rough estimation about both time and value (x and y-axis information respectively). However, they had difficulties in relational reasoning. Thus, after segmenting graphs successfully into meaningful parts, providing metrical information for only selected landmarks, which are critical for making relational judgments for that particular graph, may lead to successful communication.

Guideline. provide metrical information for landmarks, which are critical for event segmentation. Providing information for each landmark should be deliberately avoided unless it is explicitly asked by haptic explorer.

10.2.6 Temporal Congruity

Mayer (2009) highlights the importance of temporal congruity in his multi-media design principles ("*temporal contiguity principle*"). This principle is defined as "people learn better when corresponding words and pictures are presented simultaneously rather than successively". Instead of providing haptic exploration and verbal assistance independently and sequentially, providing simultaneous and congruent access to both modalities would lead to successful comprehension. As a more fine-grained design issue, haptic exploration is a fast process and therefore one of the challenges of providing verbal assistance system for haptic representations is to sustain good temporal congruity.

Principle of Temporal Congruity: sustain temporal congruity between the haptic explorer's position or his/her location of request and the verbal assistance.

Several ways of achieving this are described as follows;

10.2.6.1 Sonification

For such a system, the sonification could be employed as a useful method for providing instantaneous alert. Considering that 72.4% of the conversational utterances in the experiment presented in Chapter 9.3.4 contained information about positioning (being on the start point, or on the line etc.), providing this information to the explorer seems crucial for the assistive system; however delivering this information verbally would yield continuously speaking assistance, therefore sonification can be a good candidate to carry this kind of contextually light information. This would also help to sustain temporal congruency.

Guideline. use sonification for the alerts regarding

- Start and end points
- being out of the line or of the frame

10.2.6.2 Haptic Ostensive Expressions (*HOEs*)

The *HOE* is a reference, which involves deictic reference to the referred object by manipulating it haptically (Foster et al., 2008). The use of *HOE*s that anchors haptic explorer's location to the content provided in the verbal assistance is one of the ways of providing temporal congruity. In the investigated setting, only haptic explorers have active role for haptic exploration. Even after requesting an assistance regarding a specific point or region by pointing with *HOE*, the explorer may still continue to explore. Therefore, as the findings also showed that, the verbal assistants tend to omit uttering demonstrative pronouns and when necessary, they use temporal indicators to relate a previously mentioned expression to currently explored region. This preference shown by the verbal assistants may be for preventing that the explorers perform reference resolution incorrectly.

Therefore, in order to minimize the possibility of faulty reference resolution (this item also serves to principle of "error tolerance" stated in UDI by Burgstahler, 2009), the use of the demonstrative pronouns and the interpretation of the haptic ostensive actions performed by the explorers should be designed cautiously.

There are two types of expressions produced in response to haptic explorers' HOEs;

(i) Demonstrative Pronouns, DPs (such as here/this etc.)

(ii) Temporal Pointing, TPs (such as now).

Guidelines.

a. Use demonstrative pronouns during descriptional assistance about intermediate regions
   i.e. This is the global maximum.
b. Use temporal pointing during navigational assistance
   i.e. "Now, you are approaching"
c. Both DPs and TPs are ok for start and end points, however this information can be more successfully carried by sonification, as explained before.
   i.e. "Now, start point." or "this/here is the start point"

10.2.6.3 Taxonomic Granularity

Language production may be too slow to catch up with explorer's action if complete message is intended to be carried in each time. Therefore verbal descriptions should be reduced in an appropriate ways (i.e. being able to provide information in different taxonomic granularity). Finding which information is easy to grasp so that it can be skipped in the verbal descriptions and which information is hard to acquire so that it can be highlighted are the issues investigated in this dissertation. In short, instead of providing all possible information to the graph reader all together, providing it in varying taxonomic granularities would yield more effective environment. This item might also serve to achieve naturalness in a dialogue system.

Guideline. provide shortened expressions over time.

The excerpts given below illustrate two different way of referring to same entity in different taxonomic granularities (excerpts represent a part of Graph-II in Figure 4-6).

- The full description in the first pass:
  "*The population increases with fluctuations in the first third of the graph*".
- The short form in third pass:
  "Again, fluctuations in increasing pattern"

10.2.6.4 Spatio-Temporal Congruity

Acartürk's second principles states that "verbal annotations in graphs facilitate multimodal comprehension of graph-text constellations". This principle is supported by the guidelines that states that "select the location of the annotation appropriately". Those recommendations address the visual graph-text pairs, thus the location of an annotation on graph space is a matter of spatial congruity. Concerning active haptic graph exploration, the spatio-temporal congruity should be sustained since as elaborated before appropriate timing for assistance is integral part of successful communication.

Guideline. choose the location of the verbal assistance carefully; this is dependent on the proper event segmentation and putting event boundaries into most appropriate locations (see the next item).

10.2.6.5 Partonomic Granularity

Chapter 6 is dedicated to the investigation of the segmentation of events (both the graph and the abstract concept). As densely discussed, partonomic relations are crucial part of the event segmentation processes, and appropriate partonomic structures facilitate online comprehension and also memory for later use. According to one of the Mayer's multimedia design principles, namely the "segmenting principle", "people learn better when a multimedia lesson is presented in user-paced segments rather than as a continuous unit". Allowing people actively explore haptic graphs serves to this purpose, and providing verbal assistance with appropriate partonomic structures would enhance their conceptualization. The graph (at that, the abstract event) can be segmented in different granularity levels from fine-grained to course-grained. The granularity of the partonomic relations should be adjusted with respect to user' needs and previous verbal assistance content. The findings indicated that the haptic explorers performed segmentation mostly on fine-grained (action-based) level, rather than course-grained (conceptual) level. Therefore, the assistance system should highlight the abstract concept explicitly.

Guideline. adjust the granularity of the partonomic relations with respect to user' needs and previous verbal assistance content.

The empirical results concerning the effect of specific amodal properties on event segmentation also provide useful recommendations to sustain this principle.

- The number of the segment on the graph line (i.e. the number of direction or trend change) seems to be an indicator of how many sub-part the explorer divides the graph line into, however this parameter also interacts with the global graph shape indicating that relying on just these two parameters for the segmentation of an event would not be convenient.
- Steeper segments and more acute landmarks are used as event boundary, less error were also committed for these regions too. First, for steep segments direction is easy to grasp therefore instead of this information the rate of change may be additionally provided. There is one exception to this, if the graph line has two consecutive steep segments in opposite directions, then the direction should be emphasized.
- For medium segments, the direction is still easy to grasp, but modifier that defines the rate of change would be beneficial.
- Slighter and obtuse angles should be handled carefully Most problems were observed about the direction of slight and horizontal segments. Therefore for these regions the presence of the change and also its direction should be highlighted.

Like shape segments, shape landmarks also play crucial role in event segmentation. Haptic explorers usually use the landmarks that introduce direction change (*sp2* in Figure 10-2) as an event boundary, however but not the landmarks that introduce a change in the trend (*sp1* and *sp3*).

- Thus, the trend changing points, if they carry conceptually important information, are one of the points that need to be highlighted.

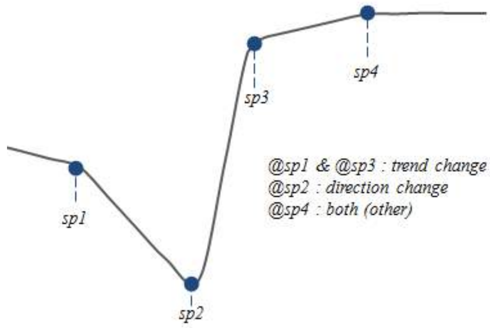


Figure 10-2 Qualitative shape landmarks

Causal attribution is a frequent phenomenon concerning graph reading, and the comprehension of such above mentioned trend changing locations is important for making causal attribution. For example, Figure 10-3 depicts an event boundary, which is suitable for causal attribution. The point where the population starts to remain stable might be conceptually important point, since one can infer that an important event happened at that time causing an abrupt change in the observed data (namely, the tourist- visit count). This was also exemplified in sentence (1), a verbal description produced by one of the participants.

(1) "Due to very famous festival is held in this city, the visit count reached to maximum value and stayed there during festival time."

Figure 10-3 A graph sample to illustrate a causal attribution

The investigation of causal attribution from haptic graphs would also provide valuable practical information for a verbally assisted system, but it was left aside in the scope of this dissertation, see Acartürk (2010) for the investigation of causal attribution in visual graph and text constellations.

The global points and local points were mentioned densely in verbal descriptions indicating that even in the haptic modality, graph readers may acquire this information but the conceptualization of other smooth points seemed to be problematic (such as *sp4* in Figure 10-2), this smooth points involve abovementioned trend changing landmarks and also the landmarks with horizontal connections.

- Therefore, same heuristics also apply to this; if they are conceptually important then they should be supported with verbal assistance.

Furthermore, according to ISO 9241-920, the haptic objects should be sufficiently simple to be recognized without performing long exploration (i.e. providing more information with zoom function, if necessary). The appropriate use of partonomic structures can be used to apply this guideline too. Since haptic explorers are good at segmenting the graph at the fine-grained, starting with course-grained verbal assistance and going into detail when it is requested would be one applicable method. However, as previously reported, for some problematic regions, in which the participants had difficulty to segment due to indistinct perceptual features, starting with fine-grained assistance and continuing with course grained content would yield more successful comprehension of the event depicted in a haptic graph.

10.2.7 Perspective-Taking and Alignment

The instruction ("*make graph reading metacognitive*") proposed by Shah and Hoeffner (2002) highlights the importance of facilitating graph domain reading addressing visual context. According to this instruction, the many of the errors in graph reading occurs when the readers approach the graph reading as just retrieval task, therefore the readers should be reminded that graph reading is an interpretation and evaluation task as well.

The occurrence of misreading has been well exemplified in haptic graph reading. As it has been densely investigated in Chapter 9, one of the critical issues is the misreading of the changes of the event depicted in the graph during right-to-left haptic exploration. Haptic explorers' first preference is to focus on basic spatial properties of the graphs by employing an action-perspective. However, correct mental representation of the abstract event depicted in the graph can be achieved through adoption of spatio-temporal perspective. After 1 or 2 graphs participants' succeeded to overcome this, however considering the scenario like the graph is presented in a multi-modal report that involves text, diagrams and also statistical graphs. In that case, even if the participant knows that the right-to-left reading is not useful, dominance of action-perspective induced by active and sequential haptic exploration may interfere and hinder the spatio-temporal characteristics of line graphs. Therefore making this knowledge explicit would prevent possible misunderstandings.

The Principle of Alignment at Spatio-Temporal Perspective: The verbal assistance system should align itself to explorers' current perspectives for the sake of communication; however the explorer's perspective is not always the most efficient perspective for conceptualizing the graph, but assistants' use of *non-embodied-observer perspective* may help to activate abstract spatio-temporal representations.

However, in addition to the need of graph-domain content, the assistance might be also needed with respect to the use of haptic device. Communicating about how to navigate on the line and current positioning on the graph frame should be taken into consideration by the verbal assistance system as well. The content of help requests and verbal assistance were categorized into two main categories w.r.t communicative intention; navigational (or instructive) and descriptive. The navigational utterances focus on the line to be explored, such as navigational help like "You should go up" and positioning question like "Am I at the beginning?". On the other hand, the descriptive utterances mainly convey content about a domain, e.g. 'bird population'. The results presented in 9.3.1.2 pointed out that the choice of subject pronoun (*potentially self-referential* ("*you*" and "*I*") versus *non-self-referential* ("*it*")) is an indicator of which perspective is adopted. In brief, the use of appropriate person pronoun may help users adopting most appropriate perspective for the communicative task at hand and activating appropriate graph schemata as well[41]. The guidelines to achieve this are listed as follows;

Guidelines:

1. During "left to right" exploration,
   a. use *non-self-referential* (3rd person) pronoun or the noun (it or the population) for a descriptional assistance about shape segments instead of using *potentially-self-referential* (2nd person) pronoun.
      i.e. The population/It increases with fluctuations in the first third of the graph.
   b. use *non-self-referential* (3rd person) or *potentially-self-referential* (2nd person) pronoun for a descriptional assistance about shape landmarks

[41] These guidelines are constructed based on the choice of subject pronoun during the collaborative activity

i.e. "It is the global minimum of 55" or "Now, you are at the global minimum of 55."

c. use *potentially-self-referential* (2nd person) pronoun for a navigational assistance (i.e. in the cases where the explorer get stucked at some point on the graph line).

i.e. "You can go down-left" , "You are out of the line"

d. use graph-domain vocabulary to describe the changes

i.e. increase, decrease, fall

2. During "right-to left" exploration,

a. don't use descriptional assistance for shape segments

b. use *non-self-referential* (3rd person) or *potentially-self-referential* (2nd person) pronoun for a descriptional assistance about shape landmarks

i.e. "This is the global minimum of 55" or "Now, you are at the global minimum of 55."

c. use *potentially-self-referential* (2nd person) pronoun for a navigational assistance.

i.e. You are now approaching to the global minimum.

d. use exploration-induced vocabulary to describe the changes

i.e. descend, reach, approach

Maintaining coherence in the use of subject pronoun with respect to communicative goal is also crucial to provide consistency and avoid unnecessary cognitive demands.

Guideline. maintain consistency in the use of subject pronoun with respect to communicative intention of the given assistance.

10.2.8 System Initiated Verbal Assistance

The items mentioned here were already discussed in the items of *Individualization* (10.2.1) and *Conceptual Properties* (10.2.5). However, these recommendations also contribute to decisions of providing system initiated assistance, hence they are touched upon again. In verbally assisted haptic graph comprehension system, explorers and verbal assistants have asymmetric roles, and haptic explorers were asked to initiate a dialogue when needed. However, as the findings indicated, there are problematic regions, which are hard to grasp in haptic modality, and the explorers may not realize that these regions are conceptually important. For such cases, the verbal assistance system should take the initiative. Such problematic regions can be derived from the calculation of the distinguishability of the qualitative ascriptions, incorporating the user' exploration patterns (i.e. detecting a need for assistance).

Guideline. provide verbal assistance without waiting haptic explorer's taking initiative for the areas, which are conceptually salient but perceptually indistinct.

Guideline. for the detection of whether the user needs help, use the visit count (in a limited window), the number of back-and-forth movements. Also individualization in terms of identifying the user and being adjusted to that particular user's haptic patterns would be highly beneficial (see Chapter 9. 3.4)

10.2.9 The Division of Labor among the Modalities

The division of labor among modalities by taking each modality's strong and weak points regarding the information that needs to be communicated is the main idea of this system. For example, according to the "modality principle" introduced by Mayer (2001, 2009), "people learn better from graphics and narration than from animation and on-screen text." Utilizing all relevant modalities in the presentation of information was also mentioned in Sjöström's guidelines on the use of haptic device and space (2002). For verbally assisted haptic graph comprehension system, the information is carried by three different modalities;

Guideline. allow haptic explorer to explore the graph actively in haptic modality, in which the rough estimation regarding amodal geometric properties and graph shape are successfully conveyed.

Guideline. use verbal assistant to provide exact metrical information and also to highlight hard to notice important regions specific for each graph, namely conceptually salient but perceptually indistinct graphical entities.

Guideline. use sonification for instantaneous alerts (such as being on the start point or out of the graph line).

Furthermore, it is also important to avoid redundancy in the communication. This issue was addressed by Mayer (2009) as a “redundancy principle” which states that “people learn better from graphics and narration than from graphics, narration, and on-screen text. The haptic explorers were very good at acquiring some type of information (such as the direction of the steep segments, or the landmarks with acute angles) therefore no need to repeat this knowledge.

Guideline. Avoid redundancy in the information presented in different modalities.

10.2.10 Limiting complexity

As a final point, limiting the complexity is another important issue, which is specifically addressed in ISO 9241-920. This item also addresses one of the Shah and Hoeffner’s (2002) instructions on graph usage, namely “reduce working memory demands”. Two guidelines can be suggested on this issue;

Guideline. Do not use too many varieties in the sonification categories.

Guideline. Do not use too many varieties in the vocabulary used to refer the changes (such as increase), states (such as peak etc.) and modifiers (such as steep). For example, if too many different expressions were used to refer the same steepness value (i.e. steep, abrupt, near vertical, and also as a negated statement like it is not slight, etc.), the participant might have difficulty in categorization and recall. Therefore a small set of vocabulary should be defined and used consistently.

10.3 General Discussion

This chapter presents *evidence-based design heuristics*, which were constructed through a series of empirical investigations. The fundamental principle of “the verbally assisted haptic graph comprehension system” was set as “to provide relevant information in a relevant way in a relevant time”. The guidelines and recommendations with the purpose of satisfying this goal were presented throughout this chapter.

Designing of such multi-modal interface with the aim of providing assistance to visually impaired people touches several main stream HCI domains, such as instructional and assistive design, haptic interfaces, dialogue systems etc. The evidence from the empirical investigations were combined with the existing guidelines in those abovementioned HCI domains in order to reach comprehensive design heuristics, of which the human cognitive processes concerning graph comprehension and communication over those representations constituted the backbone.

CHAPTER 11

General Conclusion and Future Studies

This dissertation consists of three main sections. In the first section, I present the "state of the art" of multi-modal communication, with special focus on haptic graphs and introduce the topic of verbally assisted haptic graph comprehension system. The second section is dedicated to the empirical investigations of haptic graph comprehension. The final section addresses the verbally assisted haptic comprehension system focusing on the dynamics of a task-oriented joint activity and it also contains design guidelines constructed based on the empirical findings. Latter two sections are designed as having a stand-alone structure that consists of the introduction of the research topics, the empirical investigations concerning those topics and the conclusion.

In this final chapter, I do not intend to provide summarization of the empirical research. The other way around, I intend to recapitulate the reasons of pursuing this line of research and the challenges. Next, the main contributions of this study, which were already discussed, are presented in brief. Finally, I conclude with the discussion of short-term and long-term future plans.

Statistical line graphs are one of the commonly used multimodal communicational settings when they are accompanied with verbal annotations and text in written format, or verbal descriptions and gestures in spoken communicational format. Graphs enable extraction and comparison of data points. However, graph reading is not a just data extraction or retrieval task. They can be used for making evaluation, inference and extrapolation that require reasoning over conceptual aspects of the events depicted in the graph (Shah and Hoeffner, 2002). For example, line graphs are particularly good at carrying trend information and they elicit extraction of second order entities; such as extreme values, trends, or changes in trends. This advantage can be ascribed to pattern perception processes in the human brain, such as visual chunking (see Shah, Mayer and Hegarty, 1999). Acquiring such knowledge is crucial also for blind or visually impaired people, however designing efficient environment for this purpose is a challenging issue.

The detection of higher-order concepts and relations are closely linked to shape properties of the graphs. Besides, the shape is one of the dominating criteria in labeling which is important both for online comprehension and memory. For visually impaired people, comprehension of haptic line graphs is based on local and sequential exploration processes with the goal to collect information provided by the geometrical properties of the line explored. Therefore, during haptic graph exploration, detected shape properties are used to anchor concrete graph entities to abstract event. Within the scope of this dissertation, I put particular emphasis on *the shape* of the graph line.

Graphs depict abstract concepts and transform these abstract concepts into static concrete objects by utilizing graphical entities such as graph shape, axes, data labels etc. Due to simultaneous visual perception of both global and local aspects of the graphs, we perceive visual graphs as static objects. On the other hand, although haptic graphs are also static objects that represent same abstract event, they are perceived in a sequential manner through active exploration. Therefore perception of haptic graphs resembles to event perception. To that end, the term "*event*" has meanings in two different layers when it comes to haptic graph comprehension (i) the conceptual event layer and (ii) the event-like perception of haptic graph shape. The event perception is particularly important since successful communication over graph depends on proper segmentation of the event with respect to its global and local (*perceptual* and *conceptual*) saliencies in varying partonomic and taxonomic levels. But despite its importance, it is one of the many untouched topics of haptic graph comprehension research.

Due to differences between visual and haptic modalities, informational or functional (computational) inequivalences are inevitable. To bridge the gap caused by those differences and to present coherent information to haptic graph readers, haptic graphs should be accompanied by alternative modalities.

11.1 Main Contributions

This dissertation focuses on providing visually impaired users with a verbal assistance as an accompanying modality to their haptic graph exploration. The verbally assisted haptic graph comprehension system is based on the interaction between two agents; a human explorer who perceives the graph haptically through active exploration and a verbal assistance agent that provides helps to the explorer in a instantaneous and automatic manner. The latter could be a human or an automated system. Thus, my approach for studying these issues is to incorporate ideas from Human Computer Interaction and Cognitive Science. Successful design of such system can be achieved only with an interdisciplinary perspective that incorporate several mainstream research fields such as "graph comprehension", "event segmentation", "referring expression production by humans and generation by automated systems", "dynamics of a joint activity" and "HCI aspects for efficient and effective system design". There is a huge amount of literature about each of these research areas regarding visual or language modality. However, only a few studies exist regarding the haptic modality except the last item[42].

The contributions of this dissertation are classified in the following under three parts; (i) the conceptual architecture of the task-oriented joint environment (ii) the empirical contributions for the fields of "graph comprehension", "event segmentation", "gesture-language-space research" and "dynamics of collaborative environments" and (iii) the implications, which are derived from those findings for HCI research.

11.1.1 The conceptual architecture of the task-oriented joint system

In this dissertation, a task-oriented collaborative environment with the aim of providing assistance to visually impaired people was presented. The conceptual architecture of this design was constituted by taking human cognitive processes and existing design guidelines into account. This task-oriented collaborative design of the proposed verbal assistance system, involves two main components, namely one component responsible for providing verbal-assistance (by the system) and another one for the active haptic exploration by the explorer. While the active exploration component allows the users to discover the embodied graph shape and its details by their own actions, the verbal assistance component helps the users to label those explored regions, and also helps the comprehension of conceptually important entities, which is tailored to their needs. In order to provide successful conceptualization of the graph through efficient

[42] This will be elaborated in 11.1.3.

communication, empirical studies are employed in gaining a better understanding of haptic graph comprehension and of collaborative activity dynamics towards the design of a verbally assisted system for visually impaired people.

11.1.2 Empirical Contributions

Perceiving and comprehending graphs and also communicating over them require the involvement of various sensory (representational) and communicational modalities. A systematic investigation of the interaction between modalities in communication through graphs plays an important role in the realization of an automatic verbal assistance system. Besides, the multi-modal research method used in the series of experiments consisted of both linguistic and non-linguistic methodologies[43].

The empirical findings are summarized and discussed from both theoretical and applied-research perspectives in the interim and general conclusions of the empirical investigation-oriented chapters (Chapter 5, 6, 7 and 9). In brief, the findings contribute to our understanding of several research topics and those are listed with the main findings as follows;

- Haptic graph comprehension
    - Without assistance, the haptic explorers have difficulties in conceptualization of the graph dues to having incomplete knowledge in their graph schemata, which has a key role in establishing appropriate mapping between the graphical features and the conceptual event.
    - Amodal geometric properties (such as shape, size, orientation etc.) have a strong significant effect on event segmentation and description.
- Gesture-Language-Space Relation
    - The sensory modality of graph reading has an effect on gesture production. Haptic exploration and gesture production shares underlying mechanisms of motor plans and actions, therefore the investigation of their relation arises new questions.
- Dynamics of a task oriented joint activity of haptic exploration
    - Taking initiative in requesting help and having adequate verbal assistance enriched by modifiers, in a response, seems a superb combination for a successful joint activity that enhances graph comprehension.
    - Haptic exploration patterns (the speed, the number of back-and-forth actions on the graph line, the qualitative ascriptions etc.) can be used to detect assistance need in an instantaneous and automatic manner.
    - Coming into alignment at the situation level and the appropriate choice of frame of reference with respect to the communicative goal at hand is crucial.

11.1.3 Implications

As well as the empirical contributions, the implications addressing two HCI related topics were in the focus of this dissertation: (i) the use of gesture analysis as research paradigm in HCI and (ii) developing evidence-based design guidelines based on the empirical findings. The contribution of gesture analysis as a research paradigm for HCI was discussed in Chapter 8. The analysis of gestures has been proved to be very efficient HCI tool to understand how the users conceptualize the graphs. Besides, they were valuable complement for verbal descriptions in data analysis i.e. in resolving ambiguities of referring expressions.

[43] Gestures and sketches are forms of spatial representations, so do the graphs. Therefore, incorporating those non-linguistic methods aided to draw language-independent conclusions.

Moreover, the integrated knowledge obtained from all empirical investigations was used to constitute design guidelines, in specific for verbally assisted haptic graph comprehension and in general for shape based haptic representations and verbal assistance constellations (Chapter 10). Recently, considerable attention has been paid to haptic studies in the HCI domain, since haptic devices are becoming widespread. However, none of them addresses such cognitive process oriented recommendations for a system that incorporates haptic representation and accompanying verbal assistance, which is tailored to user's needs.

11.2 Short-Term Future Studies

To my knowledge, this dissertation could be the first systematic study pursuing such integrated approach for haptic shape and verbal assistance constellations in the fields of Cognitive Science and HCI. The investigation of the haptic modality from the perspectives of both HCI and theoretical research lead to empirical findings reported here and those findings make new research questions arise. Here, several short-terms plans are summarized.

The first short-term plan concerns the investigation of the features afforded by the haptic space and the device. Successful interpretation of the event presented in the graph requires appropriate mapping between the physical properties of the graph and the conceptual domain depicted by the graph. Regarding top-down conceptual factors, having appropriate graph schemata that guide explorer about which feature is important and relevant for the comprehension of haptic graphs is critical. Regarding bottom-up perceptual factors, the geometric properties of the line shape have pivotal role in the comprehension of the abstract event and also for deciding on the content of multimodal assistance that accompanies to haptic exploration. Within the scope of this dissertation, I mainly focused on the contribution of relevant geometric graph features leaving the effect of irrelevant ones (i.e. length and depth) aside. As briefly mentioned in Chapter 1.3 and Chapter 7, length seems to be inherently correlated with the exploration time since haptic exploration is performed through actions. In more detail, the abrupt changes, which are represented by long segments, will be explored longer in time as well. It has been also shown that time dominates space in spatial reasoning concerning layout and associations for many visually impaired people (Golledge, 1993). Combining these with the matters of haptic illusions, the length property might be considered as perceptually salient property although conceptually it is irrelevant. The investigation of such perceptually salient but conceptually irrelevant features on haptic graph comprehension would provide additional design guidelines for achieving less error prone environment.

The second short-term plan concerns the cognitive modeling of haptic graph comprehension. In this dissertation, I did not go into detail of this topic since my focus was in the verbally assisted haptic graph comprehension from the HCI perspective. As discussed in Chapter 7, the empirical findings presented here have potential to extend the existing graph comprehension theories for visual graphs by shedding light into how the graph comprehension occurs in haptic modality. Chapter 1.3 provides summarization of those graph comprehension theories (Freedman and Shah, 2002; Lohse, 1993; Pinker, 1990). Although the exploration modality is different, it seems that haptic graph comprehension can fit to these general theories, but more detailed research is needed to understand the differences in the exploratory patterns for different tasks. In order to have conclusive results, one of the most preferred ways of investigating visual graph comprehension is to employ the eye tracking research paradigm. The counterpart of eye movement analyses in the haptic modality can be considered as the analyses of haptic exploration movements. Although the speed of eye movements and the speed of hand are not comparable, the use of, for example, mouse movements as an indicator of attended location is also one of the commonly used methods in HCI (e.g. Quek, et al., 2002). Both of them provide sequence of attended locations, time of exploration for each region (are of interest), speed, back and forth movements between AOIs. Following the experimental paradigm of the study presented in Chapter 6, both eye movement of visual participants and haptic exploration movements of haptic participants were recorded when they were performing a joint activity of graph exploration. The comparative analyses would provide valuable insights about how much the models of graph comprehension explain haptic graph comprehension to what extent and also would help to reach less modality-independent graph comprehension theories.

Another short-term plan is to build a multimodal corpus from the variety of behavioral data gathered through four empirical investigations. In all experiments, the participants produced post-exploration summaries and in only one experiment, they performed joint activity. The behavioral data were annotated in accordance with the annotation schemes introduced in Chapter 4.5. The reliability of the annotations was tested by conducting interrater reliability analyses. The corpus is planned to have two parts. First part (CI) consists of richly annotated data of post-exploration (*off-line*) verbal summaries of visual (with/without labels) and haptic graphs, and of speech accompanying gestures. The behavioral data (haptic exploration actions for haptic explorers and gaze data of visual graph readers) collected during on-line graph reading were also recorded and linked with the corresponding post-exploration data. Furthermore, verbal descriptions' expressivity concerning "how well the descriptions describe the graphs" enriches the content of the corpus. A second part (CII) contains behavioral data gathered from participant pairs (a haptic explorer and an observing verbal assistant) performing task-oriented joint activity of haptic graph exploration. This multi-modal data contains dialogues, and concurrent acts such as haptic exploration patterns of explorers and eye movements of verbal assistants on the graphs during the exploration. The behavioral data, which will be released in the multi-modal corpus of haptic graph comprehension, can be listed as follows;

- Referents' perceptual and conceptual properties
- Post-exploration descriptions (CI) and dialogues (CII) (annotated w.r.t semantic annotation scheme)
- Dialogue acts (such as communicational goal, turn taking acts.)
- Expressivity scores of each verbal description (CI)
- Speech-accompanying gestures (directionality, space, handedness etc.) (CI)
- Haptic exploration actions (location, speed, visit history etc. w.r.t region-of-interests) (CI, CII)
- Haptic ostensive expressions and actions (CII)
- Eye movements (location, mean duration, visit count etc. w.r.t region-of-interests) (CI, CII)

Here, several research areas, where such multi-modal corpus would be highly beneficial, are listed;

- Referring expression production and generation (i.e. the investigation of the choice of modifiers based on the contrast in the spatial properties across modalities)
- Multi-modal reference resolution
- Gesture-language relation
- Sensory modality and gesture relation
- Alignment (at linguistic and situation model level) between interlocutors
- A design of verbal assistance systems focusing on shape-based representations

Most of the experiments were conducted in Turkish and one experiment was conducted in German. Due to this difference concerning the experimental language, the empirical findings based on linguistic data were not compared and the cross-linguistic analysis was not intended within the scope of this dissertation at the first place. However, the abovementioned corpus consisting of German and Turkish verbal data would be interesting for the investigation of cross-linguistic issues as well.

11.3 Long-Term Future Studies

Within the scope of this dissertation, 1-line graph comprehension and its challenges with respect to haptic perception were addressed. A long-term planned study concerns providing visually impaired people with verbal assistance for multi-line haptic graph comprehension (Habel and Acartürk, 2012). As discussed in Chapter 1, graphs are produced based on several design decisions. The type of data (discrete or continuous) or the message (data point comparison, or extraction of trends and relations etc.) affect the decision of graph type, for example bar graphs are good at carrying discrete information, and pie charts at carrying proportional data. Additionally, line graphs are efficient for the extraction of trend information, and multi-line graphs, on the other hand, are preferred to highlight the relation between two sequence of data represented with two trend-lines. As expected, conceptualization of multi-line graphs, particularly through haptic modality, is more complex than 1-line comprehension. Concerning a perception of multi-line haptic graphs, the distinction between the two lines and switching between them are at the one side of this complexity. However, more critical issues lay down in the conceptual level. The individual lines of a multi-line graph can be explored in a similar way with 1-line graphs, which is investigated in this dissertation. However, multi-line graph reading requires different strategy than reading of data values and extraction of relations that belong to the same data line. In 2-line graphs, the relation between the two data line should have a meaning to carry, otherwise they could be given as two separate graphs.

Reading multi-line graphs requires high demands on working memory for keeping the information of salient (perceptually or conceptually) parts in mind and comparing them with sequential exploration with back-and-forth movements, plus switching between the two lines. If there is, the crossing of the lines introduces an extra challenge for the conceptualization.

Figure 11-1 illustrates two samples for "*two-line graphs*". In those samples, one line represents the monthly representation of 30-year average of precipitation, whereas the second line corresponds to monthly precipitation data for the specific year. Haptic perception is based on exploration and thus the properties of shape are foregrounded. The left-graph depicts the averaged data and the data sampled from 2014. The trend-lines show parallelism although the difference between the two lines varies in time. Thus this graph facilitates the extraction of statements like "the precipitation amount in 2014 was higher than the average (in general/ for each month)". Visual graph readers may notice this higher-order function difference as distance between the lines, however haptic exploration does not support this sort of detection.

Besides, in such cases, the extraction of conceptually important relations can be particularly overlooked due to strong effect of the perceptual organization, also known as Gestalt principles of "Parallelism" and "common fate". For instance, for this particular graph, the highest precipitation in average was observed in April, whereas it was observed in November for the year of 2014. This kind of switch may have conceptual significance and be essential for making causal attribution i.e. regarding seasonal changes. This introduces a challenge in terms of graph comprehension. When visual graph readers are primed to importance of such points, they may reach to correct interpretation. However, due to sequentiality of the haptic exploration process, this would be still challenging for haptic readers.

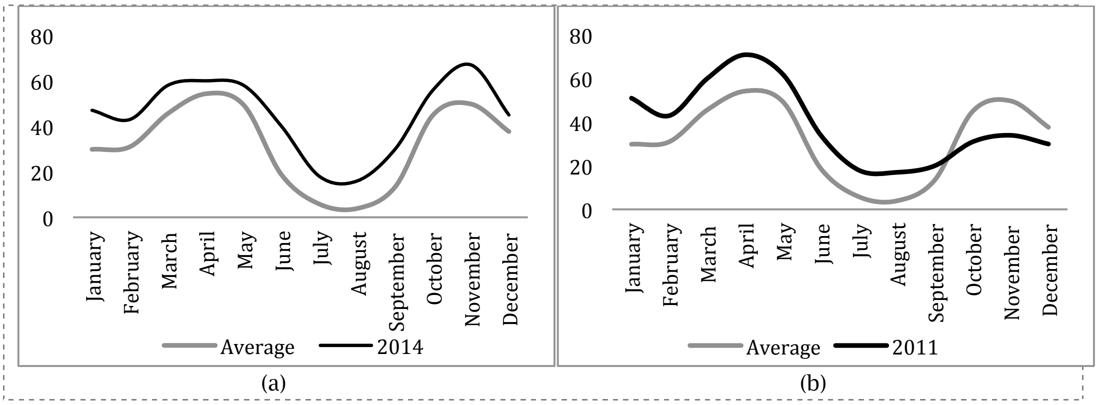


Figure 11-1 two samples for a multi-line statistical graph (the data were retrieved from Turkish State Meteorological Service and redrawn)

The empirical findings indicated that the haptic explorers tended to produce fine-grained (action-based) verbal descriptions. However, comparing shape entities that belong to one data line is different than comparing two data lines. And for the latter, course grain conceptualization of both trend lines would be more efficient in that respect. The analysis of verbal descriptions would also provide valuable insights regarding the segmentation of two relational events. The choice of description such as *in sequential manner* (describing them one by one without focusing on the relation) or *in integrated way* (describing the relation without focusing on the distributions of the individual lines) would be used as indicator of how they conceptualize the event and how well their mapping between the graph and the message is. Furthermore, the empirical findings also indicated that the speech-accompanying gestures produced for the haptically explored graphs exhibited non-segmented/smooth pattern, and the shape landmarks can be easily ignored. However, the intersection points (as illustrated in Figure 11-1b) may provide distinct landmarks for segmentation. In this example, the intersection point does not carry any distinctiveness for the average year's distribution and it introduces a slight change for the distribution of 2014. However, extracting the meaning of this crossing is important for understanding 2-line graphs. The analysis of gestures (i.e. looking at whether the explorers use two hands, each of which represents one line) might provide very valuable complementary data in understanding how haptic readers conceptualize them. One of the study from our research group, we (Alaçam, Habel and Acartürk, 2013) investigated how the presence of incongruence between consecutive (visual) bar graph pairs influences conceptualization of the represented information about precipitation. Gestures and verbal descriptions produced by the participants were used as indicators of event conceptualizations. The graphs represented average precipitation data of various cities. The first graph represented the monthly representation of 30-year average of precipitation. After the graph disappeared, the second graph that represented monthly precipitation data for the specific year was presented. Then, the participant was asked to give a verbal description by taking both graphs into account. Although this study addressed the comprehension of two events depicted in visual graphs, the findings are relevant especially for one aspect. In the condition where two lines are given simultaneously in the same haptic space, the conceptualization of the relation between two lines are highly dependent on the memory component, similar to the case which is investigated in this study. The results of this study revealed that when incongruent graph pairs are presented, the participants produced more directional gestures for the specific-year graph compared to the overall graph, while no difference observed between the two congruent graphs. The increase in the number of directional gestures was considered as a likely indicator of a different conceptualization. To sum up, the cognitive demands and the task itself concerning 2-line graph reading are quite different from 1-line graph reading. Implementing this on the investigated conceptual architecture and conducting empirical research following the current multi-modal method may extend our understanding of in general "*graph*" and in specific "*haptic graph*" comprehension.

Bibliography

Abu Doush, I., Pontelli, E., Simon, D., Son, T. C., and Ma, O. (2009, October). Making Microsoft Excel™: multimodal presentation of charts. In Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility (pp. 147-154). ACM.

Abu Doush, I., Pontelli, E., Son, T. C., Simon, D., and Ma, O. (2010). Multimodal presentation of two-dimensional charts: An investigation using Open Office XML and Microsoft Excel. ACM Transactions on Accessible Computing (TACCESS), 3(2), 8.

Acartürk, C. (2010). Multimodal comprehension of graph-text constellations: An information processing perspective. Doctoral dissertation, University of Hamburg, Hamburg. Retrieved from http://ediss.sub.uni-hamburg.de/volltexte/2010/4501/

Acartürk, C. (2012). Points, lines and arrows in statistical graphs. In P. Cox, P. Rodgers and B. Plimmer (Eds.), Diagrams 2012, Lecture Notes in Artificial Intelligence (Vol. 7352, pp. 95-101). Berlin Heidelberg: Springer-Verlag.

Acartürk, C. (2014). Towards a systematic understanding of graphical cues in communication through statistical graphs. Journal of Visual Languages and Computing, 25(2), 76-88.

Acartürk, C. (2012). Referring expressions in communication through line graphs: A comparative analysis of verbal descriptions. In Proceedings of the Workshop on Discourse Structure in Turkic Languages.The 16th International Conference on Turkish Linguistics.

Acartürk, C., Alaçam, Ö., and Habel, C. (2014). Developing a Verbal Assistance System for Line Graph Comprehension. In A. Marcus (Ed.): Design, User Experience and Usability (DUXU/HCII 2014), Part II, (pp. 373–382). Berlin: Springer-Verlag.

Acartürk, C., and Alaçam, Ö. (2012). Gestures in communication through line graphs. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 66-71).

Acartürk, C., Habel, C., and Cagiltay, K. (2008). Multimodal comprehension of graphics with textual annotations: The role of graphical means relating annotations and graph lines. In J. Howse, J. Lee and G. Stapleton (Eds.), Diagrammatic representation and inference: Lecture notes in computer science (Vol. 5223, pp. 335-343). Berlin/Heidelberg: Springer.

Acartürk, C., Habel, C., Cagiltay, K., and Alaçam, Ö. (2008). Multimodal comprehension of language and graphics: Graphs with and without annotations.Journal of Eye Movement Research, 1(3), 2.

Akıllı, G. K. (2004). A proposal of instructional design/development model for game-like learning environments: The FID2GE model (Doctoral dissertation, Middle East Technical University).

Alaçam, Ö., Habel, C., Acartürk, C. (2013a) Investigation of Haptic Line-Graph Comprehension Through Co-Production of Gesture and Language. Tilburg Gesture Research Meeting, Tilburg, Netherlands.

Alaçam, Ö., Habel, C., Acartürk, C. (2013b) Towards designing audio assistance for comprehending haptic graphs: A Multimodal Perspective. In C. Stephanidis and M. Antona (Eds.), UAHCI/HCII 2013, Part I, LNCS Lecture Notes in Computer Science (Vol. 8009, pp. 409-418). Heidelberg: Springer.

Alaçam, Ö., Habel, C., Acartürk, C. (2013c). Gesture and Language Production in Communication through Bar Graphs. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Austin, TX. pp. 1714–1719. Cognitive Science Society

Alaçam Ö., Acartürk C., and Habel C. 2014. Referring Expressions in Discourse about Haptic Line Graphs. In Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue. SemDial 2014 – DialWatt. Verena Rieser and Phillipe Muller (eds.). (pp. 7–16).

Alaçam, Ö., Habel, C. and Acartürk, C. (2014). Verbally Assisted Haptic Graph Comprehension: The Role of Taking Initiative in a Joint Activity. In the Proceedings of the 2st European Symposium on Multimodal Communication, University of Tartu, Estonia, August 6-8, 2014.

Alaçam, Ö., Acartürk, C. and Habel, C. 2015. Haptic exploration patterns in virtual line-graph comprehension. In A. Marcus (Ed.): Design, User Experience and Usability (DUXU/HCII 2015), Berlin: Springer-Verlag.

Alaçam, Ö., Habel, C. and Acartürk, C. 2015. Switching Reference-frame Preferences during Verbally Assisted Haptic-graph Comprehension. In the Proceedings of 6th International Conference on Spatial Cognition (Cognitive Processing: DOI :10.1007/s10339-015-0730-9)

Ali, N., and Peebles, D. (2013). The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. Human Factors: The Journal of the Human Factors and Ergonomics Society, 55(1), 183-203.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. Spatial Cognitionand Computation, 5, 307-331.

Alibali, M. W., Bassok, M., Solomon, K. D., Syc, S. E., and Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. Psychological Science, 10, 327-333.

Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. Languageand Cognitive Processes, 15, 593-613.

Allen, W. C. (2006). Overview and evolution of the ADDIE training system. *Advances in Developing Human Resources*, 8(4), 430-441.

Alonso-Arevalo, M. A., Shelley, S., Hermes, D., Hollowood, J., Pettitt, M., Sharples, S., and Kohlrausch, A. (2012). Curve shape and curvature perception through interactive sonification. ACM Transactions on Applied Perception (TAP), 9(4), 17.

Amedi A, Malach R, Hendler T, Peled S, Zohary E, 2001 ``Visuo-haptic object-related activation in the ventral visual pathway'' Nature Neuroscience 4 324-330

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., and Golledge, R. G. (2004). Functional equivalence of spatial representations derived from vision and language: evidence from allocentric judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(4), 801.

Baddeley, A. D., Grant, S., Wight, E., and Thomson, N. (1973). Imagery and visual working memory. In P. M. A. Rabbitt and S. Dornic (Eds.), Attention and performance V (pp. 205–217). London: Academic Press.

Baldwin, D. A., and Baird, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), E arly social cognition (pp. 215-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bard, E., Hill, R., Foster, M., and Arai, M. (2014). Tuning accessibility of referring expressions in situated dialogue. *Language, Cognition and Neuroscience*, *29*, 928-949.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and brain sciences, 22(04), 577-660.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364* (1521), 1281-1289.

Beilock, S. L., and Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. Psychological Science, 21(11), 1605-1610.

Bennett, J. (1966). Whatever the Consequences, *Analysis*, 26, 83–102.

Ben-Tal, O., Berger, J., Cook, B., Daniels, M., Scavone, G., and Cook, P. 2002. SONART: The sonification application research toolbox. In Proceedings of ICAD.

Bergen B, Chang N. (2005) Embodied construction grammar in simulation-based language understanding. In: Östman JO, Fried M (eds) Construction grammars: Cognitive grounding and theoretical extensions. Benjamins, Amsterdam, pp 147–190

Bergman, E., and Johnson, E. (1995). Towards accessible human-computer interaction. Advances in human-computer interaction, 5(1), 87-114.

Bergmann, K., Kahl, S., and Kopp, S. (2013). Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In Intelligent Virtual Agents (pp. 203-216). Springer Berlin Heidelberg.

Bergmann, K., Kahl, S., and Kopp, S. (2014). How is information distributed across speech and gesture? A cognitive modeling approach. *Cognitive Processing, Special Issue: Proceedings of KogWis* 2014.

Bergmann, K., Kopp, S., and Eyssel, F. (2010). Individualized gesturing outperforms average gesturing–evaluating gesture production in virtual humans. In Intelligent Virtual Agents (pp. 104-117). Springer Berlin Heidelberg.

Beveridge, M. E., and Pickering, M. J. (2013). Perspective taking in language: integrating the spatial and action domains. *Frontiers in human neuroscience*, *7*.

Biederman, I. (1985). Human image understanding: Recent research and a theory. Computer vision, graphics, and image processing, 32(1), 29-73.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, *94*(2), 115.

Boroditsky, L. (2000) Metaphorical structuring: Understanding time through spatial metaphors. *Cognition 75*:1–28.

Branigan, H. P., Pickering, M. J. and Cleland, A. A. (1999) Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review* 6:635–40.

Branigan, H. P., Pickering, M. J. and Cleland, A. A. (2000) Syntactic coordination in dialogue. *Cognition* 75:B13–B25.

Branigan, H. P., Pickering, M. J., Pearson, J., and McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, *42*(9), 2355-2368.

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., and King, F. J. (1975).Interservice procedures for instructional systems development. executive summary and model. Florida State Univ Tallahassee Center For Educational Technology.

Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482-1493.

Brown, L. and Brewster, S. (2003). Drawing by ear: Interpreting sonified line graphs. *In Proceedings of ICAD*. 152–156.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., and Taylor, H. A. (2009). When you and I share perspectives pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, 20(1), 27-32.

Bryant DJ (1997) Representing space in language and perception. Mind Lang 12:239–264.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cognitive Brain Research*,. 24, 355–363.

Burgstahler, S. (2007). Universal design of instruction: Definition, principles, and examples. University of Washington. Retrieved from http://www.washington.edu/doit/Brochures/Academics/instruction.html.

Burtt, H. E. (1917). Tactual illusions of movement. *Journal of Experimental Psychology*, *2*(5), 37.

Card, S. K., Moran, T. P., and Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. Communications of the ACM, 23, 396–410.

Card, S. K., Newell, A., and Moran, T. P. (1983). The psychology of human-computer interaction.

Card, S. (2008). Information Visualization. In Sears A., Jacko, J.A. (eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. pp. 509-543. Erlbaum, Mahwah NJ

Carlson, L. A. (1999). Selecting a reference frame. *Spatial cognition and computation*, *1*(4), 365-379.

Carlson, L. A. (2008). On the “whats” and “hows” of “where”: the role of salience in spatial descriptions. In Spatial Cognition VI. Learning, Reasoning, and Talking about Space (pp. 4-6). Springer Berlin Heidelberg.

Carlson, L. A., and Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition*, *106*(1), 384-407.

Carlson-Radvansky, L. A. and Jiang, Y. (1998) Inhibition accompanies reference frame selection. *Psychological Science* 9:386–91.

Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: Where is above?. *Cognition*, *46*(3), 223-244.

Carlson-Radvansky, L. A., and Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science*, 56-60.

Carlson-Radvansky, L. A., Covey, E. S., and Lattanzi, K. M. (1999). “What” effects on “where”: Functional influences on spatial relations. *Psychological Science*,*10*(6), 516-521.

Carmi, R., and Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. Vision research, 46(26), 4333-4345.

Casati, R., Varzi, A. C.. (2002). Events. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, Stanford: CSLI (www.plato.stanford.edu).

Casati, R., and Varzi, A. (2008). Event concepts. Understanding events: From perception to action, 31-53.

Casner, S. M. (1991). Task-analytic approach to the automated design of graphic presentations. ACM Transactions on Graphics (TOG), 10(2), 111-151.

Cassell, J. (1998). A framework for gesture generation and interpretation.Computer vision in human-machine interaction, 191-215.

Cassell, J. (2000). Embodied conversational agents. MIT press.

Cassell, J. T., Bickmore, T., Campbell, L., Vilhjalmsson, H. and Yan, H.. (2000). Designing embodied conversational agents. *Embodied conversational agents*, 29.

Cattaneo, Z., and Vecchi, T. (2011). Blind Vision: The Neuroscience of Visual Impairment. MIT Press.

Cheng, M. F. (1968). Tactile-kinesthetic perception of length. The American journal of psychology, 74-82.

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. Cognition,23, 43-71.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.Psychological assessment, 6(4), 284.

Clark, H. H. and Marshall, C. R. (1981) Definite reference and mutual knowledge. In: Elements of discourse understanding, ed. A. K. Joshi, I. A. Sag and B. L. Webber. Cambridge University Press.

Clark, H. H. and Murphy, G. L. (1982) Audience design in meaning and reference. In: Language and comprehension, ed. J. F. Le Ny and W. Kintsch. North- Holland. [RMK, aMJP]

Clark, H. H. and Schaefer, E. F. (1987) Concealing one's meaning from overhearers. Journal of Memory and Language 26:209–25. [aMJP]

Clark, H. H. and Schober, M. F. (1991) Asking questions and influencing answers. In: Questions about questions: Inquiries into the cognitive bases of surveys, ed. J. M. Tanur, pp. 15–48. Russell Sage Foundation. [MFS] References/Pickering and Garrod: Toward a mechanistic psychology of dialogue

Clark, H. H. and Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition* 22:1–39.

Clark, H. H. (1979) Responding to indirect speech acts. *Cognitive Psychology* 11:430–77.

Clark, H. H. (1996). *Using language.* Cambridge university press.

Clark, H. H., and Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149).

Clark, H. H., and Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, 22(2), 245-258.

Clementini, E., and Felice, P. D. (1998). Topological invariants for lines.Knowledge and Data Engineering, IEEE Transactions on, 10(1), 38-54.

Cleveland, W. S. (1985). The elements of graphing data (pp. 135-143). Monterey, CA: Wadsworth Advanced Books and Software.

Cleveland, W. S. (1990). A model for graphical perception. In proceedings of the Section on Statistical Graphcis. *American Statistical Association,* pp 1-32.

Cohen, Y. E., and Andersen, R. A. (2004). Multimodal spatial representations in the primate parietal lobe. *Crossmodal space and crossmodal attention*, 154-176.

Cohen, E. H., and Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research*, 47(22), 2825-2840.

Cohen, P. R., and Perrault, C. R. (1979). Elements of a plan based theory of speech acts. *Cognitive Science.* 3, 197-212.

Cohen, R. F., Yu, R., Meacham, A., and Skaff, J. (2005). Plumb: Displaying graphs to the blind using an active auditory interface. *In Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets'05). ACM, New York, 182–183.

Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2008). Gesturing makes learning last. Cognition, 106(2), 1047-1058.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: providing a foundation for predicting the future behavior of others. *Frontiers in human neuroscience*, *7*.

Cresswell, M. J. (1979). Interval Semantics for Some Event Expressions. In R. Bäuerle, U. Egli, and A. von Stechow (eds.), *Semantics from Different Points of View*, Berlin and Heidelberg: Springer-Verlag, pp. 90–116.

Cresswell, M. J. (1986). Why Objects Exist but Events Occur. *Studia Logica*, 45, 371–375; reprinted in Events, pp. 449–453.

Chisholm, R. M. (1970). Events and Propositions, *Noûs*, 4, 15–24; reprinted in Events, pp. 89–98.

Chisholm, R. M. (1971). States of Affairs Again. *Noûs*, 5, 179–189.

Cuijpers, R. H., Kappers, A. M., and Koenderink, J. J. (2003). The metrics of visual and haptic space based on parallelity judgements. *Journal of Mathematical Psychology*, *47*(3), 278-291.

Dale, R. (1992). Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes. MIT Press, Cambridge, MA.

Dale, R., and Reiter, E. (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19(2), 233-263.

Davidson, P. W. (1972). Haptic judgments of curvature by blind and sighted humans. *Journal of experimental psychology*, *93*(1), 43.

Davidson, P. W. (1969). The Individuation of Events, in N. Rescher (ed.), Essays in Honor of Carl G. Hempel, Dordrecht: Reidel, pp. 216–34.

Davidson, P. W., and Whitson, T. T. (1974). Haptic equivalence matching of curvature by blind and sighted humans. *Journal of experimental psychology*, *102*(4), 687.

Davis, F. D. (2006). On the relationship between HCI and technology acceptance research. Human-Computer Interaction and Management Information Systems: Applications, D. Galletta, ed, 395-401.

De Ruiter, J. (2000). The production of gesture and speech. *Language and gesture*. (pp. 284-311). Cambridge: Cambridge University Press.

Demir, S., Oliver, D., Schwartz, E., Elzer, S., Carberry, S., and McCoy, K. F (2010). Interactive SIGHT demo: Textual summaries of simple bar charts. *In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'10*). ACM, New York, 267–268.

Demir, S., Carberry, S., and McCoy, K. F. (2012). Summarizing information graphics textually. Computational Linguistics, 38(3), 527-574.

Denisova, K., Singh, M., and Kowler, E. (2006). The role of part structure in the perceptual localization of a shape.

Dillon, A. and Morris, M. (1996) User acceptance of new information technology: theories and models. In M. Williams (ed.) Annual Review of Information Science and Technology, Vol. 31, Medford NJ: Information Today, 3-32.

Ducasse, C. J., 1926, 'On the Nature and the Observability of the Causal Relation', *Journal of Philosophy*, 23, 57–68.

Duran, N. D., Dale, R., and Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*(1), 22-40.

Emiliani, P. L., and Stephanidis, C. (2005). Universal access to ambient intelligence environments: opportunities and challenges for people with disabilities. IBM Systems Journal, 44(3), 605-619.

Ellis, R., and Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. British Journal of Psychology, 91, 451–471.

Eschenbach, C., Habel, C., and Kulik, L. (1999). Representing Simple Trajectories as Oriented Curves. In FLAIRS Conference (pp. 431-436).

Eschenbach, C., Habel, C., Kulik, L., and Leßmöllmann, A. (1998, January). Shape nouns and shape concepts: A geometry for 'corner'. In *Spatial cognition* (pp. 176-202). Springer Berlin Heidelberg.

Eschenbach, C., Tschander, L., Habel, C., and Kulik, L. (2000). Lexical specifications of paths. In C. Freksa, W. Brauer, C. Habel and K. F. Wender (Eds.), Spatial Cognition II (pp. 127-144). Berlin: Springer.

Faineteau, H., Gentaz, E., and Viviani, P. (2003). The kinaesthetic perception of Euclidean distance: a study of the detour effect. *Experimental Brain Research*, *152*(2), 166-172.

Feldman, J. (2001). Bayesian contour integration. Perception and Psychophysics,63, 1171-1182.

Ferreira, V. S., Slevc, L. R., and Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions?. Cognition, 96(3), 263-284.

Ferres, L., Lindgaard, G., and Sumegi, L. 2010. Evaluating a tool for improving accessibility to charts and graphs. In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility(ASSETS'10). ACM, New York, 83–90.

Ferres, L., Lindgaard, G., Sumegi, L., and Tsuji, B. (2013). Evaluating a tool for improving accessibility to charts and graphs. ACM Transactions on Computer-Human Interaction (TOCHI), 20(5), 28.

Feyereisen, P., and Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of nonverbal behavior*, 23(2), 153-171.

Fiehler, K., Burke, M., Engel, A., Bien, S., and Rösler, F. (2008). Kinesthetic working memory and action control within the dorsal stream. *Cerebral Cortex*,*18*(2), 243-253.

Fischer, G. (2001). User modeling in human–computer interaction. User modeling and user-adapted interaction, 11(1-2), 65-86.

Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs?. Applied cognitive psychology, 14(2), 151-162.

Fischer, M. H., Dewulf, N., and Hill, R. L. (2005). Designing bar graphs: Orientation matters. Applied Cognitive Psychology, 19(7), 953-962.

Foster, M.E., Bard, E.G., Hill, R.L., Guhe, M., Oberlander, J., and Knoll, A. (2008). The Roles Of Haptic-Ostensive Referring Expressions in Cooperative, Task-based Human-Robot Dialogue. *In Proceedings of ohe 3rd ACM/IEEE International Conference on Human-Robot Interaction*, pp. 295-302. Amsterdam, March 12-15, 2008.

Freedman, E. G., and Shah, P. (2002). Toward a Model of Knowledge-Based Graph Comprehension. Paper presented at Diagrammatic Representation and Inference, Second International Conference, Diagrams 2002, Callaway Gardens, GA, USA.

Friederici, A. D., and Levelt, W. J. (1990). Spatial reference in weightlessness: perceptual factors and mental representations. *Perception and Psychophysics*,*47*(3), 253-266.

Frings, C., and Spence, C. (2013). Gestalt grouping effects on tactile information processing: when touching hands override spatial proximity. Attention, Perception, and Psychophysics, 75(3), 468-480.

Fukumura, K., van Gompel, R. P., and Pickering, M. J. (2010). The use of visual context during the production of referring expressions. The Quarterly Journal of Experimental Psychology, 63(9), 1700-1715.

Galati, A., Michael, C., Mello, C., Greenauer, N. M., and Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, *68*(2), 140-159.

Gallace, A., and Spence, C. (2011). To what extent do Gestalt grouping principles influence tactile perception?. Psychological bulletin, 137(4), 538.

Garbarini, F., and Adenzato, M. (2004). At the root of embodied cognition: Cognitive science meets neurophysiology. *Brain and cognition*, 56(1), 100-106.

Garrod, S. and Clark, A. (1993) The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes* 8:101–26.

Garrod, S., and Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., and Pickering, M. J. (2004). Why is conversation so easy?. *Trends in cognitive sciences*, *8*(1), 8-11.

Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292-304.

Gatt, A., and Belz, A. (2010). Introducing shared tasks to NLG: The TUNA shared task evaluation challenges. In E. Krahmer and M. Theune (Eds), *Empirical methods in natural language generation* (pp. 264-293). Springer Berlin Heidelberg.

Gatt, A., Krahmer, E., van Deemter, K., and van Gompel, R. P. (2014). Models and empirical data for the production of referring expressions. *Language, Cognition and Neuroscience*, 29, 899-911.

Gattis, M. (2005). Inferencing from spatial information. *Spatial Cognition and Computation*, *5*(2-3), 119-137.

Gentaz, E., Baud-Bovy, G., and Luyat, M. (2008). The haptic perception of spatial orientations. *Experimental brain research*, 187(3), 331-348.

Gentaz, E., and Hatwell, Y. (1995). The haptic oblique effect in children's and adults' perception of orientation. *Perception*, London, 24, 631-631.

Gentner, D. and Markman, A. B. (1997) Structural alignment in analogy and similarity. *American Psychologist* 52(1):45–56.

Gentner, D., and Stevens, A. L. (Ed.) (1983). Mental models. Hillsdale Nj: Lawrance Erlbaum Association.

Gentner, D., and Stevens, A. L. (2014). Mental models. Psychology Press.

Geomagic Touch (formerly Sensable Phantom Omni) User Manual. Retrieved from http://www.dentsable.com/haptic-phantom-omni.htm, retrieved on August 12, 2015.

Gibbs, R. W., JR. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, 20(4), 431-444.

Gibbs Jr, R. W. (2005). Embodiment and cognitive science. Cambridge University Press.

Gibson, J. J. (1962). Observations on active touch. *Psychological review*, *69*(6), 477.Giudice, N. A., Palani, H., Brenner, E., Kramer, K.M.: Learning Non-Visual Graphic Information using a Touch-Based Vibro-Audio Interface. In Proceedings 14th international ACM SIGACCESS conference on Computers and accessibility, pp. 103-110. ACM (2012)

Giles, H., Coupland, N. and Coupland, J. (1991) Accommodation theory: Communication, context, and consequence. *In: Contexts of accommodation: Developments in applied sociolinguistics, ed. H. Giles, J. Coupland, and N. Coupland*, pp. 1–68. Cambridge University Press.

Gillan, D. J., and Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. Human Factors: The Journal of the Human Factors and Ergonomics Society, 36(3), 419-440.

Giudice, N. A., Betty, M. R., and Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(3), 621.

Giudice, N. A., Palani, H. P., Brenner, E., and Kramer, K. M. (2012). Learning non-visual graphical information using a touch-based vibro-audio interface. In Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility (pp. 103-110). ACM.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and brain sciences*, 20(01), 41-50.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin and review*. 9, 558– 565.

Goodman, N. (1951) The Structure of Appearance, Cambridge (MA): Harvard University Press.

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. Child Development, 231-239.

Goldin-Meadow, S. (2005). Hearing gesture: How our hands help us think. Harvard University Press.Goldin-Meadow, S., and Sandhofer, C.M. (1999). Gesture conveys substantive information about a child's thoughts to ordinary listeners. Developmental Science, 2, 67–74.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. Language and cognition, 2(01), 1-19.

Goldin-Meadow, S., Alibali, M.W., and Church, R.B. (1993). Transitions in concept acquisition: Using the hand to read the mind. Psychological Review, 100, 279–297.

Goldin-Meadow, S., and Beilock, S. L. (2010). Action's influence on thought: The case of gesture. Perspectives on Psychological Science, 5(6), 664-674.

Goldin-Meadow, S., and Singer, M.A. (2003). From children's hands to adults' ears: Gesture's role in teaching and learning. Developmental Psychology, 39, 509–520.

Goldin-Meadow, S., Cook, S.W., and Mitchell, Z.A. (2009). Gesturing gives children new ideas about math. Psychological Science, 20, 267–272.

Goldin-Meadow, S., Kim, S., and Singer, M. (1999). What the teacher's hands tell the student's mind about math. Journal of Educational Psychology, 91, 720–730.

Goldin-Meadow, S., Nusbaum, H., Kelly, S., and Wagner, S. (2001). Explaining math: Gesturing lightens the load. Psychological Science, 12, 516–522.

Goldin-Meadow, S., Wein, D., and Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. Cognition and Instruction, 9, 201–219.

Goldman, A. I. (2007). A Program for "Naturalizing" Metaphysics, with Application to the Ontology of Events. *The Monist* 90: 457–479.

Goldstein, E. B. (1989). Sensation and Perception, Brooks/Cole Publishing Company.

Golledge, R. G. (1993). Geography and the disabled: a survey with special reference to vision impaired and blind populations. Transactions of the Institute of British Geographers, 63-85.

Goodman, N. (1976), Languages of Art, 2nd ed., Indianapolis: Hackett.

Goodwin, G. P., and Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological review*, 112(2), 468.

Gordon, I. A. and Morison, V. (1982) The haptic perception of curvature. *Percept. Psychophys.* 31, 446–450.

Goudbeek, M., and Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. Topics in cognitive science, 4(2), 269-289.

Grice, H.P. (1975). Logic and conversation. *In P.Cole and J. Morgan (Eds.), Syntax and Semantics*: Vol 3, Speech acts (pp.43-58). New York: Academic.

Grunwald, M., Muniyandi, M., Kim, H., Kim, J., Krause, F., Mueller, S., Srinivasan, M. A.: Human haptic perception is interrupted by explorative stops of milliseconds. *Frontiers in psychology*, 5. (2014).

Guhe, M., Habel, C., and Tschander, L. (2003). Incremental production of preverbal messages with inC. In F. Detje, D. Dörner, and H. Schaub (Eds.), Proceedings of the 5th International Conference on Cognitive Modeling (ICCM-5), (pp 123–128). Bamberg University Press.

Giuliani, M., Foster, M. E., Isard, A., Matheson, C., Oberlander, J., and Knoll, A. (2010). Situated reference in a hybrid human-robot interaction system. In Proceedings of the 6th International Natural Language Generation Conference(pp. 67-75). Association for Computational Linguistics.

Gurr, C. A. (1999). Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. Journal of Visual Languages and Computing, 10(4), 317-342.Habel, C. (2005). Verbs and directions. In L. A. Carlson and E. v. d. Zee (Eds.), Functional features in language and space (pp. 93–112). Oxford: Oxford University Press.

Gustafson, K. L., and Branch, R. M. (1997). Survey of instructional development models (3rd Ed.). Syracuse, NY: ERIC Clearinghouse on Information Resources.

Habel, C. (2003). Representational commitment in maps (pp. 69-93). London: Taylor and Francis.

Habel, C., and Acartürk, C. (2007). On reciprocal improvement in multimodal generation: Co-reference by text and information graphics. In I. van der Sluis, M. Theune, E. Reiter and E. Krahmer (Eds.), *Proceedings of MOG 2007: The workshop on multimodal output generation* (pp. 69-80). University of Aberdeen, UK.

Habel, C., and Acartürk, C. (2009). Eye-tracking evidence for multimodal language-graphics comprehension: The role of integrated conceptual representations. In C. Navarretta, P. Paggio, J. Allwood, E. Alsén and Y. Katagiri (Eds.), *Proceedings NODALIDA 2009 Workshop. Multimodal Communication* (Vol. 6, pp. 9-14). Odense.

Habel, C. and Acartürk, C. (2011). Causal inference in graph-text constellations: Designing verbally annotated graphs. *Tsinghua Science and Technology*, 16(1), 7-12

Habel, C., and Acartürk, C. (2012). Towards comprehension of haptic multiple-line graphs. In *Proceedings of the Diagrams Workshop on Accessible Graphics: Graphics for Vision Impaired People. Kent, UK.*

Habel, C., Alaçam, Ö., and Acartürk, C. (2013). Verbally assisted comprehension of haptic line-graphs: referring expressions in a collaborative activity. In *Proceedings of the CogSci 2013 Workshop on Production of Referring Expressions*, Berlin.

Habel, C., Alaçam, Ö., Acartürk, C. (under revision). Geometric Representation of Line Shape for Multimodal Comprehension of Statistical Graphs.

Habel, C., and Eschenbach, C. (1997, January). Abstract structures in spatial cognition. In *Foundations of computer science* (pp. 369-378). Springer Berlin Heidelberg.

Habel, C., and Tappe, H. (1999). Processes of segmentation and linearization in describing events. In R. Klabude and C. von Stutterheim (Eds.), Processes in language production (pp. 117–153). Opladen: Westdeutscher Verlag.

Habel, C., Pribbenow, S., and Simmons, G. (1995). Partonomies and depictions: A hybrid approach. In J. Glasgow, H. Narayanan, and B. Chandrasekaran (Eds.), Diagrammatic reasoning: Cognitive and computational perspectives (pp. 627–653). Cambridge, MA: MIT Press.

Hanson, C., and Hirst, W. (1989). On the representation of events: a study of orientation, recall, and recognition. *Journal of Experimental Psychology: General*, 118(2), 136.

Hard, B. M., Recchia, G., and Tversky, B. (2011). The shape of action. *Journal of experimental psychology: General*, 140(4), 586.

Hard, B. M., Tversky, B., and Lang, D. (2006). Making sense of abstract events: Building event schemas. *Memory and Cognition*, 34(6), 1221–1235.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Hegarty, M. (2005). Multimedia learning about physical systems. *The Cambridge handbook of multimedia learning*, 447-65.

Hegarty, M. (2011). The Cognitive Science of Visual-spatial Displays: Implications for Design. *Topics in Cognitive Science*, 3, 446–474.

Hegarty, M., Carpenter, P. A., Just, M. A. (1991). Diagrams in the comprehension of scientific texts. *Handbook of reading research.* 641-668.

Heller, M. A. (1989). Tactile memory in sighted and blind observers: the influence of orientation and rate of presentation. *Perception*, 18(1), 121-133.

Heller, M. A., and Myers, D. S. (1983). Active and passive tactual recognition of form. *The Journal of general psychology*, 108(2), 225-229.

Heller, M. A., Calcaterra, J. A., Burson, L. L., and Tyler, L. A. (1996). Tactual picture identification by blind and sighted people: Effects of providing categorical information. *Perception and Psychophysics*, *58*(2), 310-323.

Hilbert, D.: Foundations of Geometry. Open Court, La Salle, IL (1902 / rev. edition 1971)

Hobbs, J., and Pustejovsky, J. (2003). Annotating and reasoning about time and events. *In Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning* (Vol. 3).

Hoedemaker, R. S., and Gordon, P. C. (2014). Embodied language comprehension: Encoding-based and goal-driven processes. *Journal of Experimental Psychology*: General, 143(2), 914.

Hoffman, D. D., and Richards, W. A. (1982, August). Representing Smooth Plane Curves for Recognition: Implications for Figure-Ground Reversal. In *AAAI* (pp. 5-8).

Hoffman, D. D., and Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*(1), 65-96.

Hollands, J. G., and Spence, I. (1992). Judgments of change and proportion in graphical perception. Human Factors: The Journal of the Human Factors and Ergonomics Society, 34(3), 313-334.

Hollands, J. G., and Spence, I. (1998). Judging proportion with graphs: The summation model. Applied Cognitive Psychology, 12(2), 173-190.

Hollands J. G. and Spence I. (2001), The discrimination of graphical elements, Applied Cognitive Psychology, 15, 413-431.

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. Psychonomic bulletin and review, 15(3), 495-514.

Hostetter, A. B., Alibali, M. W., and Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands.Language and Cognitive Processes, 22, 313-336.

Hostetter, A. B., and Sullivan, E. L. (2011). Gesture production during spatial tasks: Its not all about difficulty. In Proceedings of the 33rd Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Huang, Y. Y., Moll, J., Sallnäs, E. L., and Sundblad, Y. (2012). Auditory Feedback in Haptic Collaborative Interfaces. *International Journal of Human-Computer Studies*, 70(4), 257-270.

Hunter, I. M. (1954). Tactile-kinesthetic perception of straightness in blind and sighted humans. *Quarterly Journal of Experimental Psychology*, *6*(4), 149-154.

Hutchins, E. L. And Palen, L. (1997). Constructing meaning from space, gesture, and speech. In Tools, and Reasoning: Essays in Situated Cognition, L. B. Resneck, R. Saljo, C. Pontecorvo, and B. Burge, Eds. Springer-Verlag, Vienna, Austria.

IBM Corporation (1992). Object-Oriented Interface Design: IBM Common User Access Guidelines. Indianapolis IN: QUE.

ISO/CD 9241-11. (2015) Ergonomics of human-system interaction -- Part 11: Usability: Definitions and concepts

Jackendoff, R. (2002). Foundations of language: Brain, meaning, grammar, evolution. Oxford, UK: Oxford University Press.

Jackendoff, R. (2012). Language as a source of evidence for theories of spatial representation. Perception-London, 41(9), 1128.

Johannsen, K., and De Ruiter, J. P. (2013). Reference frame selection in dialog: priming or preference?. *Frontiers in human neuroscience*, *7*.

John, B.E., Prevas, K., Salvucci, D. D., and Koedinger, K. (2004). Predictive Human Performance Modeling Made Easy. Proceedings of CHI 2004 (pp. 455–462). Vienna, Austria: ACM Press

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (No. 6). Harvard University Press.

Johnson-Laird, P. N. (1989). Mental models. In M.I. Posner (Ed.), Foundations

Johnson-Laird, P. N. (2001). Mental models and deduction. Trends in cognitive sciences, 5(10), 434-442.

Johnson-Laird, P. N. (2006). How we reason. Oxford University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences,* 107(43), 18243-18250.

Johnson-Laird, P. N. (2013). Human and machine thinking. *Psychology Press.*

Johnson, M. (1987). The body in the mind: The bodily basis of meaning, imagination, and reason. Chicago: University of Chicago Press.

Johnson-Laird,P.N.,and Stevenson,R. (1970). Memory for syntax. *Nature* 227, 412.

Johnson, C. J., Paivio, A. U., and Clark, J. M. (1989). Spatial and verbal abilities in children's crossmodal recognition: A dual coding approach. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *43*(3), 397.

Kaas, A. L., Stoeckel, M. C., and Goebel, R. (2008). The neural bases of haptic working memory. *In Human haptic perception: Basics and applications* (pp. 113-129). Birkhäuser Basel.

Kalia, A. A., and Sinha, P. (2012). Tactile picture recognition: Errors are in shape acquisition or object matching?. *Seeing and perceiving*, *25*(3-4), 287-302.

Kappers, A. M. (2011). Human perception of shape from touch. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 366(1581), 3106-3114.

Kappers, A. M., and Schakel, W. B. (2011). Comparison of the haptic and visual deviations in a parallelity task. *Experimental brain research*, *208*(3), 467-473.

Kendon, Adam (2008). Some reflections on 'gesture' and 'sign: Gesture, 8 (3),348-336Kendon, Adam (2009). Language's matrix. *Gesture*, 9 (3), 355–372

Kerzel, M. (2015) Event Recognition during the Exploration of Line-Based Graphics in Virtual Haptic Environments (in German). (Doctoral dissertation). University of Hamburg. Retrieved from http://ediss.sub.uni-hamburg.de/volltexte/2015/7438/pdf/Dissertation.pdf

Kerzel, M., Alaçam, Ö., Habel, C., Acartürk, C (2014). Producing Verbal Descriptions for Haptic Line-Graph Explorations [Poster abstract]. In Rieser, V., Muller, P. (Eds.) *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue* (p. 205-207). SemDial 2014 – DialWatt

Kerzel, M. and Habel, C. (2013). Event Recognition During Exploration of Hapic Virtual Environment Line-based Graphics. In T. Tenbrink, J. Stell, A. Galton and Z. Wood (eds.) Spatial Information Theory, 11th International Conference, COSIT 2013. (pp. 109–128). Berlin: Springer-Verlag.

Kessler, K., and Rutherford, H. (2009). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions.*Embodied and grounded cognition*, 108.

Kieras, D. E. (1997). Task Analysis and the Design of Functionality. The computer science and engineering handbook, 23, 1401-1423.Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), Language and gesture (pp. 162–185). New York: Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. Psychological review, 95(2), 163.

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. Journal of Memory and language, 48(1), 16-32.

Kita, S., Van Gijn, I., and Van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. *In Gesture and sign language in human-computer interaction* (pp. 23-35). Springer Berlin Heidelberg.

Klatzky, R. L., Lederman, S. J., and Metzger, V. A. (1985). Identifying objects by touch: An "expert system". Perception and psychophysics, 37(4), 299-302.

Klatzky, R. L., and Lederman, S. J. (2007). Object recognition by touch. *Blindness and brain plasticity in navigation and object perception*, 185-207.

Klatzky, R. L., Lederman, S. J., and Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of experimental psychology: general*, *116*(4), 356.

Knauff, M. (2013). Space to reason: A spatial theory of human thought. MIT Press.

Koolen, R., Gatt, A., Goudbeek, M., and Krahmer, E. (2009). Need I say more? On factors causing referential overspecification. In Proceedings of the pre-cogsci 2009 workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.

Koolen, R., Gatt, A., Goudbeek, M., and Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43, 3231-3250.

Koolen, R., Goudbeek, M., and Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive science*, 37(2), 395-411.

Kopp, S., Jung, B., Leßmann, N., and Wachsmuth, I. (2003). Max - A Multimodal Assistant in Virtual Reality Construction. KI, 17(4), 11.

Kopp, S., and Bergmann, K. (2012). Individualized gesture production in embodied conversational agents. In Human-Computer Interaction: The Agency Perspective (pp. 287-301). Springer Berlin Heidelberg.

Kopp, S., Bergmann, K., and Kahl, S. (2013). A spreading activation model of the semantic coordination of speech and gesture. In: Proceedings of the 35th Annual Meeting of the Cognitive Science Society (COGSCI 2013).

Kosslyn, S. M. (1987). Understanding charts and graphs (No. TR-8). Harvard Univ Cambridge Ma Dept Of Psychology.

Kosslyn, S. M., (1989), Understanding Charts and Graphs, Applied Cognitive Psychology, 3, 185-226.

Kosslyn, S. (1993). Elements of graph design. NewYork: W.H. Freeman and Company.

Kosslyn, S. M. (1994). Image and brain. Cambridge, MA:MIT Press.

Kosslyn, S. M. (2006). Graph design for the eye and mind. Oxford University Press.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., and Gabrieli, J. D. (1989). Evidence for two types of spatial representations: hemispheric specialization for categorical and coordinate relations. Journal of experimental psychology: human perception and performance, 15(4), 723.

Krahmer, E., and Theune, M. (2002). Efficient Context-sensitive Generation of Referring Expressions. In: K. van Deemter and R. Kibble (Eds.) Information Sharing: Reference and Presupposition in Language Generation and Interpretation. (pp. 223-264). CSLI Publications, Stanford.

Krauss, R. M., Chen, Y., and Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), Language and gesture (pp. 261-283). Cambridge: Cambridge University Press.

Kurze, M. (1995). Giving blind people access to graphics (example: Business graphics). In Proceedings of the Software-Ergonomics'95 Workshop.

Lacey, S., and Campbell, C. (2006). Mental representation in visual/haptic crossmodal memory: evidence from interference effects. *The Quarterly Journal of Experimental Psychology*, 59(2), 361-376.

Lacey, S., Campbell, C., and Sathian, K. (2007). Vision and touch: multiple or multisensory representations of objects?. Perception-London, 36(10), 1513-1522.

Lakusta, L., and Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive science*, 36(3), 517-544.

Lakoff, G., and Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Landau, B., and Jackendoff, R. (1993). Whence and whither in spatial language and spatial cognition?. *Behavioral and brain sciences*, 16(02), 255-265.

Larkin, J. H., and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive science, 11(1), 65-100.

Lassiter, G. D., and Slaw, R. D. (1991). The unitization and memory of events. *Journal of Experimental Psychology: General*. 120(1), 80–82.

Lederman, S. J., and Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive psychology*, *19*(3), 342-368.

Lederman, S. J., and Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, and Psychophysics*, *71*(7), 1439-1459.

Lee, E. J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior*, 26(4), 665-672.

Lechelt, E. C., Eliuk, J., and Tanne, G. (1976). Perceptual orientational asymmetries: A comparison of visual and haptic space. *Perception and Psychophysics*, 20(6), 463-469.

Lechelt, E. C., and Verenka, A. (1980). Spatial anisotropy in intramodal and cross-modal judgments of stimulus orientation: The stability of the oblique effect. *Perception*, 9(5), 581-589.

Levelt, W. (1981). JM (1989). Speaking. From intention to articulation. MA: The MIT Press, Cambridge.

Levinson, S. C.(2003). Space and Language in Cognition: Explorations in Cognitive Diversity. Cambridge: Cambridge University Press.

Levy, E., Zacks, J., Tversky, B., and Schiano, D. (1996). Gratuitous graphics? Putting preferences in perspective. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 42-49). ACM.

Lewis, C., Polson, P. G., Wharton, C., and Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 235-242). ACM.

Lewis, D. K. (1969) Convention: A philosophical study. Basil Blackwell/Harvard University Press.

Lidwell, W., Holden, K., and Butler, J. (2010). Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design. Rockport Pub.

Linnell, P. (1998) Approaching dialogue: Talk, interaction, and contexts in a dialogical perspective. *Studies in language and society*, *3*. Benjamins.

Logie, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, 36(3), 189-208.

Lohmann, K. (2012) Verbal Assistance with Virtual Tactile Maps : a Multi-Modal Interface for the Non-Visual Acquisition of Spatial Knowledge. (Doctoral dissertation). University of Hamburg. Retrieved from http://ediss.sub.uni-hamburg.de/volltexte/2013/6240/pdf/Dissertation.pdf

Lohmann, K., and Habel, C. (2012). Extended verbal assistance facilitates knowledge acquisition of virtual tactile maps. *In Spatial Cognition VIII* (pp. 299-318). Springer Berlin Heidelberg.

Lohmann, K., Eschenbach, C., and Habel, C. (2011). Linking spatial haptic perception to linguistic representations: assisting utterances for tactile-map explorations. *In Spatial information theory* (pp. 328-349). Springer Berlin Heidelberg.

Lohmann, K., Kerzel, M., and Habel, C. (2010). Generating Verbal Assistance for Tactile-Map Explorations1. *In MOG 2010 3rd Workshop on Multimodal Output Generation* (p. 27).

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8(4), 353-388.

Loomis, J. M., and Klatzky, R. L. (2008). Functional equivalence of spatial representations from vision, touch, and hearing: Relevance for sensory substitution. *Blindness and brain plasticity in navigation and object perception*, 155-184.

Loomis, J. M., and Lederman, S. J. (1986). Tactual perception. *Handbook of perception and human performances*, 2, 2.

Loomis, J. M., Klatzky, R. L., and Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception, 20*(2), 167-177.

Loomis, J. M., Klatzky, R. L., Avraamides, M., Lippa, Y., and Golledge, R. G. (2007). Functional equivalence of spatial images produced by perception and spatial language. In *Spatial processing in navigation, imagery and perception* (pp. 29-48). Springer US.

Louw, S., Kappers, A. M. L. and Koenderink, J. J. (2000) Haptic detection thresholds of Gaussian profiles over the whole range of spatial scales. *Experimental Brain Research.* 132, 369–374.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics* (Tog), 5(2), 110-141.

Magee, L. E., and Kennedy, J. M. (1980). Exploring pictures tactually. *Nature* 283: 287-288.

Marchetti, F. M., and Lederman, S. J. (1983). The haptic radial-tangential effect: Two tests of Wong's "moments-of-inertia" hypothesis. *Bulletin of the Psychonomic Society*, *21*(1), 43-46.

Markman, A. B. and Gentner, D. (1993) Structural alignment during similarity comparisons. Cognitive Psychology 25:431–67.

Marr, D. (1982). Vision: A computational approach. San Fransico, Freeman.

Massad, C. M., Michael, H., and Newtson, D. (1979). Selective perception of events. Journal of Experimental Social Psychology, 15(6), 513–532.

Mazzarella,E., Hamilton, A., Trojano, L., Mastromauro, B., and Conson, M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *The Quarterly Journal of Experimental Psychology* . 65, 2447– 2460.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

McNeill, D. (2000). Analogic/analytic representations and cross-linguistic differences in thinking for speaking. Cognitive Linguistics, 11(1/2), 43-60.

McNeill, D. (2008). Gesture and thought. University of Chicago Press.

Meathrel, R., and Galton, A. (2001, August). A hierarchy of boundary-based shape descriptors. In *International Joint Conference On Artificial Intelligence* (Vol. 17, No. 1, Pp. 1359-1364). Lawrence Erlbaum Associates Ltd.

Melinger, A., and Kita, S. (2007). Conceptualisation load triggers gesture production. Language and Cognitive Processes, 22, 473-500.

Melinger, A., and Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. Gesture, 4, 119-141.

Mennie, N., Hayhoe, M., and Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*,179, 427–442.

Meyer, J., Shinar, D., and Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 268-286.

Millar, S. (1999). Memory in touch. Psicothema, 11(4), 747-767.

Millar, S. (1978). Aspects of information from touch and movement. In G. Gordon (ed.) Active Touch. London: Pergamon Press.

Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and perception*. Belknap Press.

Moll, J., and Sallnäs, E. L. (2009). Communicative Functions of Haptic Feedback. In: M. E. Altinsoy, U. Jekosch, and S. A. Brewster (Eds.), Haptic and Audio Interaction Design. (pp. 1-10). Springer, Berlin Heidelberg.

Moll, J., and Pysander, E. L. S. (2013). A haptic tool for group work on geometrical concepts engaging blind and sighted pupils. ACM Transactions on Accessible Computing (TACCESS), 4(4), 14.Nielsen, Jakob (1993). Usability Engineering. San Diego CA: Academic Press.

Morris, M. W., and Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory and Cognition*, 18(4), 407-418.

Nesbitt, K. (2005). "Structured Guidelines to Support the Design of Haptic Displays." GOTHI'05 - Guidance on Tactile and Haptic Interactions, October 24-26, 2005 Saskatoon, Saskatchewan, CANADA.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 2 8 , 28-38.

Newtson, D., and Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436-450.

Newtson, D., Engquist, G. A., and Bois, J. (1977). The objective basis of behavior units. Journal of Personality and social psychology, 35(12), 847.

Nielsen, J. (1994). *Usability engineering*. Elsevier.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner and A. L. Stevens (Eds.), Mental models (pp. 7–14). Hillsdale, NJ: Erlbaum. 7-14.

Norman, Donald (1988). The Design of Everyday Things. New York: Doubleday/Currency.

O'Malley, M. K., and Gupta, A. (2008). Haptic interfaces. HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces", Edited by Kortum, P., Elsevier Inc., USA, 1. (pp. 25-73)

Oliveira, F. C., Quek, F., Cowan, H., and Fang, B. (2012). The Haptic Deictic System—HDS: Bringing blind students to mainstream classrooms. Haptics, IEEE Transactions on, 5(2), 172-183.

Özçalışkan, S., and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. Cognition, 96.

Özyürek, A. (2010). The role of iconic gestures in production and comprehension of language: evidence from brain and behavior. In Gesture in Embodied Communication and Human-Computer Interaction (pp. 1-10). Springer Berlin Heidelberg.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130296.

Özyürek, A., Kita, S., Allen, S. E., Furman, R., and Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. Gesture, 5(1-2), 219-240.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In Rosch, E., and Lloyd, B. B. (eds.), Cognition and Categorization, Erlbaum, Hillsdale, NJ, pp. 259–303.

Papeo, L., Corradi- Dell'Acqua, C., and Rumiati, R.I. (2011). "She" is not like "I": the tie between language and action is in our imagination. *Journal of cognitive neuroscience*,. 23, 3939–3948.

Pearce, E., Smith, C. (1998). *Fodor's World Weather Guide*. Random House: New York.

Pechmann, T. (1989). Incremental speech production and referential overspecification. Linguistics, 27(1), 89-110.Ping, R and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. Cognitive Science. Volume34, Number 4, 602-619.

Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In Proceedings of the 13th annual ACM international conference on Multimedia (pp. 683-689). ACM.

Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427.

Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.

Pickering, M. J., and Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228.

Pickering, M.J., McLean, J.F., and Gambi, C. (2012).Do addressees adopt the perspective of the speaker? ActaPsychol. (Amst) 141, 261–269.

Pinker, S. (1990). A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73-126.

Polyn, S. M., and Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences,* 12, 24–30.

Pont, S. C., Kappers, A. M. L. and Koenderink, J. J., 1997 Haptic curvature discrimination at several regions of the hand. *Percept. Psychophys.*

Pont, S. C., Kappers, A. M. L. and Koenderink, J. J. 1999 Similar mechanisms underlie curvature comparison by static and by dynamic touch. *Percept. Psychophys.* 61, 874–894.

Prather S C, Votaw J R, Sathian K, 2004 ``Task-specific recruitment of dorsal and ventral visual areas during tactile perception'' Neuropsychologia 42 1079 -1087

PRBO. Waterbird Census at Bolinas Lagoon, Marin County, CA. Public report by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science. (2012) http://www.prbo.org/cms/366, retrieved on January 29, 2012.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). Human-computer interaction. Addison-Wesley Longman Ltd..

Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition *Journal of cognitive neuroscience.* 17, 884–892.

Pulvermüller,F. (2005).Brain mechanisms linking language and action. *Nature Reviews Neuroscience*,. 6, 576–582.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, 41(1), 47-81.

Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., and Radev, D. R. (2003). TimeML: Robust specification of event and temporal expressions in text. New directions in question answering, 3, 28-34.

Pustejovsky, J., Lee, K., Bunt, H., and Romary, L. (2010). ISO-TimeML: An International Standard for Semantic Annotation. In LREC.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X. F., Kirbas, C., McCullough, K. And Ansari, R. (2002). Multimodal human discourse: gesture and speech. ACM Transactions on Computer-Human Interaction (TOCHI), 9(3), 171-193.

Ramloll, R., Yu, W., Brewster, S., Riedel, B., Burton, M., and Dimigen, G. (2000, November). Constructing sonified haptic line graphs for the blind student: first steps. In Proceedings of the fourth international ACM conference on Assistive technologies (pp. 17-25). ACM.

Ramsey, F. P. (1927). Facts and Propositions. *Proceedings of the Aristotelian Society*, Suppl. Vol. 7, 153–170.

Ratwani, R. M., and Trafton, J. G. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. Psychonomic bulletin and review, 15(4), 757-762.

Rauh, R., Kulik, L.: The influence of linear shapes on solving interval-based configuration problems. In In C. Freksa, W. Brauer, C. Habel, K. Wender (eds.), Spatial Cognition II LNCS , vol. 1404, pp. 239-252. Springer, Heidelberg (2000)

Reigeluth, C. M. (1983). Meaningfulness and instruction: Relating what is being learned to what a student knows. Instructional Science, 12(3), 197-218.

Reiter, E. (1994). Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible?. In Proceedings of the Seventh International Workshop on Natural Language Generation (pp. 163-170). Association for Computational Linguistics.

Reiter, E., Dale, R., and Feng, Z. (2000). Building natural language generation systems (Vol. 33). Cambridge: Cambridge university press.

Riedel, B. (2001). Perception of gradient in haptic graphs: A comparison of virtual and hysical stimuli. In Conference Proceedings to Eurohaptics.

Renshaw, J. A., Finlay, J. E., Tyfa, D., and Ward, R. D. (2004). Understanding visual influence in graph design through temporal and spatial eye movement characteristics. *Interacting with computers*, 16(3), 557-578.

Reynolds, J. R., Zacks, J. M., and Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, 31(4), 613-643.

Richards, W., and Hoffman, D. D. (1985). Codon constraints on closed 2D shapes. *Computer Vision, Graphics, and Image Processing*, *31*(3), 265-281.

Richardson, D. C., Dale, R., and Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological science*, *18*(5), 407-413.

Richardson, D. C., Dale, R., and Spivey, M. J. (2007). Eye movements in language and cognition: A brief introduction. In M. Gonzalez- Marquez, S. Coulson, I. Mittelberg, and M. J. Spivey (Eds.), *Methods in cognitive linguistics.* Amsterdam: John Benjamins.

Rogers, Y., Rutherford, A., and Bibby, P. A. (Eds.). (1992). Models in the mind. Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E. (1999). Principles of categorization. *Concepts: core readings*, 189-206.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. Review of Educational Research, 71(3), 365-392.

Rubin, J. M., and Richards, W. A. (1985). Boundaries of visual motion.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. (1986). Sequential thought processes in PDP models. V, 2, 3-57.

Ryle, G. (1949) The Concept of Mind, London: Hutchinson.

Sachs, J.(1967). Recogition memory for syntactic and semantic aspects of connected discourse. Percept. Psychophys. 2, 437–442.

Sanders, A. F., and Kappers, A. M. (2008). Curvature affects haptic length perception. *Acta psychologica*, *129*(3), 340-351.

Seels, B and Richie, R. (1994) Instructional technology: The definitions and domains of the field. Washington: AECT.

Schwan, S., and Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, 18(1), 37-55.

Schilder, F., and Habel, C. (2001). From temporal expressions to temporal information: Semantic tagging of news messages. *In Proceedings of the workshop on temporal* *and spatial information processing*-Volume 13 (p. 9). Association for Computational Linguistics.

Schilder, F., Katz, G., and Pustejovsky, J. (2007). Annotating, extracting and reasoning about time and events. *In Annotating, Extracting and Reasoning about Time and Events* (pp. 1-6). Springer Berlin Heidelberg.

Schmidt-Weigand, F. (2006). Dynamic visualizations in multimedia learning: The influence of verbal explanations on visual attention, cognitive load and learning outcome (*Doctoral dissertation*, Universitäts bibliothek Giessen).

Shneiderman, B. (2000). Universal usability. *Communications of the ACM* ,43(5), 84-91.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational psychology review*, 14(1), 101-120.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*,47(1), 1-24.

Schober, M. F. (1995) Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes* 20(2):219–47.

Schober, M.,and Brennan,S.E. (2003). Processes of interactive spoken dis- course: the role of the partner. In *Handbook of Discourse Processes*, eds. Graesser, Gernsbacher and Goldman (Mahwah, NJ: Lawrence Erlbaum),123–164.

Sebanz, N, Bekkering, H., and Knoblich, G. (2006). Joint Action: Bodies and Minds Moving Together. *Trends in Cognitive Sciences*, 10, 70-76.

Shah, P. (2002). Graph comprehension: The role of format, content, and individual difference. In M. Anderson, B. M. M. and P. Olivier (Eds.), *Diagrammatic Representation and Reasoning*. London, New York: Springer Verlag.

Shah, P. and Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of experimental Psychology: General*, 124, 43-61.

Shah, P., and Freedman, E. G. (2011). Bar and Line Graph Comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*,3(3), 560-578.

Shah, P., and Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47-69.

Shah, P., and Shellhammer, D. (1999). The role of domain knowledge and graph reading skills in graph comprehension. Paper presented at *the Society of Applied Research in Memory and Cognition*, Boulder.

Shah, P., Freedman, E. G., and Vekiri, I. (2005). The comprehension of Quantitative Information in Graphical Displays. In P. Shah and A.Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinkin*g. New York: Cambridge University Press

Shah, P., Mayer, R. E., and Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, *91*(4), 690.

Shapiro, L. (2007). The embodied cognition research programme. *Philosophy compass*, 2(2), 338-346.

Shelton, A. L., and McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive psychology*, *43*(4), 274-310.

Shrock, S. A. (1995). A brief history of instructional development. In G. Anglin (Ed.), Instructional technology: Past, present, and future (2nd ed. pp. 11-19). Engelwood, CO: Libraries Unlimited.

Silver, P., Bourke, A., and Strehorn, K. C. (1998). Universal Instructional Design in higher education: An approach for inclusion. *Equity and Excellence*, 31(2), 47-51.

Simkin, D., and Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398), 454-465.

Singh, M., Fulvio, J. (2007). Bayesian contour extrapolation: Geometric determinants of good continuation. *Vision Research* 47, 783–798 (2007)

Sjöström, C., Danielsson, H., Magnusson, C., Rassmus-Gröhn, K. (2003). Phantom-based Haptic Line Graphics for Blind Persons. *Visual Impairment Research* 5, 13–32.

Spanger, P., Yasuhara, M., Iida, R., Tokunaga, T., Terai, A., and Kuriyama, N. (2012). REX-J: Japanese Referring Expression Corpus of Situated Dialogs. *Language Resources and Evaluation*, 46, 461-491.

Speer, N. K., Swallow, K. M., and Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, and Behavioral Neuroscience*, 3, 335–345.

Speer, N. K., Zacks, J. M., and Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, 18(5), 449-455.

Srinivasan, M. A. and Basdogan C. (1997). Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges. *Computer and Graphics* 21(4): 393-404.

Stephanidis, C. (2001). Adaptive techniques for universal access. *User modeling and user-adapted interaction*, 11(1-2), 159-179.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., and Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, 23(5), 1052-1064.

Swallow, K. M., Zacks, J. M., and Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138(2), 236.

Talmy, L. (1975). Semantics and syntax of motion. *Syntax and semantics*,4(181-238).

Talmy, L. (1983). How language structures space. In J. H. L. Pick, and L. P. Acredolo (Eds.), Spatial orientation: Theory, research and application New York: Plenum Press.

Tan, J. K., and Benbasat, I. (1990). Processing of graphical information: A decomposition taxonomy to match data extraction tasks and graphical representations. Information Systems Research, 416-439.

Tan, J. K., and Benbasat, I. (1993). The Effectiveness of Graphical Presentation for Information Extraction: A Cumulative Experimental Approach*. Decision Sciences, 24(1), 167-191.

Taylor, H. A., and Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. Journal of Memory and language, 31(2), 261-292.

Teo, L., and John, B. E. (2008). Cogtool-explorer: towards a tool for predicting user interaction. *In CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 2793-2798). ACM.

Thibadeau, R. (1986). Artificial perception of actions. *Cognitive Science*, 10 , 117-149.

Tomasino, B., Werner, C. J., Weiss, P. H., and Fink, G. R. (2007). Stimulus properties matter more than perspective: an fMRI study of mental imagery and silent reading of action phrases. *Neuroimage* 36(Suppl.2),T128–T141.

Traum, D. R., and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational intelligence*, *8*(3), 575-599.

Trickett, S. B., and Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, and N. Swoboda (Eds.), *Diagrammatic representation and inference* (pp. 286-300). Springer Berlin Heidelberg.

Trickett, S. B., Ratwani, R. M., and Trafton, J. G. (2006). Real-world graph comprehension: High-level questions, complex graphs, and spatial cognition.

Tripp, S.D. and Bichelmeyer, B. (1990). Rapid Prototyping: An alternative instructional design strategy. Educational Technology Research and Development, 38(1), 31-44.

Tschander, L. B., Schmidtke, H. R., Eschenbach, C., Habel, C., and Kulik, L. (2003). A geometric agent following route instructions. In Spatial cognition III (pp. 89-111). Springer Berlin Heidelberg.

Tufte, E. R. (1983). The visual display of quantitative information (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Tversky, B. (1990). Where partonomies and taxonomies meet. In S. L. Tsohatzidis (Ed.), *M e anings and prototypes: studies in linguistic categorization* (pp. 334-344). London: Routledge.

Tversky, B. (1991). Spatial mental models. The psychology of learning and motivation: Advances in research and theory, 27, 109-145.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In *Spatial Information Theory A Theoretical Basis for GIS* (pp. 14-24). Springer Berlin Heidelberg.

Tversky, B. (1996). Spatial perspective in descriptions. *Language and space*, *3*.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499-535.

Tversky, B., and Hemenway, K. (1984). Objects, parts, and categories. *Journal of experimental psychology: General*, 113(2), 169.

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129.

Tversky, B., and Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In Spatial information theory. *Cognitive and computational foundations of geographic information science* (pp. 51-64). Springer Berlin Heidelberg.

Tversky, B., Heiser, J., Lee, P., and Daniel, M.-P. (2009). Explanations in gesture, diagram, and word. *In K. R. Coventry, T. Tenbrink, and J. A. Bateman (Eds.), Spatial language and dialogue* (pp. 119–131). Oxford, England: Oxford University Press.

Tversky, B., Zacks, J. M., and Hard, B. M. (2008). The structure of experience. *Understanding events: How humans see, represent, and act on events*, 436-464.

Van Deemter, K., Gatt, A., van Gompel, R. P., and Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in cognitive science*, 4(2), 166-183.

Van Deemter, K., Krahmer, E., and Theune, M. (2005). Real versus template-based natural language generation: A false opposition?. *Computational Linguistics*, 31(1), 15-24.

van der Horst, B. J., and Kappers, A. M. (2008). Haptic curvature comparison of convex and concave shapes. *Perception*, *37*(8), 1137-1151.

Van Der Sluis, I., Luz, S., Breitfuß, W., Ishizuka, M., and Prendinger, H. (2012). Cross-cultural assessment of automatically generated multimodal referring expressions in a virtual world. International Journal of Human-Computer Studies, 70(9), 611-629.

Van der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. Journal of Memory and Language, 60(4), 448-463.

Van Dijk, T. A., and Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. Information systems research, 11(4), 342-365.

Vogels, I. M., Kappers, A. M., and Koenderink, J. J. (1999). Influence of shape on haptic curvature perception. *Acta psychologica*, *100*(3), 267-289.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological bulletin*, *138*(6), 1172.

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. Speech Communication, 57, 209-232.

Wall, S. and Brewster, S. (2006). Feeling what you hear: Tactile feedback for navigation of audio graphs. *In CHI'06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'06). ACM, New York, 1123–1132.

Watson, M. E., Pickering, M. J., and Branigan, H. P. (2004). Alignment of reference frames in dialogue. In *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 2353-2358). Mahwah, NJ: Lawrence Erlbaum.

Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. *In D. R. Kenshalo (Ed.), The skin senses* (pp. 195-222). Springfield, IL: Thomas.

Weka 3: Data Mining Software in Java," http://www.cs.waikato.ac.nz/ml/weka

Welch, R. B. and D. H. Warren (1980). Immediate Perceptual Response to Intersensory Discrepancy. *Psychological Bulletin* 88(3): 638-667.

Wickens, C. D., and Carswell, C. M. (1995). The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(3), 473-494.

Wijntjes, M. W. A., Sato, A., Hayward, V. and Kappers, A. M. L. 2009 Local surface orientation dominates haptic curvature discrimination. *IEEE Trans. Haptic* 2, 94–102.

Wilkes-Gibbs, D. and Clark, H. H. (1992) Coordinating beliefs in conversation. *Journal of Memory and Language* 31:183–94.

Willems, R. M., Hagoort, P., and Casasanto, D. (2010). Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychological Science*, *21*(1), 67-74.

Wu, P., Carberry, S., Elzer, S., and Chester, D. (2010). Recognizing the Intended Message of Line Graphs. In: Goel, A.K., Jamnik, M., and Narayanan, N.H. (eds.) Diagrammatic Representation and Inference. (pp. 220–234). Springer, Heidelberg.

Yoon, S. O., Koh, S., and Brown- Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin and Review*. 19, 699–707.

Yu, J. (2014). A Haptic-audio Human-Computer Interface : Acquisition of Indoor Spatial Knowledge by Visually Impaired People. (Doctoral dissertation). University of Hamburg. Retrieved from http://ediss.sub.uni-hamburg.de/volltexte/2014/7073/

Yu, J., Habel, C (2012). A Haptic-Audio Interface for Acquiring Spatial Knowledge about Apartments. In: Magnusson, C.; Szymczak, D. and Brewster, S. (eds.). *Haptic and Audio Interaction Design* 2012. LNCS, vol. 7468, pp. 21–30. Springer, Berlin.

Yu, W. and Brewster, S. (2002). Comparing two haptic interfaces for multimodal graph rendering. *In Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'02*). IEEE Computer Society, Los Alamitos, CA, 3–9.

Yu, W., and Brewster, S.A. (2003). Evaluation of Multimodal Graphs for Blind People. *Journal of Universal Access in the Information Society* 2, 105-124.

Yu, W., Ramloll, R., and Brewster, S. A. 2001. Haptic graphs for blind computer users. *In Proceedings of the1st International Workshop on Haptic Human-Computer Interaction*. Springer-Verlag, 41–51.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979-1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001a). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), 651–655

Zacks, J. M., Kumar, S., Abrams, R. A., and Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, 112(2), 201-216.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*,133(2), 273.

Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80-84.

Zacks, J. M., Swallow, K. M., Vettel, J. M., and McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain research*, 1076(1), 150-162.

Zacks, J., and Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory and Cognition*, 27, 1073-1079

Zacks, J. M., and Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, 127(1), 3-21.

Zacks, J. M., and Tversky, B. (2005). Multiple systems for spatial imagery: transformations of objects and bodies. *Spatial Cognition and Computation*, 5(4), 271-306.

Zacks, J., and Tversky, B. (2012). Granularity in taxonomy, time, and space. *Motion Encoding in Language and Space.*

Zacks, J. M., Tversky, B., and Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130(1), 29.

Zacks, J., Levy, E., Tversky, B., and Schiano, D. (2002). Graphs in Print. In M. Anderson, B. Meyer, and P. Olivier (Eds.), Diagrammatic Representation and Reasoning. pp. 187–206. Springer, London.

Zeng, L., and Weber, G. (2010). Audio-haptic browser for a geographical information system. In K. Miesenberger, J. Klaus, W. Zagler, and A. Karschmer (Eds.), *Computers helping people with special needs*, part II (pp. 466–473). Berlin, Heidelberg: Springer.

Zhang M,Weisser V D, Stilla R, Prather S C, and Sathian K. (2004). Multisensory cortical processing of object shape and its relation to mental imagery. *Cognitive, Affective, and Behavioral Neuroscience* (4). pp. 251- 259

Zhao, H., Plaisant, C., Shneiderman, B., and Lazar, J. (2008). Data sonification for users with visual impairment: a case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1), 4.

Zwaan, R. A. and Radvansky, G. A. (1998) Situation models in language comprehension and memory. *Psychological Bulletin.* 123:162–85.

Zwaan, R. A., and Taylor, L. J.(2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General.* 135, 1–11.

Zwickel, J.(2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin and Review* . 16, 1089–1093.

APPENDIX A: The Demographic Data Forms

(In German/Turkish/English)

Teilnehmer Nr. (Doğum yılı / Date of Birth):______

Alter (Yaş, Age): ______

Geschlecht (Cinsiyet/Gender): Weiblich: ___ Männlich ___

Händigkeit (Handedness): Rechts ___ Links: ___

Muttersprache(Anadil/Mother-tongue): __________

Ausbildung (Eğitim durumu/ Educational Background/State)

Student/in : ___	Absolvent/in: ___	Sonstiges (bitte erläutern): ___
(Lisans/undergraduate)	(Yükseklisans/graduate)	(Diğer/other)

Fachbereich/ Studienfach (Bölüm/Department of study): ______________________

APPENDIX B: The Sample for a Consent Form (In German)

Teilnahme am Experiment

„Multimodal Communication Through Graphs“

Lieber Teilnehmer des Experiments. Mit Ihrer Teilnahme unterstützen Sie die Wissenschaft. Wir sind uns sicher, dass Sie unsere Experimente als interessant und nicht als unangenehm empfinden. Dennoch möchten wir Sie vorab über Ihre Rechte informieren.

Allgemeine Information über die Studie: Es gibt zwei Teilexperimente in dieser Session; alle notwendigen Informationen werden Ihnen in den Anleitungsfolien direkt vor den Experimenten gegeben. Nachdem Sie die Anleitung gelesen haben, werden Ihnen in einer Eingewöhnungsphase einige Beispiele gezeigt, die das Hauptexperiment simulieren. Sobald Sie bereit sind, beginnt das Hauptexperiment. Während des Hauptexperiments können wir Ihre Fragen nicht beantworten. Deshalb bitten wir Sie, eventuelle Fragen während der Eingewöhnungsphase zu stellen.

(a) Abbruch des Experiments und Bedenken

Sie dürfen jederzeit und ohne Angabe von Gründen das Experiment unter- oder abbrechen. Bitte sprechen Sie jegliche Bedenken vor und während des Experiments mit dem Versuchsleiter ab.

(b) Datenschutz

Im Laufe des Experiments werden persönliche Daten von Ihnen aufgenommen. Dies schließt Daten wie Ihr Geburtsdatum und Geschlecht ein, sowie Video- und/oder Tonaufzeichnungen. Diese Daten werden unter allen Umständen vertraulich behandelt und nicht in Zusammenhang mit Ihrem Namen gespeichert.

(c) Ziel und Ergebnisse des Experiments

Leider ist es bei der Art Experimente, die wir durchführen, nicht möglich, Sie vorher über das Ziel aufzuklären. Der Versuchsleiter wird Sie nach Ablauf des Experiments über das Ziel des Experiments aufklären. Wenn Sie wünschen, über Ergebnisse des Experiments informiert zu werden, geben Sie bitte unten Ihre E-Mail-Adresse an.

Mit meiner Unterschrift erkläre ich, die oben stehenden Bedingungen zur Kenntnis genommen und verstanden zu haben und erkläre mich mit den Bedingungen einverstanden.

____________________ ______._______. 2012 ___________________________

(Name) (Datum) (E-Mail-Adresse – freiwillig für Ergebnisse)

____________________ ______._______. 2012

(Versuchleiter) (Datum)

APPENDIX C: The Drawing Sheet Sample

Experiment: Multimodal Communication Through Line Graphs

Drawing Sheet

Please draw the graph here as much as you remember.

(__________)

(this field will be filled by experimenter)

APPENDIX D: The Spatial Term Survey

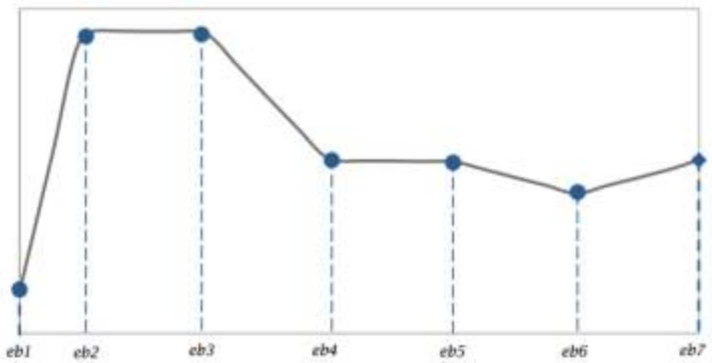
Fuzzy Spatial Prepositions and Haptic Ostensive Expressions			
		1-2-3-4-5 Ne kadar anlamlı? How meaningful is this?	Anlamı ne? What does it mean?
1.	Arkaya dogru (backward)		
2.	Artıyor (it is increasing)		
3.	Asagı gidiyor (it is going downward)		
4.	Bası (it's the beginning)		
5.	Baslangıc noktasında mıyım? (Am I at the starting point?)		
6.	Cerceve icinde miyim? (Am I inside of the frame?)		
7.	Cerceve uzerinde miyim? (Am I on the frame?)		
8.	Cizgi icinde miyim? (Am I inside of the line?)		
9.	Cizgi üzerinde miyim? (Am I on the line?)		
10.	Cukur yapmıs (It made a cavity)		
11.	Dalgalı bir tepe (wavy peak)		
12.	Dalgalı yukselmis (It has a wavy increase)		
13.	Dip yapmıs (It made a deep)		
14.	Dusuyor (It is falling)		
15.	Düz mu (Is it flat?)		
16.	Egim var mı? (Is there a incline/slope?		
17.	Egimli mi? (Is it inclined?)		
18.	Eksene yapısık mı? (Is it connected to the frame?)		
19.	Grafigin dısında mıyım? (Am I outside of the graph?)		
20.	Grafigin üzerinde miyim? (Am I on the graph?)		
21.	Grafik buradan mı baslıyor? (Is the graph starting here?)		
22.	Grafigin icinde miyim? (Am I inside of the graph?)		
23.	Ikinci basamaktan sonra (after the second step)		
24.	Kırılma yapmıs (It made a fraction)		
25.	M seklinde (M shape)		
26.	Maximummus (maxima)		
27.	Minimum noktalardan biri (one of the minima points)		
28.	Öne gidiyor (it is going front)		
29.	Oradan asagıya git (go downward from there)		
30.	Sagdan devam et (keep going right)		

31.	Sagdan sonu (the right end)		
32.	Soldan sonu (the left end)		
33.	Sonu (the end)		
34.	U seklinde (U shape)		
35.	Yatay mı (Is it horizontal)		
36.	Yukarı gidiyor (It goes upward)		
37.	Bastan sona (from the beginning to the end)		
38.	Tepeler arası (in between the peaks)		
39.	ileri gidiyor (it is going forward)		
40.	Geri gidiyor (It is going backward)		

APPENDIX E: The location of event boundaries and their referring percentages

* indicates the significant difference.

Graph-1



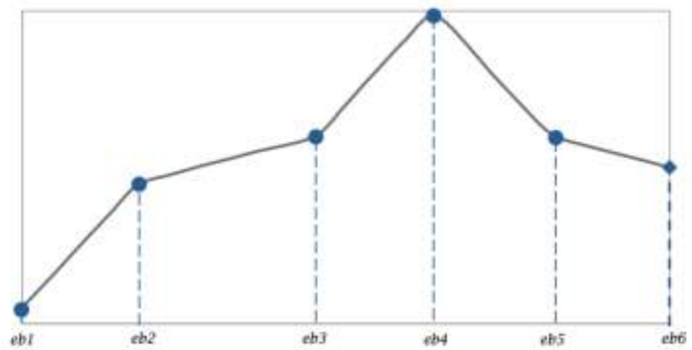
Effect of Data Labels						
Event Start	***eb1***	***eb2***	***eb3****	***eb4***	***eb5***	***eb6***
VG- with labels	80	73.3	100	53.3	33.3	6.7
VG- without labels	93.3	86.7	66.7	60	40	20

End of Global Event	***eb3***	...	***eb7***
VG- with labels			100
VG- without labels	8.3		93.3

Effect of Modality						
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***	***eb6***
VG- without labels	93.3	86.7	66.7	60	40	20
HG -without labels	100	58.3	91.7	25	16.7	16.7

End of Global Event	***eb3***	...	***eb7***
VG- without labels	8.3		93.3
HG -without labels	8.3		91.7

Graph-2



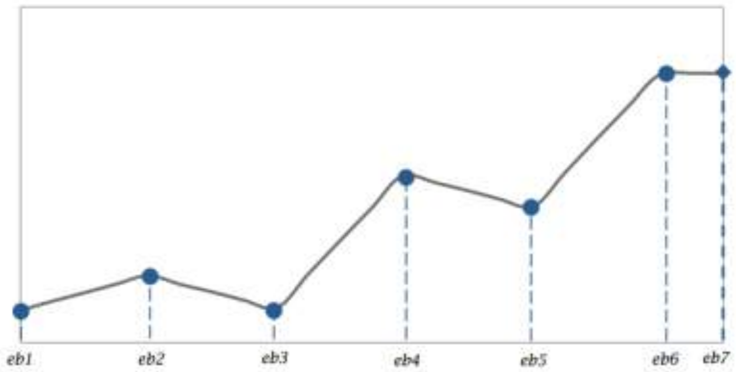
Effect of Data Labels					
Event Start	***eb1***	***eb2***	***eb3****	***eb4***	***eb5***
VG- with labels	100	28.6	21.4	100	14.3
VG- without labels	100	53.3	66.7	86.7	20

End of Global Event	***eb4***	***eb5***	***eb6***
VG- with labels			100
VG- without labels	6.7	-	93.3

Effect of Modality					
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***
VG- without labels	100	53.3	66.7	86.7	20
HG -without labels	93.3	53.3	53.3	93.3	20

End of Global Event	***eb4***	***eb5***	***eb6***
VG- without labels	6.7	-	93.3
HG -without labels	6.7	-	93.3

Graph-3



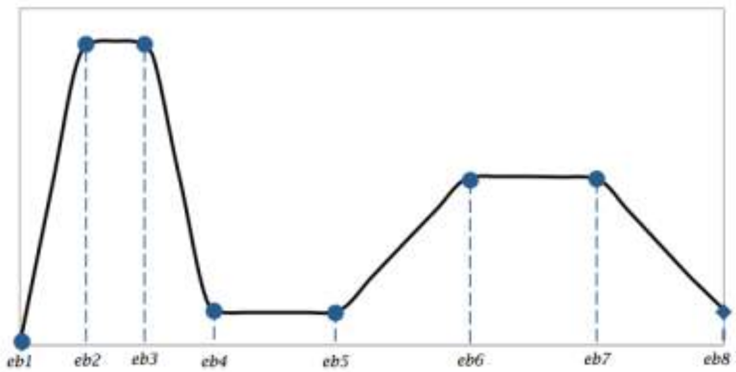
Effect of Data Labels						
Event Start	***eb1***	***eb2****	***eb3***	***eb4****	***eb5***	***eb6***
VG- with labels	53.8	15.4	46.2	15.4	53.8	23.1
VG- without labels	73.3	46.7	53.3	46.7	60	33.3

Effect of Modality						
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***	***eb6***
VG- without labels	73.3	46.7	53.3	46.7	60	33.3
HG -without labels	75	56.3	81.3	62.5	75	37.5

End of Global Event	***eb5***	***eb6***	***eb7***	***eb8***
VG- with labels		6.7	6.7	53.8
VG- without labels		6.7	6.7	80

End of Global Event	***eb5***	***eb6***	***eb7***	***eb8***
VG- without labels		6.7	6.7	80
HG -without labels	6.3	0	12.5	81.3

Graph-4



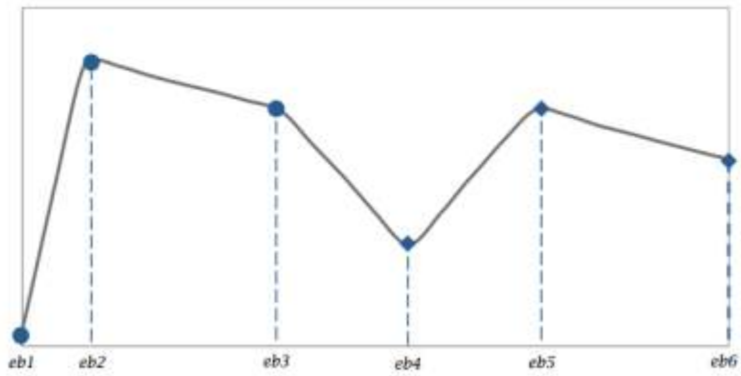
Effect of Data Labels							
Event Start	***eb1****	***eb2****	***eb3***	***eb4***	***eb5***	***eb6***	***eb7***
VG- with labels	71.4	64.3	92.9	50	87.5	50	35.7
VG- without labels	100	73.3	80	60	93.3	53.3	66.7

Effect of Modality							
Event Start	***eb1***	***eb2****	***eb3***	***eb4***	***eb5***	***eb6***	***eb7***
VG- without labels	100	73.3	80	60	93.3	53.3	66.7
HG -without labels	100	26.7	100	60	86.7	40	73.3

End of Global Event	***eb5***	***eb6***	***eb7***	***eb8***
VG- with labels		35.7	14.3	42.9
VG- without labels		6.7	6.7	86.7

End of Global Event	***eb5***	***eb6***	***eb7***	***eb8***
VG- without labels		6.7	6.7	86.7
HG -without labels	6.7	-	6.7	93.3

Graph-5



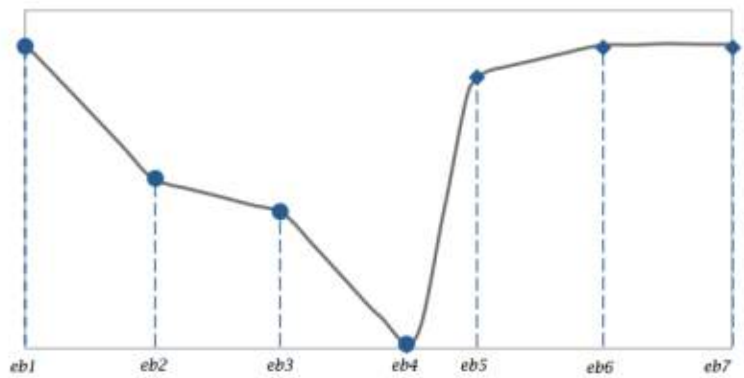
Effect of Data Labels					
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5****
VG- with labels	92.3	92.3	38.5	92.3	23.1
VG- without labels	93.3	93.3	66.7	86.7	66.7

Effect of Modality					
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***
VG- without labels	93.3	93.3	66.7	86.7	66.7
HG -without labels	100	100	50	100	92.9

End of Global Event	***eb5***	***eb6***	***eb7***
VG- with labels	7.7	15.4	46.2
VG- without labels	13.3	-	80

End of Global Event	***eb5***	***eb6***	***eb7***
VG- without labels	13.3	-	80
HG -without labels		7.1	92.9

Graph-6



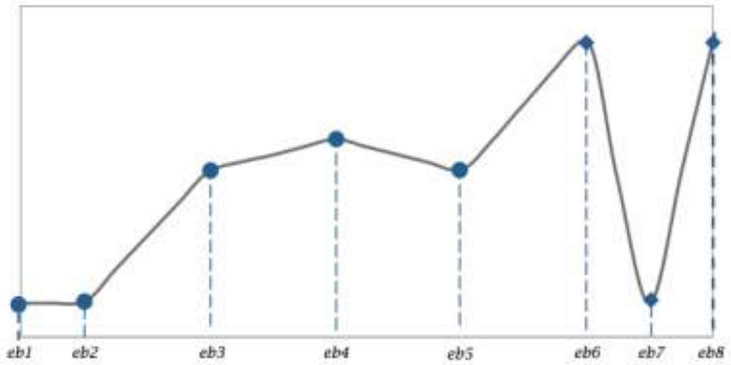
Effect of Data Labels						
Event Start	***eb1***	***eb2***	***eb3****	***eb4***	***eb5***	***eb6****
VG- with labels	100	15.4	15.4	100	69.2	0
VG- without labels	87.5	31.3	31.3	93.8	43.8	31.3

Effect of Modality						
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***	***eb6***
VG- without labels	87.5	31.3	31.3	93.8	43.8	31.3
HG -without labels	100	26.7	33.3	100	20	46.7

End of Global Event	***eb5***	***eb6***	***eb7***
VG- with labels	15.4	7.7	76.9
VG- without labels	6.3	-	87.5

End of Global Event	***eb5***	***eb6***	***eb7***
VG- without labels	6.3	-	87.5
HG -without labels			100

Graph-7



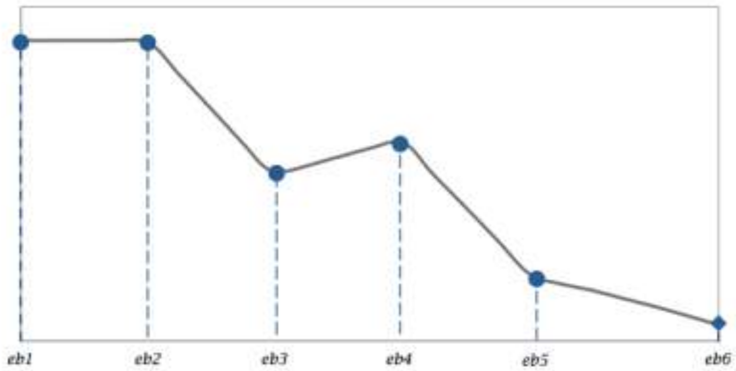
Effect of Data Labels							
Event Start	eb1	eb2	eb3	eb4	eb5*	eb6	eb7
VG- with labels	62.5	82.3	12.5	12.5	31.3	93.8	75
VG- without labels	86.7	73.3	26.7	26.7	80	80	60

Effect of Modality							
Event Start	eb1*	eb2	eb3	eb4*	eb5	eb6	eb7
VG- without labels	86.7	73.3	26.7	26.7	80	80	60
HG -without labels	50	75	12.5	68.5	56.3	87.5	81.2

End of Global Event	eb6	eb7	eb8
VG- with labels		25	75
VG- without labels		40	60

End of Global Event	eb6	eb7	eb8
VG- without labels		40	60
HG -without labels	6.3	6.3	87.5

Graph-8



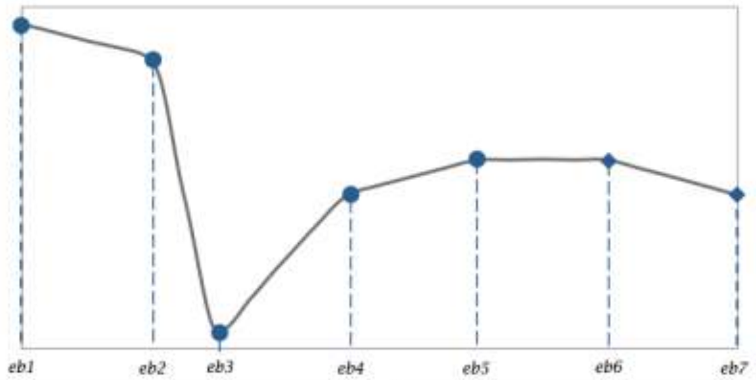
Effect of Data Labels					
Event Start	eb1	eb2	eb3	eb4	eb5
VG- with labels	62.5	75	56.3	75	31.3
VG- without labels	87.5	87.5	56.3	68.7	37.5

Effect of Modality					
Event Start	eb1	eb2*	eb3	eb4	eb5
VG- without labels	87.5	87.5	56.3	68.7	37.5
HG -without labels	66.7	53.3	66.7	53.3	33.3

End of Global Event	eb5	eb6
VG- with labels	18.8	68.8
VG- without labels		93.8

End of Global Event	eb5	eb6
VG- without labels		93.8
HG -without labels		80

Graph-9



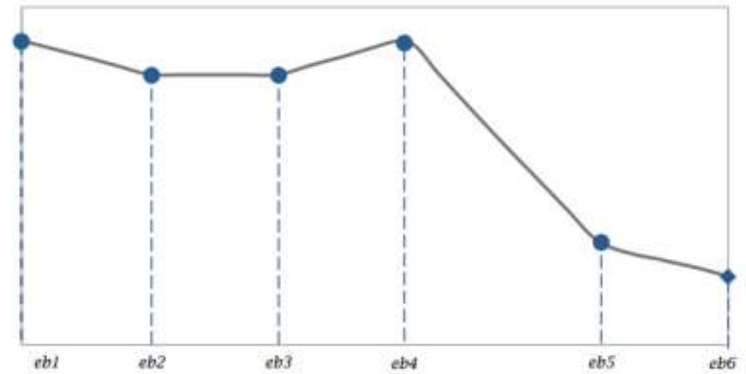
Effect of Data Labels						
Event Start	***eb1***	***eb2***	***eb3***	***eb4****	***eb5***	***eb6***
VG- with labels	93.8	81.3	100	12.5	12.5	0
VG- without labels	81.3	75	93.8	43.8	37.5	18.8

Effect of Modality						
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***	***eb6***
VG- without labels	81.3	75	93.8	43.8	37.5	18.8
HG -without labels	100	46.7	93.3	40	33.3	26.7

End of Global Event	***eb6***	***eb7***
VG- with labels	6.3	93.7
VG- without labels	6.3	87.5

End of Global Event	***eb6***	***eb7***
VG- without labels	6.3	87.5
HG -without labels		100

Graph-10



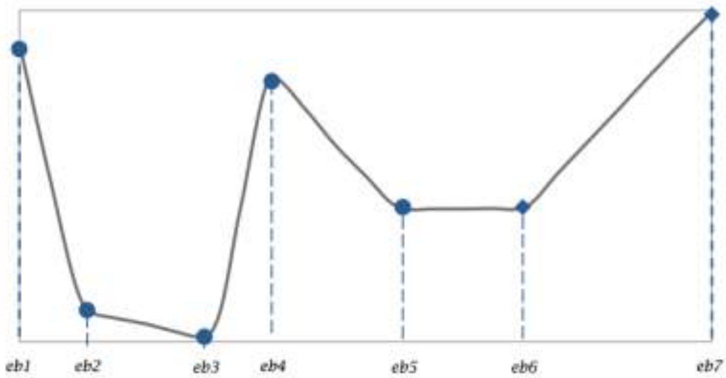
Effect of Data Labels					
Event Start	***eb1***	***eb2****	***eb3***	***eb4***	***eb5***
VG- with labels	100	6.3	12.5	100	18.8
VG- without labels	92.9	42.9	43	92.9	35.7

Effect of Modality					
Event Start	***eb1***	***eb2***	***eb3***	***eb4***	***eb5***
VG- without labels	92.9	42.9	43	92.9	35.7
HG -without labels	80	13.3	73.3	80	33.3

End of Global Event	***eb7***
VG- with labels	100
VG- without labels	92.9

End of Global Event	***eb7***
VG- without labels	92.9
HG -without labels	93.3

Graph-11



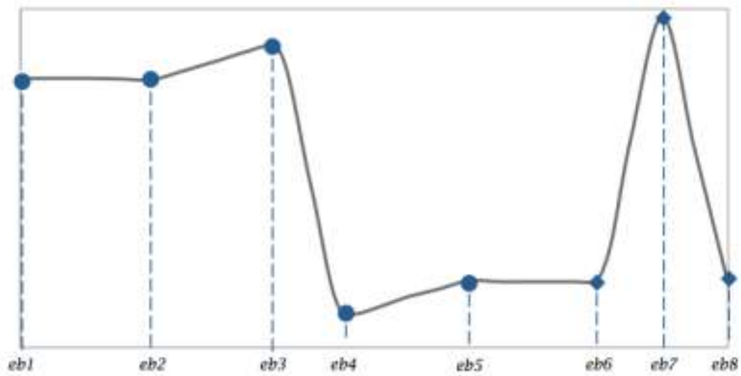
Effect of Data Labels						
Event Start	eb1	eb2*	eb3	eb4	eb5	eb6
VG- with labels	86.7	20	86.7	60	53.8	66.7
VG- without labels	93.8	62.5	87.5	81.3	43.8	50

Effect of Modality						
Event Start	eb1	eb2*	eb3	eb4	eb5	eb6
VG- without labels	93.8	62.5	87.5	81.3	43.8	50
HG -without labels	100	26.7	100	93.3	26.7	80

End of Global Event	eb4	eb5	eb6	eb7
VG- with labels			6.7	86.7
VG- without labels	6.3	6.3	6.3	75

End of Global Event	eb4	eb5	eb6	eb7
VG- without labels	6.3	6.3	6.3	75
HG -without labels	6.7	-	-	93.3

Graph-12



Effect of Data Labels							
Event Start	eb1	eb2*	eb3	eb4	eb5	eb6	eb7
VG- with labels	86.7	13.3	93.3	60	0	66.7	66.7
VG- without labels	93.3	46.7	73.3	66.7	13.3	66.7	60

Effect of Modality							
Event Start	eb1	eb2	eb3	eb4	eb5	eb6	eb7
VG- without labels	93.3	46.7	73.3	66.7	13.3	66.7	60
HG -without labels	85.7	28.6	92.9	64.3	14.3	71.4	85.7

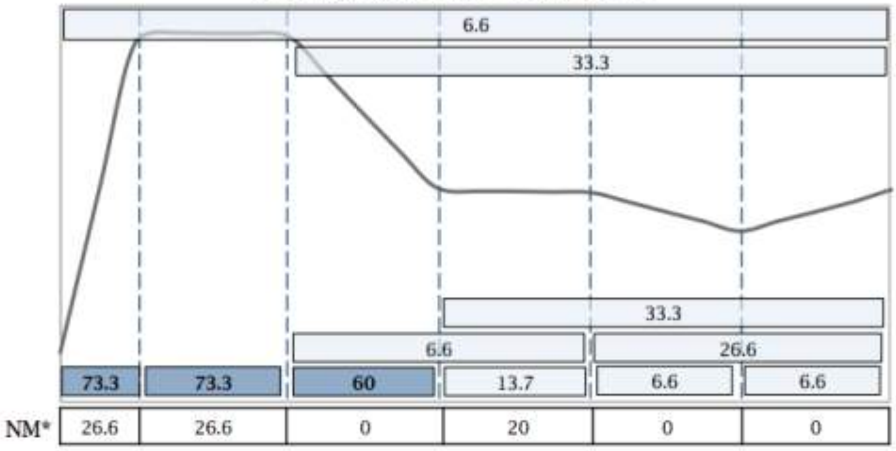
End of Global Event	eb6	eb7	eb8
VG- with labels	6.7	20	66.7
VG- without labels		40	53.3

End of Global Event	eb6	eb7	eb8
VG- without labels		40	53.3
HG -without labels		7.14	85.7

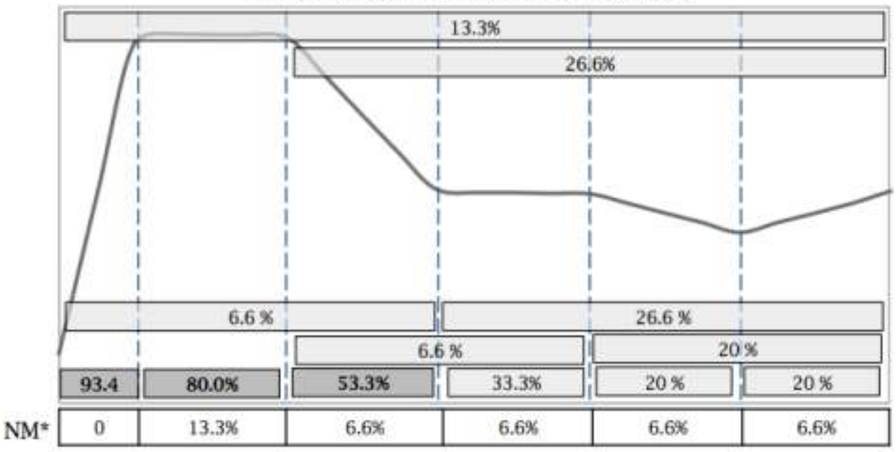
APPENDIX F: Reference Scopes Rates for each segments

Graph-1

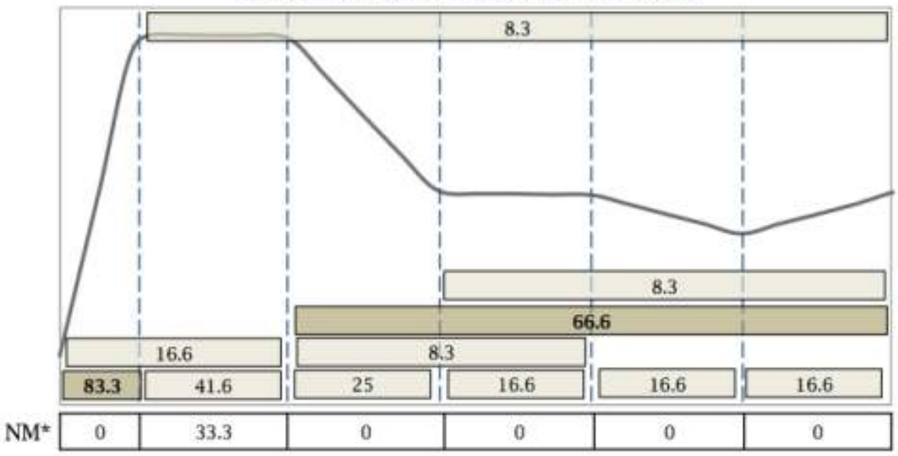
Visual graphs with data labels



Visual graphs without data labels



Haptic graphs without data labels

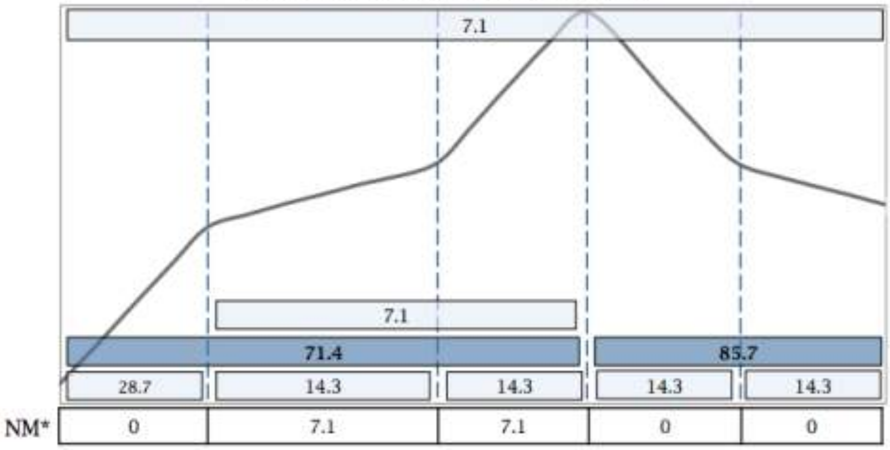


*NM: referring percentages for none-mentioned (NM) segments

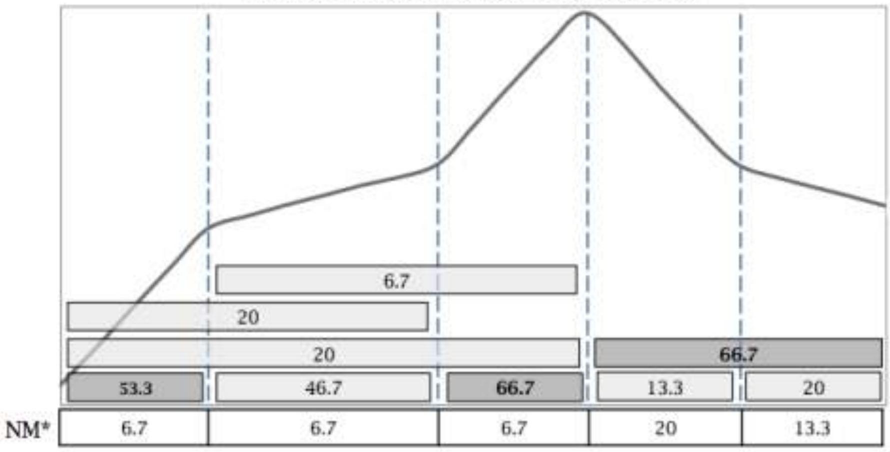
Darker background is used to highlight the segments which are mentioned by the majority of the participants (above 50%) in their post-exploration verbal descriptions.

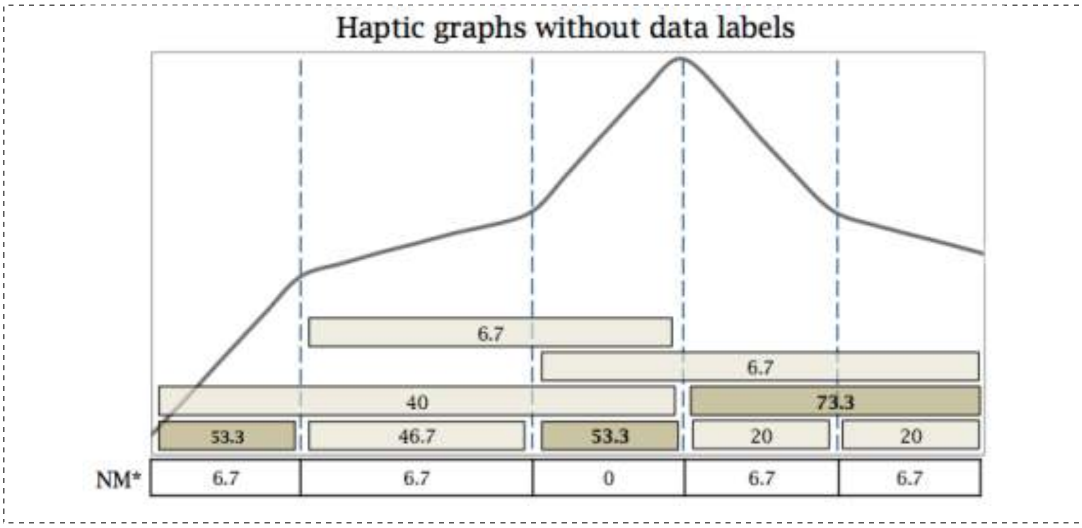
Graph-2

Visual graphs with data labels

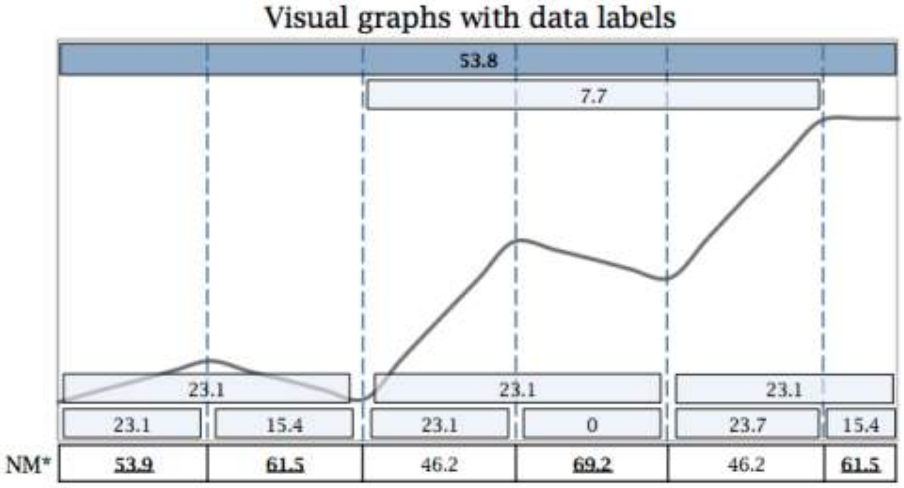


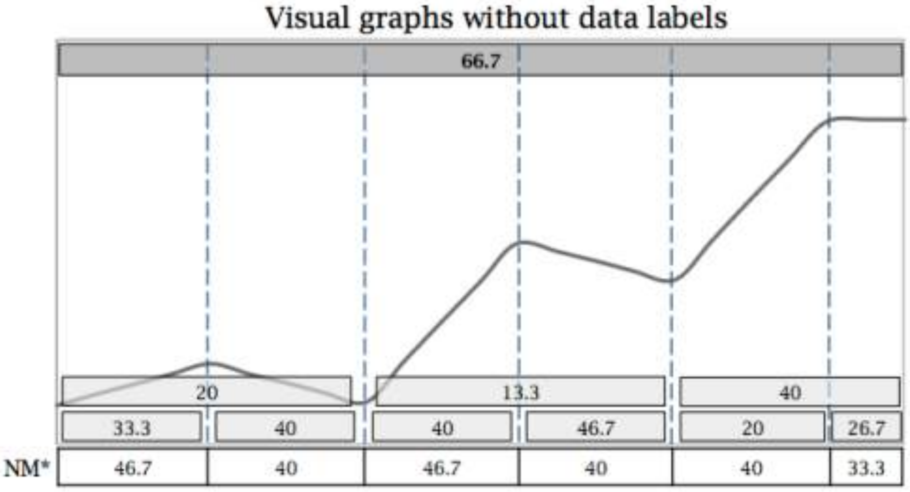
Visual graphs without data labels

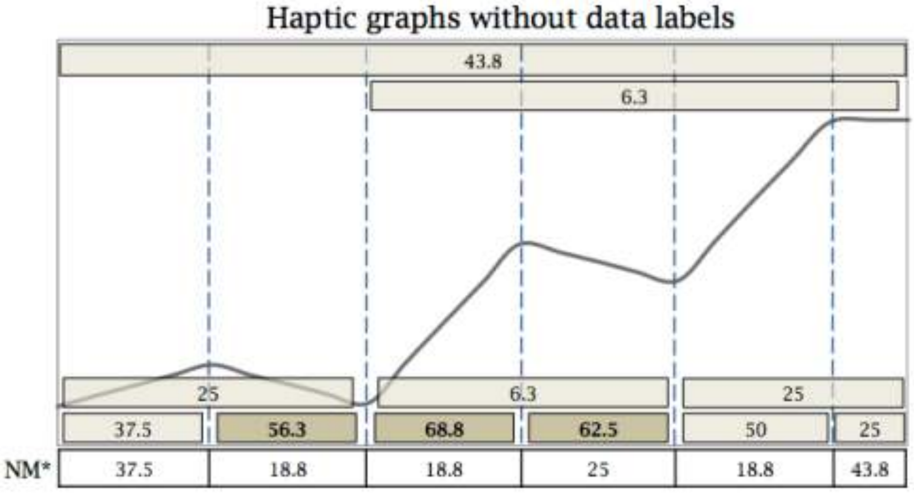



Haptic graphs without data labels
6.7
6.7
40
73.3
53.3
46.7
53.3
20
20
NM*
6.7
6.7
0
6.7
6.7

Graph-3

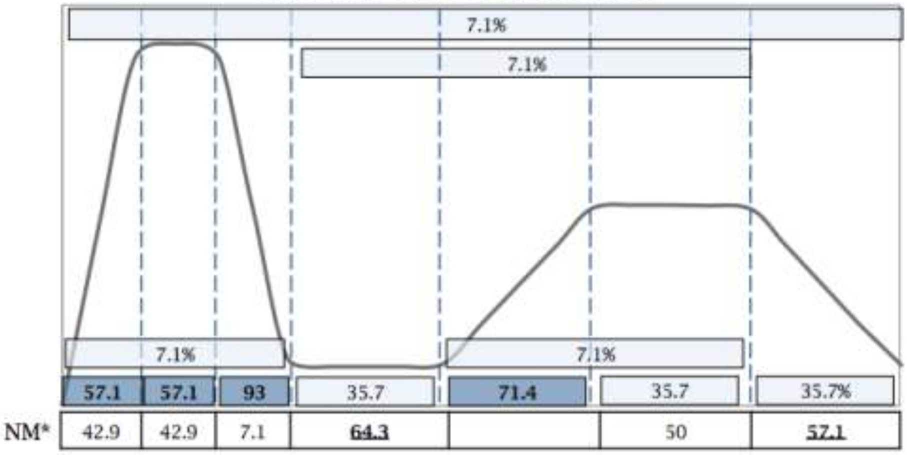

Visual graphs with data labels
53.8
7.7
23.1
23.1
23.1
23.1
15.4
23.1
0
23.7
15.4
NM*
53.9
61.5
46.2
69.2
46.2
61.5


Visual graphs without data labels
66.7
20
13.3
40
33.3
40
40
46.7
20
26.7
NM*
46.7
40
46.7
40
40
33.3

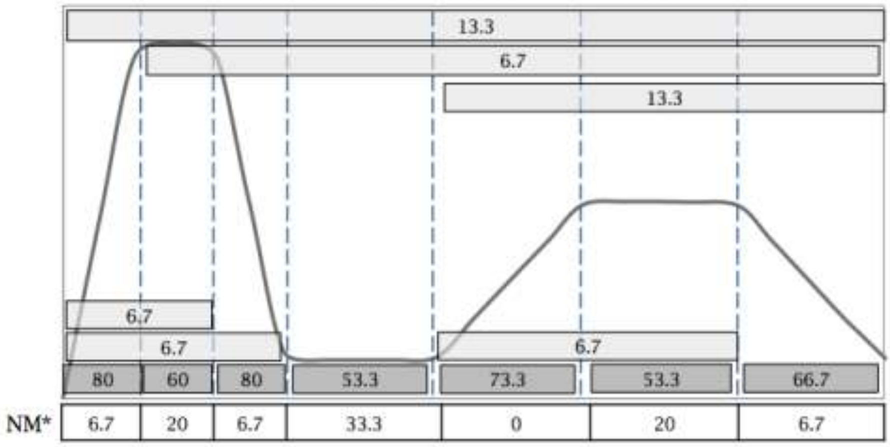

Haptic graphs without data labels
43.8
6.3
25
6.3
25
37.5
56.3
68.8
62.5
50
25
NM*
37.5
18.8
18.8
25
18.8
43.8

Graph-4

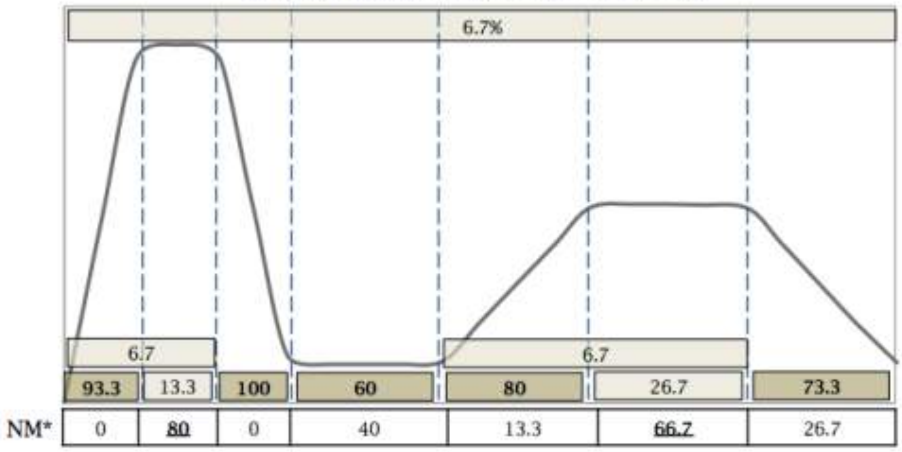
Visual graphs with data labels



Visual graphs without data labels

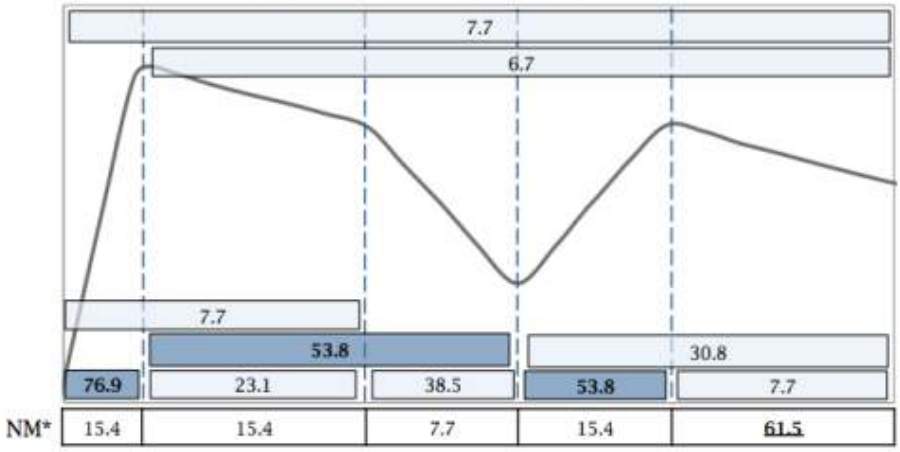


Haptic graphs without data labels

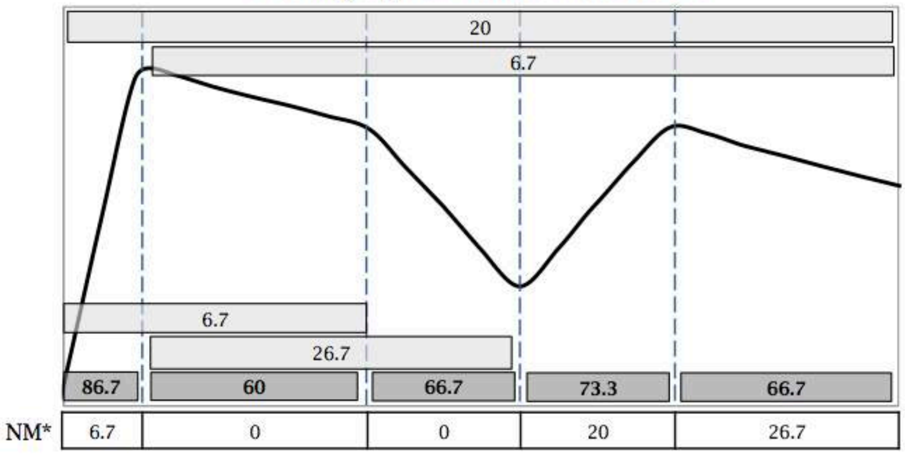


Graph-5

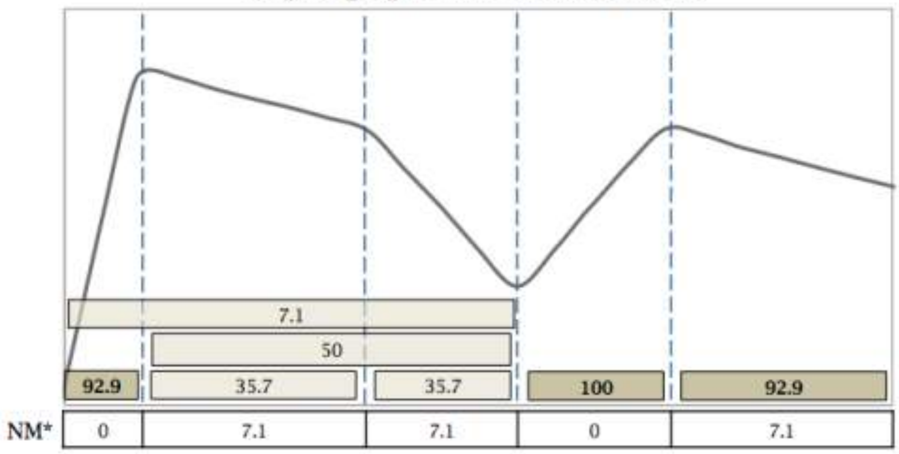
Visual graphs with data labels



Visual graphs without data labels

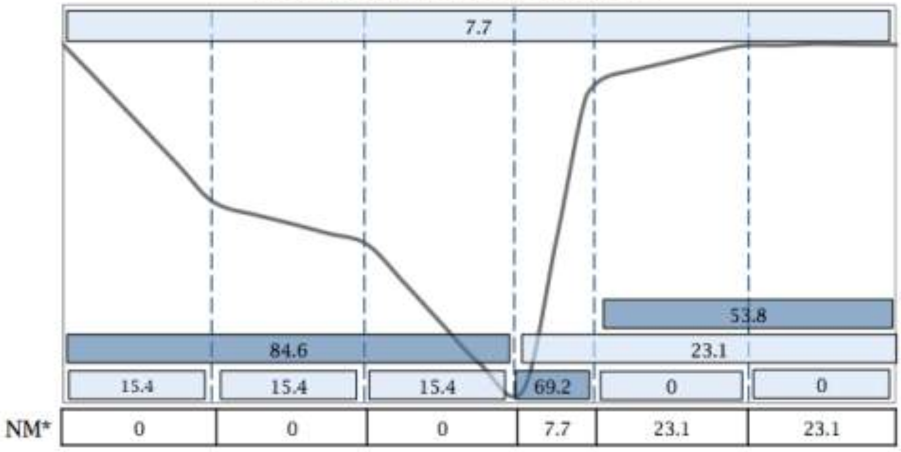


Haptic graphs without data labels

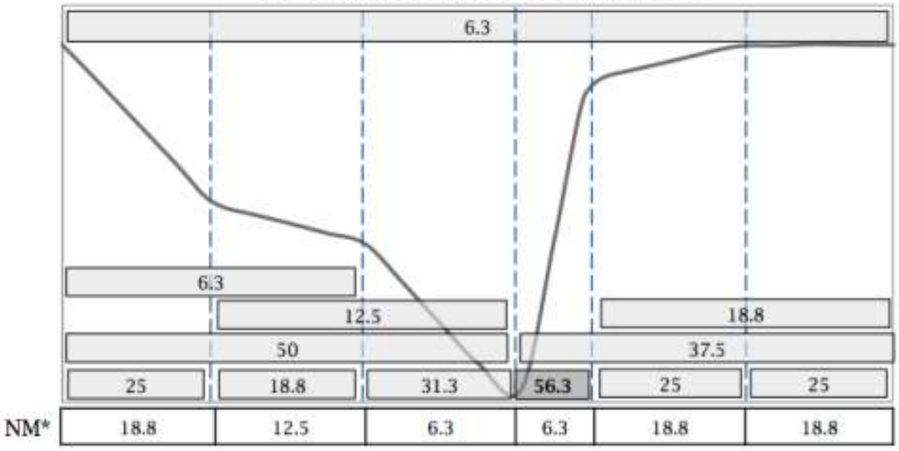


Graph-6

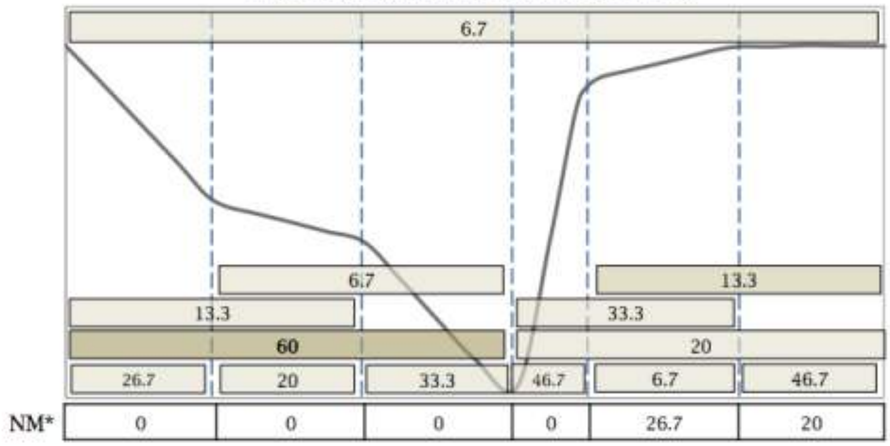
Visual graphs with data labels



Visual graphs without data labels

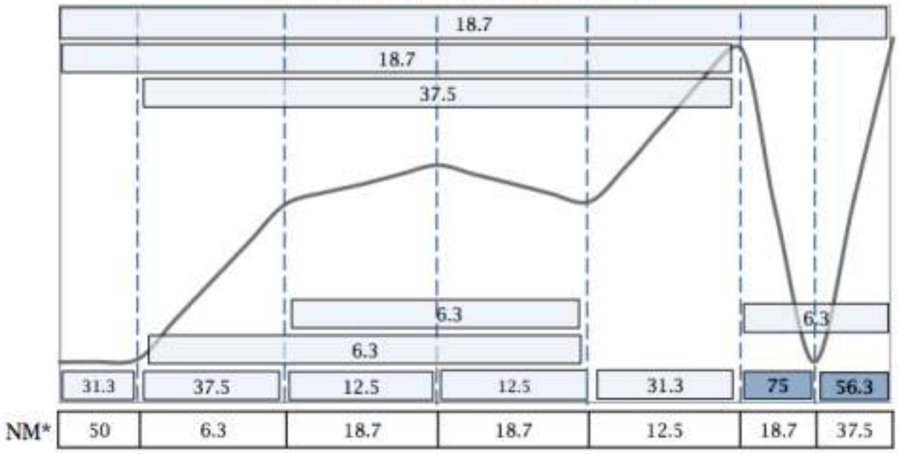


Haptic graphs without data labels

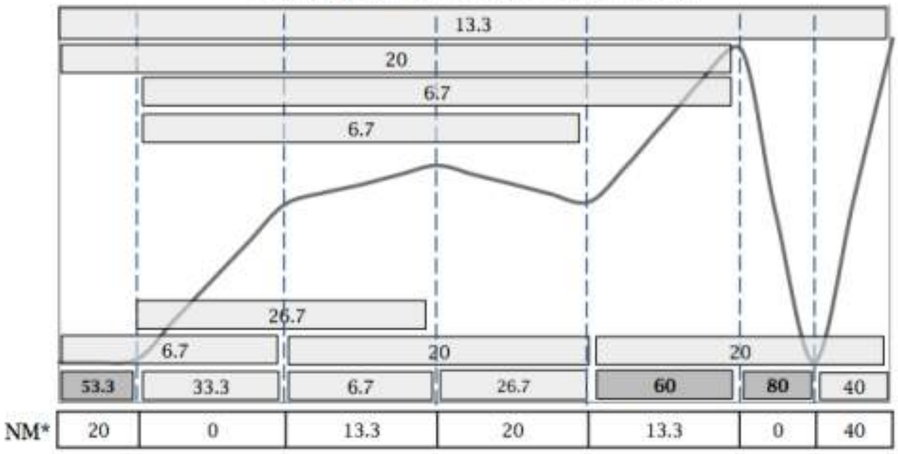


Graph-7

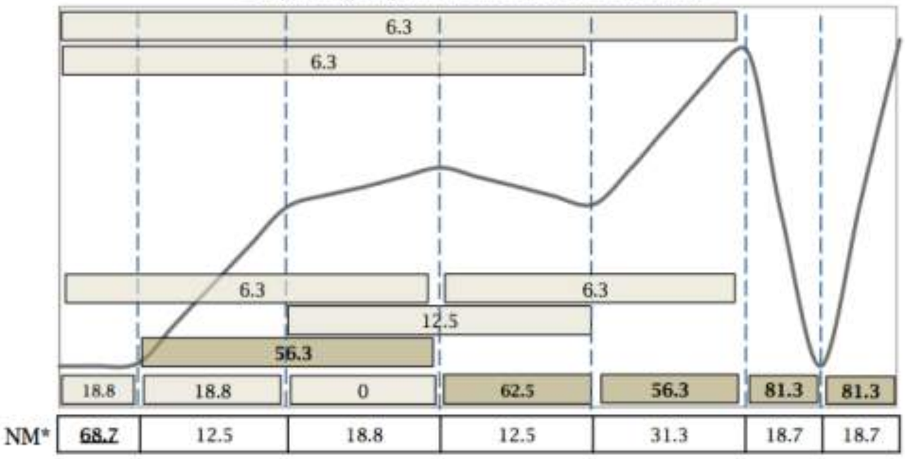
Visual graphs with data labels



Visual graphs without data labels

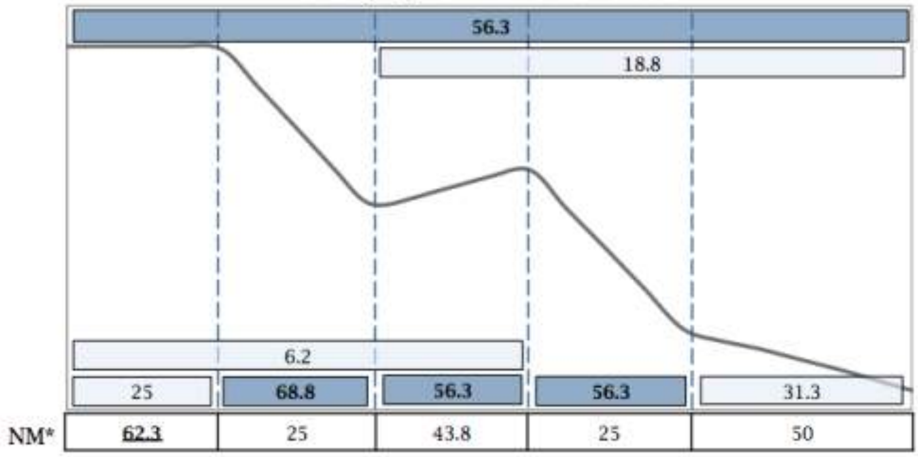


Haptic graphs without data labels

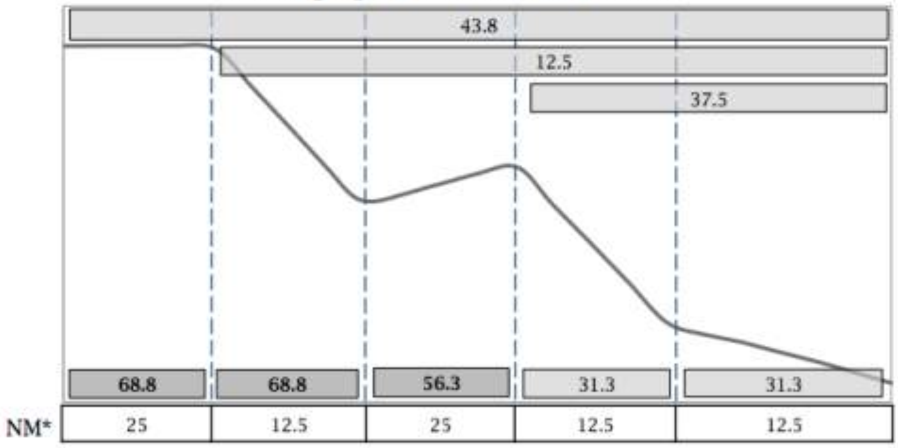


Graph-8

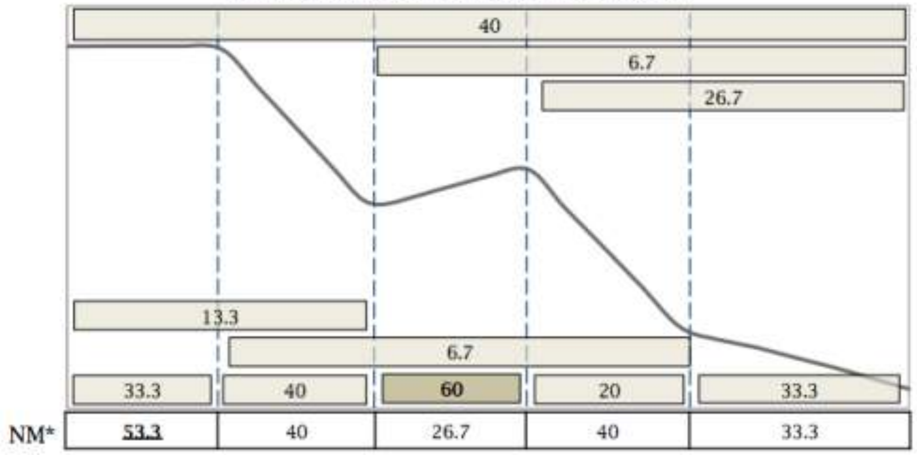
Visual graphs with data labels



Visual graphs without data labels

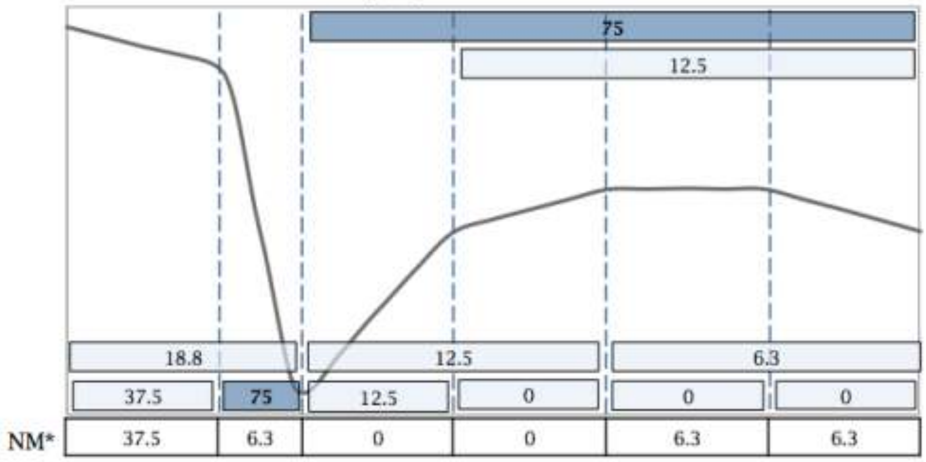


Haptic graphs without data labels

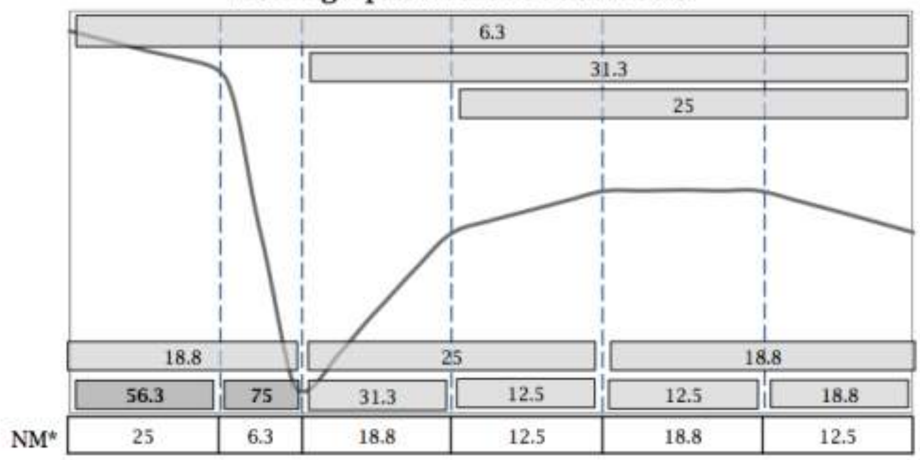


Graph-9

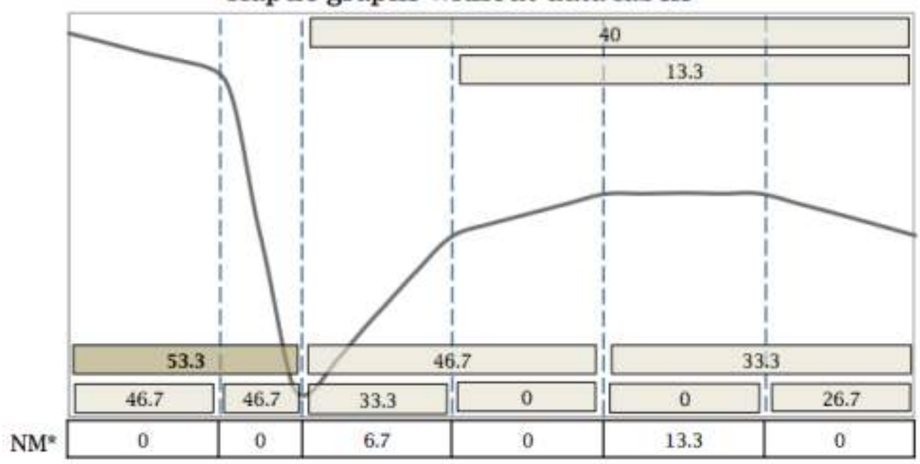
Visual graphs with data labels



Visual graphs without data labels

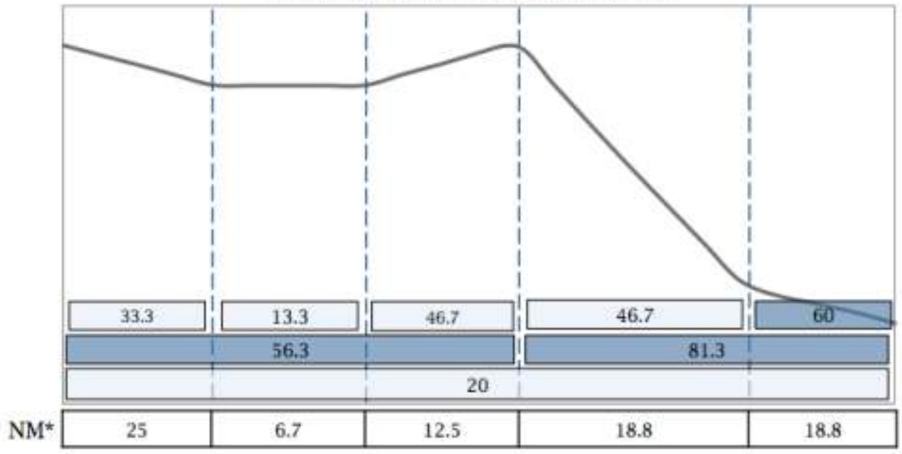


Haptic graphs without data labels

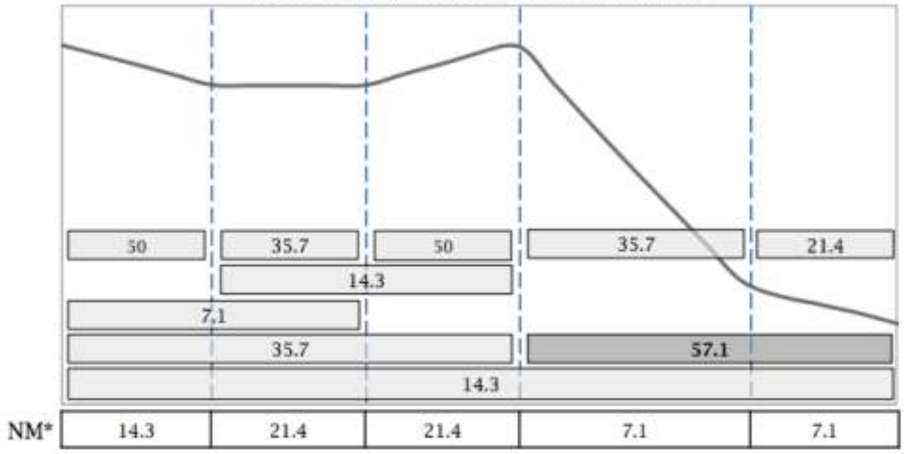


Graph-10

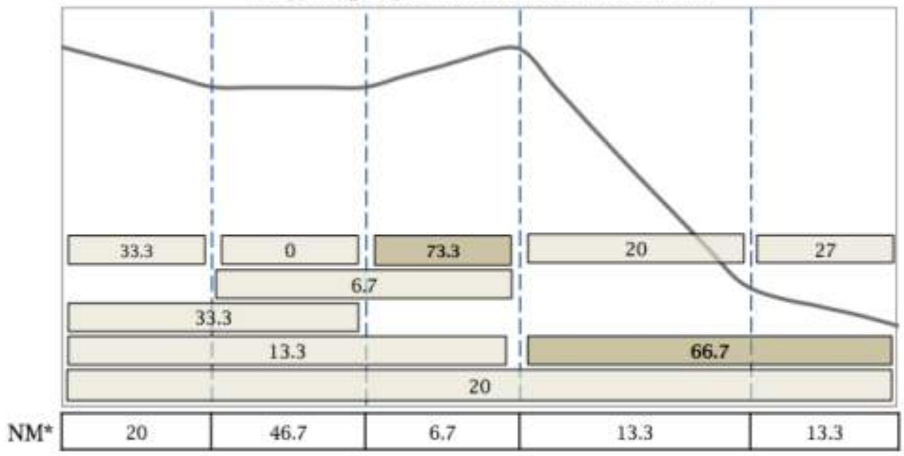
Visual graphs with data labels



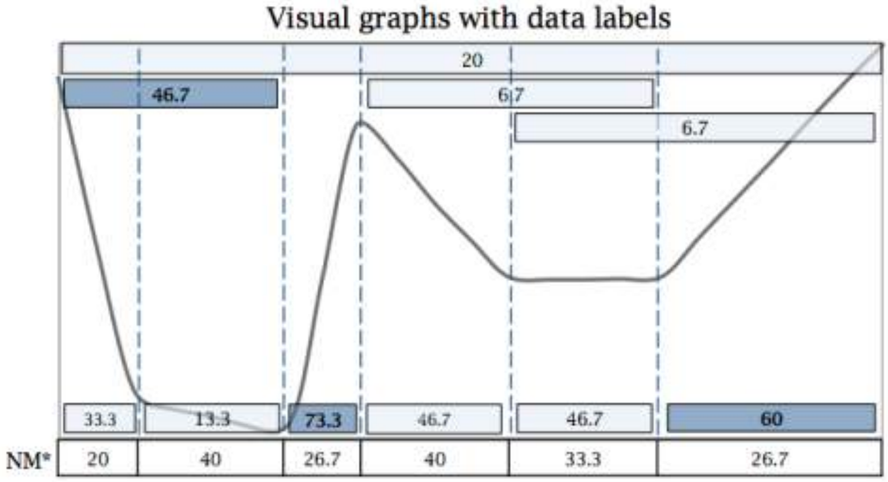
Visual graphs without data labels

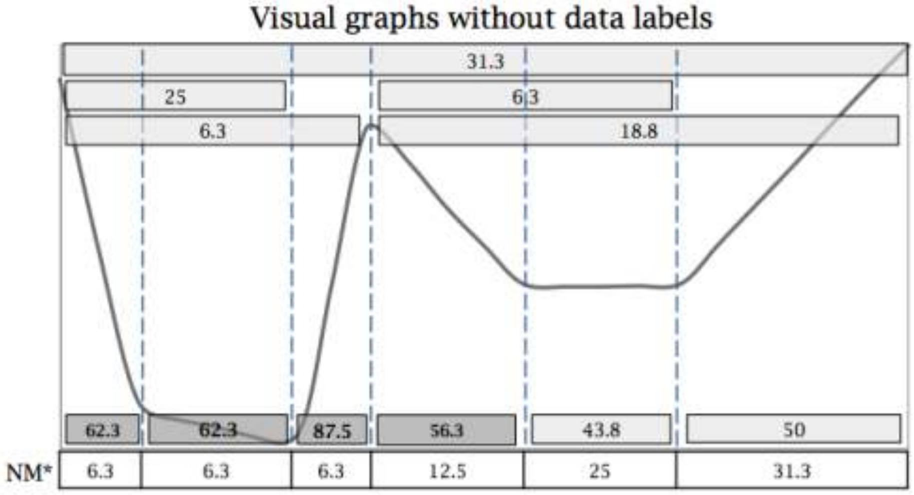


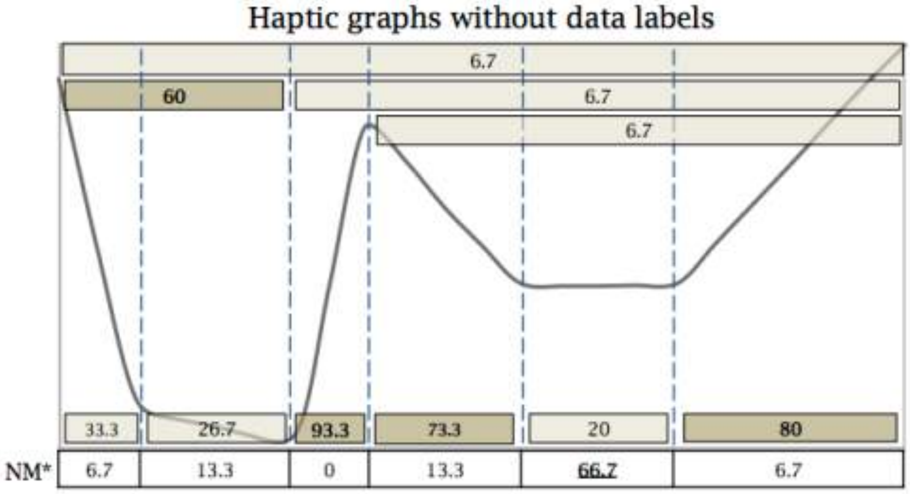
Haptic graphs without data labels



Graph-11

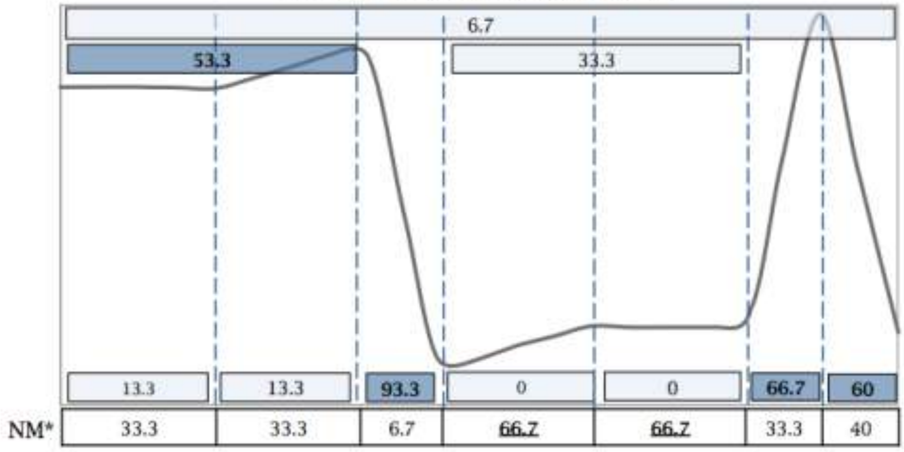

Visual graphs with data labels
20
46.7
6.7
6.7
33.3
13.3
73.3
46.7
46.7
60
NM*
20
40
26.7
40
33.3
26.7


Visual graphs without data labels
31.3
25
6.3
6.3
18.8
62.3
62.3
87.5
56.3
43.8
50
NM*
6.3
6.3
6.3
12.5
25
31.3

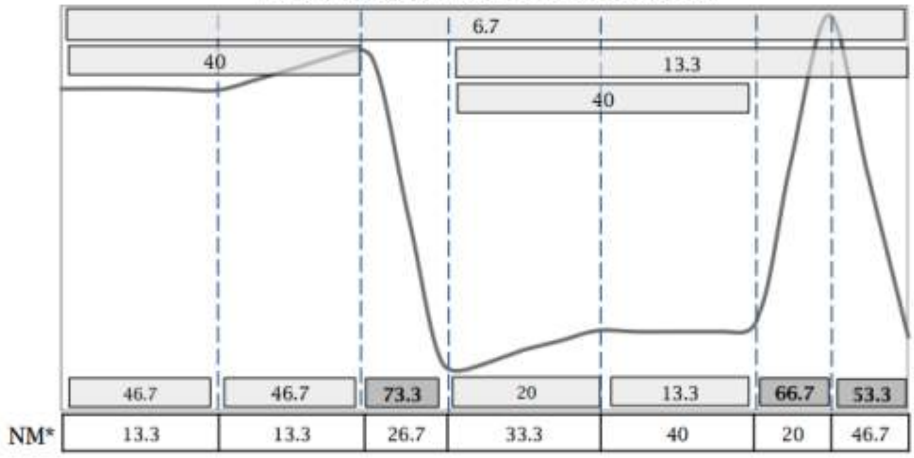

Haptic graphs without data labels
6.7
60
6.7
6.7
33.3
26.7
93.3
73.3
20
80
NM*
6.7
13.3
0
13.3
66.7
6.7

Graph-12

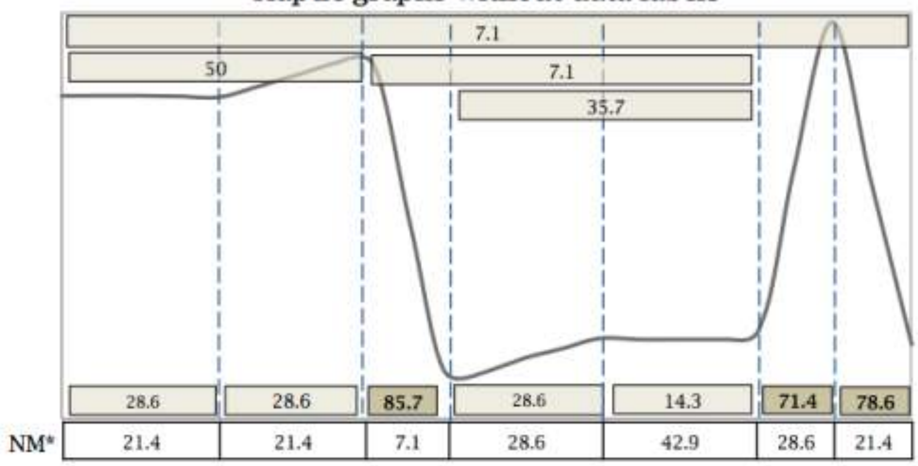
Visual graphs with data labels



Visual graphs without data labels



Haptic graphs without data labels



APPENDIX G: Frequency Tables for Time and Event Denoting Expressions

Time-Denoting Expressions (Table-2)												
	Graph-1			Graph-2			Graph-3			Graph-4		
	A	B	C	A	B	C	A	B	C	A	B	C
Total # of Time-Denoting Expression	59	68	46	51	65	57	38	61	74	64	84	83
Reference to seasons												
Summer	1	0	0	2	1	5	1	0	4	4	1	3
Spring	1	0	1	0	1	2	0	0	2	4	0	2
Winter	2	0	1	0	1	3	1	1	1	0	2	2
Fall	0	1	0	0	0	1	2	1	2	0	2	59
Total	4	1	2	2	2	9	4	1	9	8	4	8
Reference to months (explicit)												
Jan.	7	1	1	7	3	2	5	1	4	7	2	3
Feb.	10	3	0	3	0	1	2	0	1	9	4	1
Mar.	1	0	2	4	1	2	3	0	1	9	4	2
Apr.	6	2	0	0	1	0	5	0	5	5	1	0
May	4	1	0	0	0	0	10	0	0	1	0	2
Jun.	1	0	1	4	1	3	3	0	2	5	1	2
Jul.	0	0	0	2	1	0	2	0	0	3	0	1
Aug.	1	0	1	14	1	0	0	0	3	1	1	0
Sep.	1	0	1	2	0	4	5	0	0	3	2	3
Oct.	0	0	0	0	0	0	0	1	0	2	1	1
Nov.	0	0	1	2	1	1	0	0	0	0	1	0
Dec.	3	2	1	2	1	2	4	1	5	0	0	3
Total	28	8	6	32	9	10	23	3	16	27	10	11
Reference to general trend	2	2	0	1	2	1	5	10	6	1	0	0
Spatial Reference	0	3	5	0	5	3	1	2	4	0	1	2
Reference End points												
Start point	1	1	1	1	5	5	0	1	2	0	0	0
End Point	2	4	2	3	2	4	3	4	6	1	3	2
Total	3	5	3	4	7	9	3	5	8	1	3	2
Other vague descriptions												
around	0	0	0	1	3	2	0	0	1	0	0	0
for some time	1	6	3	0	2	0	1	3	3	0	6	6
then	15	13	12	8	21	6	7	7	10	13	32	14
between, during, within	3	3	0	1	0	1	0	0	0	1	0	0
towards, till	7	3	5	10	5	8	2	2	3	2	2	10
from to	0	0	1	2	0	0	3	0	3	0	0	0
No temporal Tag	8	20	10	6	15	19	3	15	20	14	26	22

A: Visual graphs with data labels

B: Visual Graphs without data labels

C: Haptic graphs without data labels

Time-Denoting Expressions (Table-2)												
	Graph-5			**Graph-6**			**Graph-7**			**Graph-8**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Total # of Time-Denoting Expression	54	71	69	56	66	70	71	73	76	60	66	50
Reference to seasons												
Summer	5	0	4	2	1	4	1	0	2	1	0	2
Spring	0	0	3	1	1	1	0	1	2	2	0	4
Winter	0	1	2	1	1	2	0	0	3	0	0	2
Fall	0	1	23	0	1	55	0	0	23	1	0	2
Total	5	1	11	3	2	8	1	1	10	3	0	7
Reference to months (explicit)												
Jan.	8	1	3	10	2	6	7	2	1	8	3	3
Feb.	12	3	3	1	0	1	5	2	1	2	2	0
Mar.	2	0	0	2	2	4	3	1	3	5	4	2
Apr.	0	0	2	0	0	1	1	2	4	1	0	0
May	4	0	0	1	2	0	1	4	1	8	0	0
Jun.	0	3	4	0	4	1	1	1	1	0	1	0
Jul.	9	1	0	22	1	3	0	0	2	4	1	0
Aug.	4	0	1	5	1	2	4	1	2	4	1	2
Sep.	3	1	1	0	2	0	3	1	3	0	2	0
Oct.	2	1	1	0	2	0	8	3	1	0	1	0
Nov.	0	1	0	0	0	0	15	2	1	0	1	0
Dec.	3	0	2	3	2	3	5	3	2	8	3	3
Total	36	8	12	37	13	16	41	15	12	28	12	8
Reference to general trend	1	1	0	0	1	2	2	0	1	6	4	4
Spatial Reference	0	0	4	1	2	3	0	2	3	0	0	0
Reference End points												
Start point	0	0	0	0	0	0	0	0	0	0	0	0
End Point	0	1	4	2	1	9	0	3	3	2	1	3
Total	0	1	4	2	1	9	0	3	3	2	1	3
Other vague descriptions												
around	1	2	0	2	0	1	0	0	0	0	0	0
for some time	1	3	4	0	1	2	1	0	1	6	3	3
then	10	32	21	9	20	14	14	28	18	6	22	13
between, during, within	0	2	0	1	0	1	0	0	2	0	0	0
towards, till	4	6	5	8	7	9	3	7	5	6	2	3
from to	6	0	0	6	0	5	1	0	4	2	0	1
No temporal Tag	6	13	21	4	17	20	12	22	19	5	17	12

Time-Denoting Expressions (Table-3)												
	Graph-9			**Graph-10**			**Graph-11**			**Graph-12**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Total # of Time-Denoting Expression	54	67	59	49	52	49	76	87	74	64	72	77
Reference to seasons												
Summer	0	0	2	0	0	2	2	1	3	2	1	2
Spring	0	1	1	0	0	1	2	0	2	0	1	2
Winter	0	0	1	1	1	2	5	1	1	1	2	1
Fall	0	0	1	0	0	0	0	0	68	0	0	85
Total	0	1	5	1	1	5	9	1	7	3	4	8
Reference to months (explicit)												
Jan.	11	1	3	8	1	2	10	2	3	9	1	4
Feb.	4	2	2	2	0	0	7	1	1	1	1	0
Mar.	11	2	1	3	1	1	4	4	4	3	2	2
Apr.	15	6	2	2	1	0	3	2	2	1	1	0
May	3	1	0	2	0	1	6	0	0	11	1	3
Jun.	3	1	1	1	1	0	4	2	2	7	0	1
Jul.	1	0	0	14	1	2	4	0	3	4	1	1
Aug.	0	0	0	1	1	1	3	2	0	3	2	4
Sep.	1	2	0	1	1	0	2	2	3	3	1	2
Oct.	1	1	2	3	0	0	1	4	4	4	1	2
Nov.	1	1	1	0	1	0	0	2	1	6	1	1
Dec.	3	4	5	3	1	2	6	4	2	3	4	5
Total	37	17	14	31	5	8	35	16	15	40	13	19
Reference to general trend	0	1	1	1	3	4	2	2	0	0	0	2
Spatial Reference	1	2	3	1	1	3	0	0	1	0	0	0
Reference End points												
Start point	0	1	0	0	2	0	1	1	1	0	0	0
End Point	2	1	2	4	3	1	4	1	4	1	2	7
Total	2	2	2	4	5	1	5	2	5	1	2	7
Other vague descriptions												
around	0	1	0	0	0	3	1	1	0	1	0	1
for some time	0	0	1	0	2	4	0	4	7	1	4	2
then	10	20	24	5	13	12	12	21	21	13	26	25
between, during, within	0	0	1	0	0	0	0	3	0	5	1	0
towards, till	4	6	7	5	4	3	6	4	3	2	3	4
from to	3	0	0	3	2	0	1	0	3	3	1	1
No temporal Tag	5	19	9	10	15	11	16	28	12	8	15	23

Basic Event-Denoting Expressions (Table-1)												
	Graph-1			**Graph-2**			**Graph-3**			**Graph-4**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Basic												
State	12	24	18	15	16	10	18	33	36	16	32	28
Act	51	54	42	38	58	55	28	41	59	47	72	73
Manner	26	31	19	16	28	15	13	11	12	9	38	14
Value												
Value To	2	7	6	5	7	5	5	5	2	10	9	12
Value from	1	2	2	2	4	0	1	1	0	0	1	1
only value	15	13	5	12	11	7	5	12	10	23	16	14
Total	19	26	19	23	26	13	12	23	18	38	32	31
Shape	0	1	1	2	1	0	3	3	0	3	0	5
No additional Attributes	12	14	6	10	10	18	14	16	26	13	21	28

Basic Event-Denoting Expressions (Table-2)												
	Graph-5			**Graph-6**			**Graph-7**			**Graph-8**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Basic												
State	18	30	26	17	20	24	31	37	29	25	29	26
Act	40	54	65	43	64	62	56	55	75	39	52	44
Manner	17	35	18	19	29	29	26	37	26	18	20	17
Value												
Value To	8	8	3	10	7	11	12	8	6	5	6	6
Value from	1	4	2	1	2	1	1	1	0	0	0	1
only value	17	10	4	14	16	13	8	14	8	12	16	3
Total	30	24	10	28	32	35	22	23	18	20	24	12
Shape	2	3	3	0	0	1	0	4	1	6	1	0
No additional Attributes	9	14	24	8	15	18	15	15	31	13	10	10

Basic Event-Denoting Expressions (Table-3)												
	Graph-9			**Graph-10**			**Graph-11**			**Graph-12**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Basic												
State	16	25	23	18	21	29	20	33	27	15	34	21
Act	45	60	50	40	40	40	65	69	62	51	54	66
Manner	15	21	19	16	17	18	14	35	16	16	24	26
Value												
Value To	11	13	6	5	4	4	15	15	5	18	7	7
Value from	1	2	0	0	2	1	2	1	1	0	0	0
only value	19	14	8	14	13	4	15	21	14	14	19	11
Total	34	35	19	20	25	11	33	45	27	36	35	22
Shape	2	6	1	0	3	1	2	2	1	0	0	1
No additional Attributes	8	8	17	4	8	14	26	18	28	8	12	22

Verbs and nouns (Table-1)												
	Graph-1			**Graph-2**			**Graph-3**			**Graph-4**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
increase (rise, increase, incline etc.)	13	19	14	17	29	26	16	23	31	16	22	32
decrease	17	18	15	14	16	16	3	14	14	17	23	24
up & downs	0	3	0	0	1	0	4	4	3	0	1	0
Remain stable												
stay	8	6	2	2	3	4	3	0	3	2	10	9
continue	14	11	3	1	4	2	2	6	5	7	15	2
get stable	0	0	0	0	0	0	0	0	0	0	0	2
go straight	0	0	0	0	0	0	0	0	0	0	0	0
Total	22	17	5	3	7	6	5	6	8	9	25	13
reach	3	3	5	1	4	4	4	4	2	4	4	4
come	0	0	0	0	0	0	0	0	0	0	0	0
do/have	1	4	1	1	0	0	2	5	8	8	8	6
start	5	9	7	2	4	3	4	7	3	4	12	8
end	0	3	1	4	3	1	1	2	6	3	2	5
Curves												
be curved	0	0	1	0	0	0	0	0	0	0	0	0
jump	0	0	0	0	0	0	2	0	0	0	0	0
do a peak	0	0	0	0	1	1	0	0	2	0	0	0
turn	0	0	0	0	0	0	0	0	0	0	0	2
TOTAL	0	0	1	0	1	1	2	0	2	0	0	2
Steep												
get steep	0	0	1	0	0	0	0	0	0	0	0	0
get sharp	0	0	0	0	0	0	0	0	0	0	0	0
get strong	0	0	0	0	0	0	0	0	0	0	0	0
total	0	0	1	0	0	0	0	0	0	0	0	0
Other verbs												
draw	0	0	0	0	0	0	0	0	0	0	0	0
slow down	0	0	0	0	0	0	0	0	0	0	0	0
compensate	0	0	0	0	0	0	0	0	0	0	0	0
recover	0	0	0	0	0	0	0	0	0	0	0	0
Others												
Relation	0	2	7	4	3	7	1	4	9	3	12	8
Negation	3	5	4	1	0	3	1	0	3	2	1	5

Verbs and nouns (Table-2)												
	Graph-5			Graph-6			Graph-7			Graph-8		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
increase (rise, increase, incline etc.)	18	23	27	11	18	18	32	32	33	8	10	5
decrease	14	29	27	24	25	26	19	19	25	27	30	24
up & downs	0	1	2	0	0	1	2	6	0	0	1	1
Remain stable												
stay	0	1	0	0	1	2	1	1	1	2	1	0
continue	5	3	6	5	7	5	5	6	1	3	11	9
get stable	0	1	6	0	3	4	0	1	4	1	0	1
go straight	0	0	0	0	0	0	0	0	0	0	0	0
Total	5	5	10	5	11	11	6	8	6	6	12	10
reach	4	2	2	4	7	3	4	3	5	2	2	1
come	0	0	0	0	0	0	0	0	0	0	0	0
do/have	5	8	5	8	7	6	10	10	9	7	9	5
start	5	10	11	6	12	9	6	5	11	6	10	8
end	1	3	3	1	0	2	1	3	1	2	1	4
Curves												
be curved	0	0	0	0	0	1	0	0	1	0	0	0
jump	0	0	0	0	0	0	0	1	0	0	0	0
do a peak	0	0	0	0	0	0	0	0	0	0	0	0
turn	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	0	0	0	0	0	1	0	1	1	0	0	0
Steep												
get steep	0	0	0	0	0	0	0	0	1	0	0	0
get sharp	0	0	0	0	0	0	0	0	0	0	0	0
get strong	0	0	0	0	0	0	0	0	0	0	0	0
total	0	0	0	0	0	0	0	0	1	0	0	0
Other verbs												
draw	0	0	0	0	0	1	0	0	0	1	0	0
slow down	0	0	0	0	0	2	0	0	0	0	0	0
compensate	0	0	0	0	0	0	0	0	0	0	0	3
recover	0	0	0	0	1	0	0	0	0	0	0	0
Others												
Relation	5	11	7	0	6	7	0	2	4	2	5	3
Negation	2	5	3	0	0	3	0	2	1	1	1	3

Verbs and nouns (Table-3)												
	Graph-9			Graph-10			Graph-11			Graph-12		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
increase (rise, increase, incline etc.)	11	19	17	4	7	13	25	21	26	14	23	20
decrease	17	27	19	24	24	24	20	30	26	22	23	29
up & downs	0	1	0	1	0	1	0	0	1	0	0	0
Remain stable												
stay	1	0	1	0	2	0	0	5	7	4	1	4
continue	6	3	9	9	6	8	7	6	0	8	15	12
get stable	0	3	6	0	0	0	0	2	0	0	0	0
go straight	0	0	0	0	0	0	0	0	0	0	0	0
Total	7	6	15	9	8	8	7	13	7	12	16	16
reach	5	4	1	3	2	1	8	4	3	2	1	3
come	0	0	0	0	0	0	0	0	0	0	0	1
do/have	8	11	4	5	6	4	6	10	7	4	13	4
start	7	12	10	7	9	9	10	14	11	4	9	8
end	1	3	3	2	0	5	2	1	2	0	2	1
Curves												
be curved	0	0	1	0	0	1	0	0	1	0	0	0
jump	0	0	0	0	0	0	0	0	0	0	0	0
do a peak	0	0	0	0	0	0	0	0	0	0	2	0
turn	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	0	0	1	0	0	1	0	0	1	0	2	0
Steep												
get steep	0	0	0	0	0	0	0	0	0	0	0	0
get sharp	0	0	0	0	0	0	0	0	0	0	0	1
get strong	0	0	0	0	1	0	0	0	0	0	0	0
total	0	0	0	0	1	0	0	0	0	0	0	1
Other verbs												
draw	0	0	0	0	0	0	0	0	0	0	0	0
slow down	0	0	0	0	0	0	0	0	0	0	0	0
compensate	0	0	0	0	0	0	0	0	0	0	0	0
recover	0	0	0	0	1	0	0	2	0	0	0	0
Others												
Relation	2	5	5	2	3	7	1	7	5	2	1	2
Negation	1	5	1	2	0	3	1	4	3	1	2	0