

**Timing in a multisensory context with music and movements:**

**In search of an internal clock**

Dissertation zur Erlangung des Grades der Doktorin der Philosophie (Dr. phil.)

an der Fakultät für Geisteswissenschaften der Universität Hamburg

Dissertation to obtain the degree of Doctor of Philosophy (Dr. Phil.)

at the Faculty of Humanities of the University of Hamburg

Vorgelegt von/Presented by

Xinyue Wang

Hamburg, 2022

**Author** Xinyue Wang  
Department of Systematic Musicology  
University of Hamburg

**Supervisor** Prof. Dr. Clemens Wöllner  
Department of Systematic Musicology  
University of Hamburg

**Examiner** Prof. Dr. Zhuanghua Shi  
Department of Psychology  
Ludwig-Maximilian-University of Munich

**Examiner** Dr. Nicholas Ruth  
Department of Systematic Musicology  
University of Hamburg

**Date of Disputation** August 9<sup>th</sup>, 2022

## **Acknowledgment**

I am really grateful for the support of my supervisor, Prof. Dr. Clemens Wöllner. He has provided me with generous, unsparing guidance on how to conduct scientific studies with integrity. Despite the difficulty of the COVID-19 pandemic, I have always felt supported in completing my academic training.

I am also grateful for the help of Dr. Birgitta Burger and Prof. Dr. Zhuanghua Shi, co-authors and mentors on the research projects. They have guided me on scientific thinking and persistence in times of challenges. I want to thank Dominik Leiner and Sebastian Schwarz for their support in implementing the SoSci Survey experiments. A huge thank you to my study participants from all over the world. You have contributed tremendously to the projects.

I couldn't have achieved the goals without the valuable comradeship of Emma Allingham, David Hammerschmidt, Frithjof Faasch, Mia Kuch, and Dr. Nicholas Ruth. Thank you for the constructive discussions that inspired many creative ideas.

I thank my dear friends Jiaxin, Liudmila, Xiangming, and Eunice for being my heroes and showing me what potential life could bear. I also express my deepest gratitude to my parents for supporting me with the weekly calls, even though the pandemic has separated us for over two years. Finally, thank you to Adrian for inspiring me with your never-ending love and support.

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>7</b>
1.1	Timing theories .....	7
1.2	The effects of sensory modalities .....	9
1.3	Tempo manipulations affect time perception .....	11
1.4	The effects of emotions.....	12
1.5	Sensorimotor synchronization .....	13
<b>2</b>	<b>Aims and objectives.....</b>	<b>13</b>
<b>3</b>	<b>Methods .....</b>	<b>14</b>
3.1	Study 1 .....	14
3.2	Study 2 .....	14
3.2.1	Participants.....	14
3.2.2	Apparatus .....	15
3.2.3	Procedure .....	15
3.2.4	Analyses.....	15
3.3	Study 3 .....	16
3.3.1	Participants.....	16
3.3.2	Apparatus .....	16
3.3.3	Procedure .....	16
3.3.4	Analyses.....	17
3.4	Study 4 .....	17
3.4.1	Participants.....	17
3.4.2	Apparatus .....	17
3.4.3	Procedure .....	17
3.4.4	Analyses.....	18
<b>4</b>	<b>Main results.....</b>	<b>19</b>
4.1	Study 1 .....	19
4.2	Study 2 .....	19

4.2.1	Visual versus auditory tempo.....	19
4.2.2	Individual modality reliance .....	20
4.3	Study 3 .....	20
4.3.1	Effects of tempo and modality .....	20
4.3.2	Contributions of movement features changes.....	20
4.4	Study 4 .....	21
4.4.1	Influence of tapping versus no-tapping.....	21
4.4.2	Influence of the tapping speed and stability .....	21
<b>5</b>	<b>Discussion.....</b>	<b>22</b>
5.1	Visual driving effect .....	22
5.2	Information specificity .....	24
5.3	Effects of visual vs. audiovisual stimuli on perceived durations.....	24
5.4	Tempo manipulations: Absence of temporal effects .....	25
5.5	Tapping changes the subjective time .....	26
<b>6</b>	<b>Limitations and future directions .....</b>	<b>28</b>
<b>7</b>	<b>Conclusion.....</b>	<b>29</b>
<b>8</b>	<b>References .....</b>	<b>31</b>
	<b>Summary (English).....</b>	<b>39</b>
	<b>Zusammenfassung (Deutsch).....</b>	<b>41</b>
	<b>List of Publications and Author Contributions.....</b>	<b>44</b>
	<b>Statutory Declaration/ Eidesstattliche Erklärung.....</b>	<b>45</b>
	<b>Study 1.....</b>	<b>46</b>
	<b>Study 2.....</b>	<b>69</b>
	<b>Study 3.....</b>	<b>91</b>
	<b>Study 4.....</b>	<b>95</b>
	<b>Appendix .....</b>	<b>151</b>



# 1 Introduction

The perception of time contributes profoundly to the construction of how we perceive the world. We move by specific tempi and estimate the time at which to react to the trajectory of a ball. Time perception holds an even more prominent role when listening and producing music. As Berger (2014) has observed, in Schubert's *String Quintet in C major, D.956*, the context of fast and lively rhythmic patterns in slow and still contexts led to distortions in the perceived durations. As complex auditory and sometimes audiovisual signals, music provides the ideal playground for manipulating time perception. According to Nichols (2011), the French composer Ravel also observed similar distortions in the "felt" (subjective) time when *Boléro* was conducted at different speeds. The investigation into how we perceive time is expected to contribute to clarifying the mechanisms that act as the foundation of human timing performances; it also provides insights into one's time experience in real-life scenarios such as concerts, sports, or games.

To answer this question, the dissertation intends to evaluate how subjective time is experienced. Time perception has no direct link to a sensory organ yet is facilitated by all sensory modalities. Interestingly, evidence suggested that the perceived durations varied by the sensory modality in which they were presented (Appelqvist-Dalton et al., 2022; Mayer et al., 2014). When the drummers' movement in performing a long note was combined with the soundtrack of a short note, participants appeared to judge the sound longer than the objective durations (Schutz & Lipscomb, 2007). Furthermore, non-musical attributes of the visual presentation, as an integral part of piano recitals, affected the overall performance quality judgments (Wapnick et al., 2009). Having these examples in mind, the modalities in which time is perceived have inevitably affected one's temporal judgments. The current dissertation aims to determine the internal clock model adopted within the seconds (supra-second) range by examining the modality-(in)dependent timing performances with ecologically valid stimuli. Furthermore, it aims to set a foundation for the optimal clock model that integrates cognitive engagements and modality-specific timing for future studies.

## 1.1 Timing theories

The theories surrounding human timing mechanisms have seen two main streams of internal clock models: The intrinsic (distributed) and the central (dedicated) model (Wittmann, 2013). The intrinsic clock model assumes that there is no such cognitive module as a timer independent of sensory processing; instead, time perception reflects the implicit function that

varies within neural networks by different sensory modalities (Buonomano & Maass, 2009). The central clock model, on the other hand, postulates a universal, independent timer dedicated to the human timing function (Treisman, 1963). The pacemaker-counter device that construes the basis of the central clock model hypothesizes that time perception is formulated as the following: The temporal pulses are emitted by the pacemaker, recorded and compared to the reference duration at the counter device before finally being judged as the subjective duration (Allman & Meck, 2012; Gibbon, 1977). As an extension of the pacemaker-counter theory, the Attentional Gate Model (Zakay & Block, 1995) emphasizes the role of attention by hypothesizing that recording the temporal units is equivalent to them passing by the gate of attention: When the “gate” opens wider (more attention), more units are recorded. Consequently, the duration is perceived to be longer, and vice versa. The intrinsic model warrants discrepancies among timing performances across modalities, whereas the central clock model supports consistent performances. Considering findings on both models, Study 1 as a review was carried out to systematically investigate the theoretical base of the two timing mechanisms and the modality-relevant evidence that supported or challenged them.

It should be noted that the thesis intends to discuss specifically time perception within the seconds-to-minutes (supra-second) range. The duration-dependent nature of a variety of timing mechanisms has been proposed by earlier research (for reviews, see Buhusi & Meck, 2005; Wittmann, 2013). The intrinsic clock model has been associated with millisecond (sub-second) range timing where sensory processing could be directly involved (Burr et al., 2009; Motala et al., 2018), whereas the central clock model has been linked to seconds-to-minutes (supra-second) range timing, assuming that cognitive processes such as the working memory (WM) and attention come into play as the pacemaker-counter mechanism postulates a comparison of the attended temporal units to the reference duration stored in WM (Wittmann, 2013). Nevertheless, a number of studies have found modality-related disparities in human timing performances in the supra-second range (e.g. Escoffier et al., 2010; Warm, Stutz, & Vassolo, 1975; for a review, see Ivry & Schlerf, 2008). The findings led to the question of whether the intrinsic clock model could explain and predict the supra-second range timing behaviors. In this vein, the studies of this project set out to investigate the role of modality in tempo judgments in this timescale.



## 1.2 The effects of sensory modalities

Examining the evidence for both clock models with multisensory stimuli as inputs sheds light on their validity. The effects of modalities in time perception, when present, indicate that the timing mechanism is distributed by separate neural networks dedicated to different sensory processing; when not present, suggest that it functions independently and affects timing performances across modalities consistently. Evidence that supports the intrinsic clock model highlights the modality specificity (for a review, see Wang & Wöllner, 2019). As the review suggested, audition has been identified in several studies as the dominant modality over vision in time perception tasks. For instance, The temporal ventriloquism effect indicated that when perceiving a series of auditory and visual stimuli, the temporal location of visual inputs was affected by the temporal location of the auditory ones (Burr et al., 2009). Similarly, the auditory driving effect (Shipley, 1964) has revealed that when playing simultaneously, the flutter rate of sounds lowered the perceived rate of visual flickers. In contrast, visual inputs did not have similar effects on the auditory inputs in either study. The advantage of audition over vision in time perception tasks was also observed in rhythmic pattern identification (Guttman et al., 2005), duration estimation (Ortega et al., 2014), and tempo (Chen et al., 2018).

Alternatively, some research proposed that the discrepancy between audition and vision was in the opposite direction: Vision dominates temporal processing over audition in certain circumstances. Schutz and Lipscomb (2007) found that when tempo-incongruent audiovisual presentations of marimba performances were presented to viewers, the tempo judgments of drumming sounds (auditory inputs) were biased by the musician's gestures (visual inputs) rather than auditory tempo. The finding, coined as the Schutz-Lipscomb effect, has been replicated with duration estimation and the temporal-order judgment tasks among young (18 to 29 years old) and old (65 to 78 years old) groups (Bak et al., 2021). These studies support other findings where the effects of visual stimuli affected auditory stimuli in tasks such as tempo discrimination (Su & Jonikaitis, 2011) and timing tasks in the form of sensorimotor synchronization (Su, 2014). The above evidence indicated a modality effect on various time perception tasks was present in favor of either audition or vision, supporting a distributed clock model.

Ecological validity is an essential factor in the emerging role of vision against audition in time perception. As mentioned in the previous literature, past studies that found

auditory dominance in temporal judgments over vision have mainly adopted visual stimuli such as flashing lights (Burr et al., 2009; Repp & Penel, 2002) and Gabor patches that consisted of even grids (Guttman et al., 2005). The stimuli were highly abstract and artificial. According to Lewkowicz's (2001) arguments, ecologically valid stimuli should represent the relationship between experiment subjects and their perceptual experiences in the day-to-day environment. In this case, light dots and patterned patches did not fully portray the nature of visual encounters. Thus, findings of audition dominance in time perception cannot be entirely relied upon. Instead, examples of visual outweighing the auditory effects in temporal processing have seen the usage of actual music performances (e.g. Schutz & Lipscomb, 2007), which could resemble the daily perceptual experience to participants better than artificial stimuli.

Attempts have been made to follow up on the nature of visual inputs in time perception. For example, a comparison between finger tapping, moving bars, and flashes suggested that finger tapping elicited higher accuracy in visuomotor synchronization than flashes (Hove et al., 2010), indicating that environment-compatible visual inputs provided more information for temporal judgments than incompatible ones. Taking a step further, a study comparing an auditory distractor (pure tone metronome) with a bouncing ball as a visual distractor in a sensorimotor synchronization task has found no main effects of either modality (Hove et al., 2013). Apart from finger tapping and ball bouncing, ecologically valid visual stimuli also include human movements in naturalistic scenes (e.g. Boltz, 2005) and point-light displays (PLDs) (Johansson, 1973), which offer temporal cues that are common and compatible with our daily perceptual experiences. Boltz (2005) has found no evidence of auditory dominance in duration reproduction tasks when presenting participants with naturalistic scenes in auditory and visual modes. When adopting PLDs as visual stimuli, Grahn (2012) found that beat-based structures similarly enhanced participants' performances in rhythm discrimination with audition and vision. Overall, the findings did not support an auditory dominance in the presence of ecologically valid visual stimuli.

Parallel to the modality-specific evidence, past research has also revealed inter-related timing performances across sensory modalities as transferrable training effects. In a study where a duration discrimination task was adopted, participants exhibited better performances with auditory and visual stimuli after training sessions with auditory sequences (Bratzke et al., 2012). Similarly, participants synchronizing with musical excerpts while responding to visual targets placed in or out-of-phase with the auditory rhythmic patterns showed that the

in-phase visual targets were discriminated with higher accuracy (Bolger et al., 2013). Bolger et al.'s (2013) study aligns with earlier research that, even when not attending to the background of the piano music, participants responded faster to the visual stimuli than in the silent condition (Escoffier et al., 2010). The studies revealed that the cross-modal transfer of training effect followed the order from audition to vision but not from vision to audition. In this case, audition appeared to dominate the temporal processing, thus supporting the intrinsic clock model. Furthermore, there is also evidence of no cross-modal (audition-vision) transfer with duration reproduction tasks (Motala et al., 2018).

Despite the asymmetry in the evidence that favors the intrinsic model, no conclusive evidence has been found on which internal clock model is more applicable in human timing scenarios. The uncertainty has, therefore, called for an investigation comparing performances in time perception tasks with different sensory modalities to provide more evidence of the validity of the two internal clock models. To answer the questions, Study 2 in this thesis investigated participants' modality preferences in tempo judgment tasks with audiovisual tempo-incongruent stimuli.

### **1.3 Tempo manipulations affect time perception**

The components of tempo, apart from event frequency, also affect the internal clock speed. London (2011) has asked a question in an earlier investigation of whether "beat rate itself is a transparent measure of musical speed" (p. 44). According to this study, tempo is a multifaceted concept that, apart from the beat rate itself, encompasses rhythmic structures, metrical levels, and salience of the event segmentations. Similar findings with visual stimuli (moving patches) suggested that visual tempo had multiple components, including the segmentation salience (Verghese & Stone, 1996), besides event frequency alone. It may be inferred that the internal clock speed is under the influence of more than the apparent tempo of auditory and visual events. As previously discussed, visual tempo could affect time perception more strongly than auditory tempo, given that the inputs are highly plausible to people's perceptual experiences. This possibility, therefore, gives rise to the exploration in Study 3 of how tempo manipulations, with the same presentation beat rates and different kinematic features, might affect human timing performances in the visual aspect. This might provide us with insights into the potential mechanism of the distributed clock model and, thus, the search for an internal clock.

Past research has seen the effects of tempo acceleration and deceleration on subjective time. One study, for instance, showed that slow-motion clips of movies, ballet performances, and sports scenes were perceived as shorter than the videos adapted to real-time speed (Wöllner et al., 2018), suggesting that the internal clock speed was synchronized to the decelerated scenarios. As research showed, tempo changes were associated with the movement features via the emotional states; for example, happy movements with high movements were often found with fast movement velocity (de Meijer, 1989). It points to the possibility that tempo could be composed of multiple facets. Subsequently, changes in visual features such as movement complexity and fluidity in tempo-manipulated inputs could affect the perception of time in addition to the beat rates the PLDs followed. Previous studies have also implied that such visual feature changes (e.g. Aubry et al., 2008) led to variations in subjective time. In Study 3, investigating the roles of these visual features in temporal judgments could help us to further our understanding of which components are particularly important to entrain the internal clock speed in an ecologically valid visual context.

#### **1.4 The effects of emotions**

Apart from sensory modalities, emotions share a close bond with time perception by exerting extensive influences on the perceived time and reflecting the temporal attributes of audiovisual inputs. Emotional arousal has been linked to the pace of the internal clock by many previous studies. According to the pacemaker-counter mechanism, higher emotional arousal increases the internal clock speed by higher emission of temporal units, thus leading to longer perceived duration (for a review, see Grondin, 2010). Emotional contents, nevertheless, demand attentional resources. As a result, the temporal pulses are under-recorded in the counter device, leading to reduced accumulation of the internal clock “ticks” that consequently translate to duration underestimation. Emotional music and sine waves neutral sounds were judged to be similar in arousal level, while the musical stimuli were underestimated in duration compared to neutral sounds (Droit-Volet et al., 2010). In this respect, participants may have diverted attention to the stimuli rather than keeping the time, therefore reporting the time shorter than with the neutral sounds.

Furthermore, viewing pictures of liked (high arousal) and disliked (low arousal) food also elicited changes in duration estimation: Pictures of the preferred food were more likely to be judged longer than the ones disliked (Gil et al., 2009). Interestingly, when valence and arousal were aligned, e.g. low valence and high arousal, the latter had stronger influences on

duration overestimation (Angrilli et al., 1997). Some argued that emotional valence only affects the perception of time to a small extent, as researchers have found with experiments where participants underestimated durations when listening to pleasant compared to unpleasant sounds (Droit-Volet et al., 2013). Similar to the effect of arousal, the allocation of attention to the timing processing was also assumed to play a central role in the influences of emotional valence on subjective time.

### **1.5 Sensorimotor synchronization**

The previous studies (Study 1, 2, and 3) aimed at tackling the disputes over the presence of modality specificity in time perception, while Study 4 set out to understand how the internal clock model functions in a highly ecologically valid context where motor involvement comes into play. Instead of passively perceiving the inputs, sensorimotor synchronization with auditory (Eerola et al., 2006) and visual beats (Huang et al., 2018) is common across age groups and is in various forms, such as tapping, walking, and free full-body movements. Tapping to external rhythms has been found to lead to duration underestimation (Hammerschmidt & Wöllner, 2020) and faster passage of time (Wöllner & Hammerschmidt, 2021) by distracting the attentional resources allocated to the timing tasks. Attention has been a key concept to the dedicated clock model (Block, 2003) and the central clock model, i.e., the pacemaker-counter theory (Treisman, 1963) where attention acted as a gatekeeper to the temporal units recorded. A consensus of both models suggested that, as attentional resources are limited, attending to tasks other than time judgment tasks would lead to less accurate estimation of durations, whereas attending to time-related tasks resulted in higher temporal sensitivity and longer perceived time (for a review, see Grondin, 2010). Altogether, investigating the effects of sensorimotor synchronization on time perception as a type of proactive timing could shed light on our internal clock mechanisms in relation to attention and the involvement of sensorimotor synchronization.

## **2 Aims and objectives**

This dissertation project aimed to look for evidence for an internal clock that serves as a human timing mechanism. In Study 1, the theoretical background was to be examined thoroughly. Past internal clock models and their evidence were compared on behavioral and neurological levels. Recent findings that support the intrinsic internal clock have been pointed out in particular. In Study 2, the study set out to look for evidence of modality-

specific discrepancies in timing performances. More specifically, participants were asked how they judged the durations of audio-visual incongruent stimuli compared to the reference durations. Having identified the advantages of visual against auditory timing, Study 3 aimed to examine the effects of variations in the visual cues on the perception of time. To identify the timing accuracy with uni- and bi-modal inputs, duration judgments with visual and audiovisual presentations were compared in Study 4. The effects of variations in the kinematic features of a dancer's performance in tempo-shifted point-light displays on the perceived durations were investigated in addition to emotion ratings as the secondary impacts. Finally, the study extended the context of the previous findings to an ecologically valid audiovisual setting, where the effects of tapping to a drummer's performances that varied in tempo and complexities were looked into more closely.

## **3 Methods**

### **3.1 Study 1**

A narrative literature review was adopted as the main research method of this study. This format was intended to provide an overview of the main theories of human timing mechanisms with supporting evidence in different time ranges (sub-second, supra-second, second to minutes, and minutes to hours). An emphasis was placed on how the central versus intrinsic (distributed) clock theory shaped sensory-modality-specific research evidence. The review aimed to identify a research gap that allowed further studies to support either internal clock theory by identifying the consistency of timing performances across modalities (audition and vision).

### **3.2 Study 2**

#### **3.2.1 Participants**

The study was expected to follow the thread from Study 1 to verify whether the modality effect was present in the seconds-range timescale. For this purpose, 24 participants (12 females,  $M_{Age} = 24.21$  yrs,  $SD_{Age} = 4.68$ ) were recruited from the university campus and the Institute of Systematic Musicology at the University of Hamburg for the study. A prevalence of musical training was observed, as participants had, on average, 7.65 years of professional training and 10.04 years of practice. Recruitment of the participants was in line with the requirements of the Ethics Committee of the Faculty of Humanities at the University of Hamburg. Each participant was compensated 10 Euros for their participation.

### **3.2.2 Apparatus**

A human actor's movements were adopted to enhance the plausibility or ecological validity of the visual stimuli to our perceptual experiences. Isochronous drumbeats and point-light displays (PLDs) of a person jumping from left to right with hands moving up and down of 9 tempi (60 to 180BPM, 15BPM per step) were synthesized from recordings, respectively. Drumbeat excerpts and PLDs of each tempo were combined in Adobe Premiere Pro CC 2017 (Adobe Systems, San Jose, CA, USA) to produce the 5-second audiovisual stimuli that would be presented to the participants, leading to a total of 81 stimuli. Therefore, only 9 stimuli were tempo-congruent in both sensory modalities.

The PLDs were recorded with the 11-camera motion capture system (Qualisys Oqus, Qualisys AB, Göteborg, Sweden) at 200 frames per second. The MATLAB Motion Capture (MoCap) Toolbox (Burger & Toiviainen, 2013) was adopted to time-shift the PLD, originally recorded at 120 BPM, to the other 8 tempi. The drumbeats were created on Drumbit (<http://drumbit.app>) from a bass drum. The experiment was presented on a Dell U2414Hb monitor (Dell Technologies Inc., Round Rock, TX, USA) through the software OpenSesame (Mathôt et al., 2012). The soundtracks were played via a Sennheiser HD600 headphone set (Sennheiser, GmbH, Hanover, Germany).

### **3.2.3 Procedure**

Participants were invited to fill in the consent form. In the experiment, they were first presented with a tempo-congruent slow (60BPM) and a fast (180BPM) reference stimulus and asked to remember both as tempo “anchors” to compare with upcoming stimuli. Three blocks of trials, each consisting of the full set of randomized stimuli, were presented following the anchors. After each stimulus, participants were required to indicate their judgments of the overall tempo in its similarity to either anchor tempo by pressing on a keyboard. The slow and fast reference stimuli were replayed every 9 trials.

### **3.2.4 Analyses**

Chi-square analyses and logistic regressions were adopted to separately examine the followings: 1) Comparisons of numbers of the “fast” judgment for different auditory and visual tempi. 2) The ratio of “fast” against all responses as a logistic function of auditory and visual tempi. The points of subjective equality (PSEs), which is the audiovisual tempo when participants were equally likely to judge “slow” and “fast” were also obtained. Lastly, Pearson’s correlations were used to explore 3) the link between audiovisual tempo

discrepancies and perceived naturalness. The analyses were conducted in R (Version 3.5.3; Core Team, 2019).

### **3.3 Study 3**

#### **3.3.1 Participants**

Study 3 investigates the differences between uni- and bimodal effects on the subjective duration and emotions in the context of tempo-shifted movements. In this study, 62 participants (29 females,  $M_{Age} = 29.23$  yrs,  $SD_{Age} = 8.83$  yrs) were recruited from online survey platforms. According to participants' self-report with the Goldsmiths Music Sophistication Index (Müllensiefen et al., 2014) and Dance Sophistication Index (Rose et al., 2020), the prevalence of professional music training over 10 years (9.8%) and dance training over 4 years (6.5%) has been low. The study was conducted in accordance with the guidelines of the Ethics Committee of the Faculty of Humanities at the University of Hamburg. Two 30-Euro prizes were drawn in a lottery and awarded to two participants.

#### **3.3.2 Apparatus**

The 10-second stimuli of human movement PLDs were presented in visual-only and audiovisual forms in order to compare sensory modalities that elicited differences in judgments. The movements were in line with the ones adopted in the previous study (Wang et al., 2021); however, they were recorded in different original tempi: 86BPM, 130BPM, and 195BPM. It is intended to be consistent with the stimuli adopted in Study 2, such that future comparisons for the purpose of modality differences (within unimodal vs. unimodal and bimodal inputs) can be made. Each movement was accelerated and/or decelerated to match the other two tempi, thus creating a total of 9 stimuli. The audiovisual stimuli encompassed the 9 visual stimuli, synchronized with drumbeats from the bass drum (<http://drumbit.app>). 8 audiovisual catch trials that differed in lengths (5s and 10s) were also produced to ensure that participants' attention to stimuli durations was consistent.

#### **3.3.3 Procedure**

The experiment was conducted fully online via the platform SoSci Survey (<https://www.sosicisurvey.de/>). Participants were presented with two blocks of randomized stimuli, each including the full set of tempo-original and -shifted, visual-only, and audiovisual stimuli ( $N = 18$ ). 4 catch trials were also presented in each block. Following the display of a stimulus, participants were asked to estimate its duration in seconds and rate their felt emotional arousal, valence, as well as the naturalness of the stimulus on a 7-point scale.



### 3.3.4 Analyses

Linear mixed models (LMMs) were adopted to answer the following research questions: 1) Whether tempo-shifted movements were perceived differently from tempo-original ones regarding durations, emotion status, and naturalness. 2) Whether presentation modalities elicited perceptual differences in the judgments above. 3) How tempo shifts induced changes in the kinematic features contributed to the perceptual judgments. The LMMs took into account inter-participant variability. The analyses were conducted in R (Version 3.5.3; Core Team, 2019) using the package lme4 (Bates et al., 2015).

## 3.4 Study 4

### 3.4.1 Participants

This study aimed to examine the influences of motor involvements and, subsequently, attention effects on the internal clock speed with seconds-range intervals. To this end, 109 participants (67 females,  $M_{Age} = 26$  yrs,  $SD_{Age} = 7.04$  yrs) were recruited for the online experiment from the survey platform Prolific (<https://prolific.co/>) and the Institute of Systematic Musicology at the University of Hamburg. A small number of participants (3.7%) have received professional music training for over 10 years. The study was approved by the Ethics Committee of the Faculty of Humanities at the University of Hamburg. Participants have received an hourly compensation of €8.85/hour or course credits.

### 3.4.2 Apparatus

To provide a highly ecologically valid/naturalistic context, drumming performances were used instead of the lab-designed jumps in the two previous studies. A total of 9 audiovisual stimuli that were created from PLDs of a drummer (male, 27 years, over 20 years of professional music training) playing 9 rhythms. The rhythms varied in three complexities (simple, medium, complex) and three tempi (slow, medium, fast). The motion-capture recordings were created from the same system (Qualisys Oqus 700) and MoCap toolbox (Burger & Toiviainen, 2013) in MATLAB as in previous studies. Each stimulus lasts 15 seconds, whereas 4 catch trials varied in duration (8s and 30s).

### 3.4.3 Procedure

The experiment was conducted fully online via the platform SoSci Survey (<https://www.soscisurvey.de/>). Participants were presented with two blocks of stimuli, each comprising a full set of randomized stimuli and 2 catch trials ( $N = 11$ ). In the first block, participants were presented with a stimulus in each trial. They were then asked to estimate its

duration and judge the PoT as well as the felt expressiveness on a 7-point scale (for an example of the experiment interface, see Appendix). Before the second block, participants familiarized themselves with the even, isochronous tapping as the correct approach. In the second block, they were prompted to tap at a comfortable pace with each stimulus by pressing the spacebar. Participants then estimated the performance duration, PoT, and expressiveness.

#### **3.4.4 Analyses**

Two streams of linear mixed models (LMMs) were conducted separately to answer the research questions. In the first stream of analyses, data from both tapping and no-tapping trials were included to explore: 1) Whether the effects of tapping (motor involvement) were present on the perceptual judgments. 2) Whether there were effects of varying tempo and complexity as the attributes of rhythmic beats on the perceptual judgments. Interactions between the musical attributes (tempo, complexity) and tapping were also explored. In the second stream of models where only the tapping trials were included, it was examined 3) whether the effects of tapping speed and stability affected the perceptual judgments. For both lines of analyses, I investigated 4) whether the music training affected the judgments. The models Similar to Study 2, the analyses were conducted in R (Version 3.5.3; Core Team, 2019) using the package lme4 (Bates et al., 2015).

## **4 Main results**

### **4.1 Study 1**

Two types of internal clock models that have been investigated intensively were identified: The intrinsic and central clock models. The intrinsic model postulates that the human timing mechanism is distributed among multiple sensory modalities, thus leading to modality-specific timing performances. In contrast, the central clock model supports a universal timing mechanism that leads to universal timing performances across modalities. Establishing the modality-related difference as a fundamental gap between the two internal clock models entailed the investigation of audiovisual evidence for and against both theories. Timing performances with auditory stimuli compared to those with visual stimuli exhibited higher accuracy and sensitivity, supporting the modality-specific timing mechanism, i.e., the intrinsic clock model. However, studies also show that the training effect of time perception tasks was transferrable across modalities, suggesting that the internal clock(s), if not universal, should be at least inter-related. Alternatively, different timing mechanisms may be fit for various time ranges. A flexible clock model might allow, as we have observed, the intrinsic clock model for the sub-second range and the central clock model for the supra-second (seconds to minutes) range. Audiovisual evidence, nevertheless, is still called for to determine the timing models and the circumstances to which they are best applied.

### **4.2 Study 2**

#### **4.2.1 Visual versus auditory tempo**

A chi-square test showed that, in general, participants tended to judge the tempo of visual rather than auditory stimuli as “fast” across all tempi. The proportion of “fast” against all responses was significantly different for slow and fast tempi. More specifically, when presented with fast visual or auditory tempi, participants were more likely to judge “fast” with the fast visual tempo, whereas when presented with slow visual or auditory tempi, participants were more likely to judge “fast” with the slow auditory tempo (in other words, more “slow” judgments with the slow visual tempo). To further validate our points, a two-dimensional logistic regression was conducted to examine the likelihood of “fast” against “slow” judgments in different visual and auditory tempo conditions. Significant main effects of visual and auditory tempo on the proportion of “fast” judgments suggested that not only

did both modalities contribute to the judgments, per unit change in the visual tempo contributed more to that in auditory tempo.

#### **4.2.2 Individual modality reliance**

Participants were categorized into three types based on the results of their individual logistic regression results: Audition- ( $N_A = 7$ ), vision- ( $N_V = 16$ ), or bimodal-reliant ( $N_{AV} = 1$ ). The log ratio between auditory and visual tempo coefficients was adopted as an indicator. When the log falls between -0.05 to 0.05, participants were identified as bimodal-reliant; when the log falls below -0.05, they were identified as vision-reliant; when the log is higher than 0.05, they were identified as audition-reliant. A Chi-square test among the number of participants between types suggested that the number of vision-reliant participants was significantly higher than audition- and bimodal-reliant ones. Taken together, the results indicated a clear reliance on visual tempo compared to auditory tempo.

### **4.3 Study 3**

#### **4.3.1 Effects of tempo and modality**

Linear mixed models showed the contributions of presentation modality, stimuli tempo, manipulation, and their interactions to the judgments of emotional arousal, valence, duration estimation, and naturalness. For emotional arousal, fast tempo, acceleration, and audiovisual presentation were significantly rated significantly higher than slow, non-acceleration, and visual-only stimuli. For emotional valence, visual-only presentations were perceived as significantly higher (more positive) than audiovisual ones. A three-way interaction among manipulation, modality, and tempo revealed that, in slow conditions, the modality effect (visual-only higher than audiovisual stimuli) was again present. In contrast, acceleration was rated higher in valence than original and deceleration in medium conditions. For naturalness, significant main effects of tempo and manipulation were present, as well as the interaction effect between both. In slow and fast conditions, tempo-original stimuli were rated more natural than tempo-manipulated ones (slow - deceleration, fast - acceleration). In medium conditions, accelerated stimuli were rated more natural than decelerated ones. For duration estimation, no significant effects were found.

#### **4.3.2 Contributions of movement features changes**

A set of linear mixed models was conducted to examine the contributions of movement complexity and fluidity to the judgments of emotional arousal, valence, duration estimation, and naturalness while having tempo as a control variable. For emotional arousal, a significant

main effect of complexity and the interaction effect between fluidity and complexity were found. Post-hoc analyses suggested that increases in fluidity predicted lower arousal when movement complexity was high, while the effect was not present when complexity was low. For emotional valence, movements higher in complexity and fluidity were perceived as lower (less positive). For naturalness, movements higher in complexity and fluidity were also perceived to be less natural. For duration estimation, no significant effects were found. In general, emotional valence and arousal appeared to be more sensitive to tempo manipulations and modality differences than duration judgments.

## **4.4 Study 4**

### **4.4.1 Influence of tapping versus no-tapping**

Linear mixed models were conducted to examine participants' time and expressiveness judgments when tapping and not tapping to the stimuli. Contributions of tempo, complexity, and participants' music training (MT) were also taken into account. A significant interaction effect between tempo and tapping was found for duration estimation. Post-hoc analyses revealed that participants perceived the slow and medium tempi differently: Tapping trials were perceived significantly slower than no-tapping trials. For PoT judgments, tapping trials also passed faster than no-tapping trials. In addition, the more complex the rhythms, the faster the PoT. For the perceived expressiveness, higher tempo was associated with higher expressiveness. Furthermore, significant interaction effects between 1) tapping and MT and 2) complexity and MT revealed that only highly musically trained participants perceived the tapping trials as more expressive than no-tapping ones. They also perceived the complex stimuli as more expressive than the simple rhythms.

### **4.4.2 Influence of the tapping speed and stability**

Linear mixed models were conducted with only the tapping trials in order to examine the effects of participants' tapping speed and stability on their perceptual judgments, in addition to the effects of tempo, complexity, and MT. For PoT, the main effects of tapping stability and tempo were found: The more unstable the tapping and the faster the stimuli, the faster the PoT. However, the interaction between tapping speed and stability did not yield significant differences in post-hoc analyses. That is to say, the time-related judgments were strongly influenced by the general motor involvement but less so by the speed and stability of the tapping behaviors. A significant main effect of tempo was found for expressiveness: The faster the tempo, the more expressive the stimuli were perceived. No effects of tapping speed

and stability were found. For duration estimation, no significant effects of any predictors were found.

## **5 Discussion**

The findings from Study 1 to 4 have provided consistent evidence for the following points: Firstly, the intrinsic clock model that underlines the modality discrepancies in temporal processing might also be applicable to seconds-range intervals. Secondly, visual dominance over audition has been found in tempo judgments, supporting the speculations from Study 1 of a seconds-range intrinsic clock. In efforts to explore the uni- versus bi-modal effects, Study 3 found no differences between the two on duration estimation. However, tempo-shifted visual stimuli elicited variations in emotional valence and arousal, suggesting that the kinematic features as potential sub-dimensions of tempo may affect the internal clock speed that wasn't reflected explicitly. Finally, Study 4 revealed that motor involvement could reduce the internal clock speed by potentially distracting attentional resources on time perception. The results further supported the role of attention in seconds-range timing, raising questions of whether an intrinsic clock could also encompass cognitive engagements with long intervals. The details are discussed in the following.

### **5.1 Visual driving effect**

In Study 2, ecologically valid visual stimuli contributed more than the auditory stimuli to tempo judgments. In a tempo-incongruent context, not only did most participants prefer to rely on the visual tempo for judgments, but per unit changes in the visual tempo also elicited greater changes in the log-likelihood of the “fast” response ratio than the auditory tempo. Regardless of the auditory tempo, participants were more likely to judge “fast” when the visual tempo was fast and “slow” when the visual tempo was “slow”. The observations aligned with the intrinsic clock model by showing modality specificity, validating the local processing of the temporal units to the sensory modalities.

The intrinsic clock model hypothesizes modality-specific timing for two reasons: 1) Variations in the state of neural mechanisms that supply the perceived temporal information are local to different modalities (Karmarkar & Buonomano, 2007), and 2) As the intrinsic model found most of its evidence in the sub-second range, it has been assumed that millisecond timing was “sensory-based” (p. 564) rather than involved with attention and WM, which were regarded the center of the central clock model (Grondin, 2010). Study 2

found a modality effect with 5-second stimuli. It indicates the applicable range of the intrinsic clock model could be extended to the supra-second range.

A possible explanation can be found with the time-based expectancy effect (Thomaschke et al., 2015), which stemmed from the state-dependent network theory (Karmarkar & Buonomano, 2007) as an integral mechanism to the intrinsic clock model. More specifically, the time-based expectancy effect refers to faster neural and behavioral responses towards familiarized intervals with specific sensory features as an encoded state change than the unexpected intervals with new features (Thomaschke et al., 2015). A study found a time-based expectancy effect with both sub- and supra-second-ranged stimuli in a time-event correlation task (Aufschnaiter et al., 2020), suggesting that the internal clock might also be supported within the seconds range. Such findings, together with results from Study 2, could contribute to the possibility of a supra-second scale timing mechanism that entails local processing of temporal information. Alternatively, the effect indicates that the sensory-based timing and cognitive processes, including attention and WM, may be equally involved with the internal clock model. It has, therefore, raised the possibility of a new clock model that could integrate both sides in the supra-second range, which calls for future explorations of its neural mechanisms and corresponding behavioral evidence

The effects of visual stimuli on tempo judgments further confirmed that vision could drive the perception of time when ecologically valid stimuli were applied. The finding challenged the modality-appropriateness theory (Ward, 1994) that vision was deemed more appropriate for spatial processing while audition was for temporal judgments. This might be due to high salience with valid visual inputs. As the Colativa effect (Colavita, 1974) suggested, when visual and auditory inputs were presented simultaneously, visual inputs were perceived as more salient. Similarly, in this case, the visual aspect of the audiovisual stimuli could receive more attention than the auditory tempo. The reliability of the sensory modalities in temporal processing might be a possible explanation for the visual driving effect. In accordance with the information reliability theory (Andersen et al., 2004), the reliability of modalities in human timing represented the amount of signal in relation to noise: The more reliable the modality, the more weight it would be assigned to in temporal processing, such as in duration estimation tasks (Hartcher-O'brien et al., 2014). Therefore, previous findings on auditory dominance in time-related judgments (e.g. Guttman et al., 2005; Shipley, 1964) may have taken advantage of the lower reliability in its visual counterparts compared to the auditory inputs. A study that evaluated auditory and visually

induced fission and fusion effects (two closely occurring events accompanied by one inducer were perceived as one, while one event accompanied by two inducer was perceived as two) also suggested that the visual cues were less relied upon potentially due to lower temporal resolution with vision, and hence lower salience of the visual stimuli (Apthorp et al., 2013). It is, therefore, inevitable to discuss the ecological validity that might be essential to the temporal resolution of vision, hence its reliability in tempo judgments.

## **5.2 Information specificity**

Study 2 hypothesized that visual trajectories that were plausible to human perceptual experiences (left-right jumps performed by a human actor) would also provide high specificity, allowing vision to exert greater influences on tempo judgments than audition. As the results suggested, visual stimuli were more relied on than auditory stimuli regarding tempo judgments. The specificity of the visual information was partially reflected by the perceived naturalness, as a tempo effect on the naturalness rating was present with vision but not with audition: The faster the tempo, the more natural it was perceived. Consistent with previous studies (Hove et al., 2010, 2013), the findings pointed to a dominant role of vision in time-related judgments when the trajectories of the stimuli were compatible with viewers' expectations based on their perceptual experiences. It also leads to a possible connection that whichever modality is perceived as more ecologically valid is also more reliable and, in accordance with the information reliability theory (Andersen et al., 2004), could be assigned more weights in temporal judgments.

## **5.3 Effects of visual vs. audiovisual stimuli on perceived durations**

Following the evidence from Study 2 that an internal clock model might be supported, Study 3 further extended the comparison between the effects of unimodal (visual) and bimodal (audiovisual) stimuli on subjective time. Participants showed no significant differences in duration estimation between uni- and bi-modal presentations. However, audiovisual stimuli were rated higher in arousal but less positive in valence than visual stimuli in the perceived durations. The absence of additive effect with bimodal inputs was opposed to past findings that congruent multisensory signals were usually perceived as more salient and precisely; they could also elicit stronger neural responses than unimodal inputs (Van der Burg et al., 2011; Van Wassenhove et al., 2007). However, it has been argued that more attention to bimodal than unimodal inputs did not necessarily translate to better temporal performances, such as discrimination of temporal locations (e.g. Apthorp et al., 2013). What hinders the



process in between might be that attention was allocated to the more salient party in a multisensory context. Even though the audiovisual stimuli were higher in salience than the unimodal ones, the discrepancy between the two modalities within the audiovisual inputs potentially still exists. In Study 3, a significant difference between the effects of visual and audiovisual stimuli was not found, which pointed to the possibility that vision was a more salient modality than audition. Thus, it was more attended to than the auditory component overall. This would align with the finding from Study 2 that visual tempo has a dominant role in the overall tempo judgments of audiovisual presentations. However, the question of how bimodal information is integrated in the optimal fashion remains.

#### **5.4 Tempo manipulations: Absence of temporal effects**

Study 3 shed light on the sub-dimensions of tempo by examining the factors that could elicit variations in emotion and temporal judgments when the presentation tempo remained the same. The findings revealed that movement complexity and fluidity impacted the perceived emotional valence and arousal but not directly on the estimated durations.

The effects of kinematic feature changes were not present for duration estimation, possibly due to the lower sensitivity of this measure to internal clock speed variations compared to PoT judgments. A previous study found that tempo changes affected the perceived PoT but not duration judgments in a seconds-range task; the effects on PoT and duration estimation were only significant in a minutes-range task (Droit-Volet et al., 2017). The finding pointed out that the timescale that allowed the involvement of cognitive processes, such as the retrieval and comparison of memory, should be met in order to elicit changes in the subjective duration. Study 3 shared the feature of a seconds-range timescale in its task with Droit-Volet et al.'s (2017) experiment. Alternatively, the absence of effect could be due to the limited extent of kinematic feature changes. A study with tempo-shifted disco music (from 105BPM to 125BPM) has not seen changes in the perceived durations but rather in duration reproduction, which did not demand as much cognitive processing as in the duration estimation task (Hammerschmidt et al., 2021). Together with the absence of kinematic effects on duration estimation in Study 3, the studies elucidate a potential characteristic of the intrinsic clock model that: 1) The clock speed changes could be reflected by lower-level temporal judgments with greater sensitivity than by higher-level judgments; 2) Tempo as well as kinematic features changes to a minimum extent are needed to entrain the internal clock speed, in accordance with the dynamic attention theory (Jones & Boltz, 1989).

Considering this possibility, Study 4 further suggested the asymmetry of lower-level vs. higher-level temporal judgments was present in an audiovisual context.

With the same presentation tempo, increased movement complexity led to higher arousal and less positive valence. In comparison, increases in fluidity were associated with lower arousal (only when complexity was high) and also less positive valence. The changes in emotion measurements potentially reflect the influences of kinematic changes on subjective time. Previous evidence has supported robust connections between emotional arousal and tempo (e.g. Droit-Volet et al., 2010; Droit-Volet et al., 2013): Faster tempo was associated with higher arousal and slower tempo with lower arousal. Considering the association between tempo and the internal clock speed (when entrained to the external tempo) (Wang & Wöllner, 2019), it might be speculated that high arousal is related to faster clock speed. Recent evidence found a direct correlation between higher psychological and physiological arousal with longer perceived duration than when the arousal was low (Appelqvist-Dalton et al., 2022), further suggesting that arousal could be associated with clock speed. However, the link between emotional valence and the internal clock bears ambiguity. As discussed in Droit-Volet et al.'s works, a comparison between the effects of happy vs. sad music supported the idea that valence did not induce significant changes in duration judgments (2010). In contrast, pleasant vs. unpleasant music suggested that stimuli of positive valence were judged shorter (2013). However, the extent of changes in fluidity and complexity might not have reached the lower threshold for significant changes in the perceived durations, similar to earlier findings on the minimum tempo change (Hammerschmidt et al., 2021). Based on the findings, it might be speculated that more complex movements were associated with faster clock speed, while more fluid movements with slower clock speed. However, the speculation has to be interpreted with caution, as no direct evidence from Study 3 has suggested that the changes in kinematic features (were strong enough to) affect the internal clock speed.

## **5.5 Tapping changes the subjective time**

Having the understanding of modality-specific temporal processing from Study 2 and 3, Study 4 moved on to investigate the performances of an internal clock model when proactive timing in an ecologically valid, audiovisual context was involved. In this study, motor involvements have been found to induce slower passage of time and shorter duration perceived in slow and medium tempo conditions. The findings are consistent with results

from past studies, where the tapping task parallel to the timing tasks led to duration underestimation and faster PoT (Hammerschmidt, Wöllner, et al., 2021; Wöllner & Hammerschmidt, 2021). This was potentially due to diverted attention resources allocated to tapping from the timing. According to the Attentional Gate Model (Block et al., 2010), attention on the temporal units emitted directly influences the internal clock speed by controlling the units that pass through the “gate”. A simultaneous task that needs attending could reduce the recorded temporal pulses, leading to fewer pulses accumulating in the pacemaker-counter device (Fraisse, 1978; Gibbon, 1977). In this vein, participants in Study 4 were distracted from the duration estimation and PoT judgments when keeping an isochronous beat to the drumming performances. However, it should be noted that the speculated mechanism of attention allocation supports a central clock model, which has been found to be the most applicable in seconds-range timing (Buhusi & Meck, 2005). The high-level cognitive control that involves attention, rather than low-level, sensory-based timing in the milliseconds range, has been believed to be a feature of the seconds-range timing (for a review, see Grondin, 2010). Considering evidence that was found in Study 2 and 3 regarding a possible intrinsic clock model in seconds-range timescale, future studies should explore the role of attention when temporal information is processed locally to each sensory modality.

It is noteworthy that the ecological validity in Study 4 was potentially higher than in Study 2 and 3, where the left-right jump was a lab-designed movement that was aimed at motion continuity and motor control (for details, see Allingham et al., 2021). Study 4 has employed the point-light displays of a drummer performing three rhythms varying in tempo and complexity, which could be more natural and common to participants’ perceptual experiences. The transferability of temporal judgment tasks from the artificial lab setting to real-world, ecologically valid scenarios has been an emerging topic in time-related research (for a review, see Matthews & Meck, 2014). Nevertheless, this review reached no conclusive judgments on the impacts of naturalistic settings on subjective time. A comparison of the results from the Study 2 to 4 suggested that ecological validity may have influenced the temporal judgments, which potentially reflect the involvement of cognitive process in seconds-range timescale: With left-right jumps, tempo encoding was successful with both auditory and visual tempo. However, such movements did not lead to estimated duration changes. With the drumming performances, significant changes have been observed in both PoT and duration judgments (at slow and medium tempi), implying that more naturalistic scenes could be linked to a greater extent of changes in the internal clock speed than the less

plausible ones. As previously discussed, more temporal units could be accumulated when more attention was allocated to the timing task. In this vein, it may be hypothesized that stimuli of higher ecological validity received more attention than the ones of lower validity. This additional finding demands further investigations of whether different levels of measured ecological validity bring low- and high-level temporal encoding into play, which could translate to the seconds-range applicability of internal clock models.

## **6 Limitations and future directions**

The studies encompassed a few limitations that could be improved and further explored in future studies. To begin with, the effects of ecological validity on time-related judgments were subject to the lack of a systematic measurement. As discussed earlier (Lewkowicz, 2001), an ecologically valid stimulus should be consistent with participants' perceptual experiences. Though Study 2 and 3 have measured the perceived naturalness of the stimuli, the definition of what naturalness was for the participant bears ambiguity. In turn, whether naturalness can be translated into ecological validity was unclear. Past studies that have adopted such stimuli focused on a compatible spatial-temporal relationship of the movement trajectories in relatively simple motions such as finger tapping (Hove et al., 2010), while others applied real-life scenarios such as video game playing (Tobin et al., 2010). Future studies should consider both qualitative and quantitative approaches in exploring how ecological validity is represented across the population, especially in complex scenarios. Such findings would shed light on the explanation for the possible effects of attention on temporal judgments with ecological validity compared to artificial stimuli.

Regarding the visual advantage that was found with the tempo judgment task, the studies have no direct evidence that this unique effect, potentially associated with increased ecological validity, leads to changes in PoT and duration estimation. This was due to the absence of design in the following studies, where audiovisual incongruent stimuli were not used. Future studies could consider such paradigms where auditory and visual information of similar ecological validity measured by pre-test ratings are judged equally or not in time-related tasks.

Another limitation of the study is that the durations of the stimuli ranged from 5 to 15 seconds. This has restricted our observation of the human timing performances, alternatively the functions of the internal clock to the seconds-range timescale. Earlier evidence has

mainly indicated the applicability of the intrinsic model in the milliseconds range (e.g. Burr et al., 2009). Though modality-specific timing performances in the seconds range that supported the intrinsic clock model were found, it remains a question whether similar effects could be observed in longer time spans. As Tobin and colleagues (2010) pointed out, ecologically valid timing scenarios could often be composed of minutes and durations beyond in addition to experience-coherent motions. Future studies must rethink the existing internal clock models on neural and behavioral levels if a modality effect in temporal processing is found in the minute range and beyond.

## **7 Conclusion**

Altogether, this research project has gathered theoretical and empirical evidence on the topic of the intrinsic clock model as a possible internal timing mechanism. The findings contribute to validating the features of an internal clock in the seconds-range timescale from a few perspectives: 1) Whether temporal processing is universal or local; 2) Whether cross-modal timing offers an advantage over unimodal inputs through optimal integration of the temporal information; 3) Time experiences that reflect the functionality of the internal clock in a highly naturalistic, multisensory context.

The modality-specific tempo judgments supported the presence of a distributed timing system, such that temporal inputs were processed as a local flow of information. An advantage of visual inputs was also found, challenging the traditional view of auditory dominance in temporal processing in circumstances of high ecological validity. The visual advantage was further evidenced by the absence of an additive effect when comparing visual to audiovisual inputs of the same movements (left-right jumps), suggesting that vision was potentially of high salience in duration judgments. By comparing the temporal judgments with drumming performances to that with lab-designed jumps across studies, it may be deduced that stimuli of high ecological validity receive more attention in temporal judgments than the ones of low validity, thus implying the internal clock model is subject to environmental influences. The speculation, however, requires further validation where a direct comparison is made. Furthermore, the findings unanimously emphasize the role of attention in time perception, which is essential to determining the internal clock models. Past evidence supports the optimal applicability of a central clock model in a seconds range timescale where attention as a part of the cognitive processes plays a role, while the evidence from the current project leaves room for the involvement of attention in the intrinsic model.

Therefore, despite the evidence on modality effect, the debate between an intrinsic and a central clock model is still open for discussion.

The findings shed light on two aspects of future studies: Firstly, greater importance should be attached to ecological validity than in the past, as it appeared to affect the internal clock speed and, ultimately, how modalities weighed in temporal processing. Secondly, how attention affects the modality-specific timing should be studied in greater detail to clarify the mechanism of an intrinsic clock in the supra-second timescale. The studies also provide insights into how our timing behaviors could be interpreted and predicted in laboratory settings and real-world scenarios. For instance, to influence one's subjective time, the visual aspect should be assigned greater weight than the auditory one. In the end, this dissertation hopes to deepen the understanding of the theoretical grounds for the internal clock models and find evidence that contributes to its mechanism on the behavioral level.

## 8 References

- Allingham, E., Hammerschmidt, D. & Wöllner, C. (2021). Time perception in human movement: Effects of speed and agency on duration estimation. *Quarterly Journal of Experimental Psychology*, 74(3), 559–572. <https://doi.org/10.1177/1747021820979518>
- Allman, M. J. & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain*, 135(3), 656–677. <https://doi.org/10.1093/brain/awr210>
- Andersen, T., Tiippana, K. & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308. <https://www.sciencedirect.com/science/article/pii/S0926641004001636>
- Angrilli, A., Cherubini, P., Pavese, A. & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception and Psychophysics*, 59(6), 972–982. <https://doi.org/10.3758/BF03205512>
- Appelqvist-Dalton, M., Wilmott, J. P., He, M. & Simmons, A. M. (2022). Time perception in film is modulated by sensory modality and arousal. *Attention, Perception, & Psychophysics*, 1–17. <https://doi.org/10.3758/S13414-022-02464-9>
- Apthorp, D., Alais, D. & Boenke, L. T. (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. *Journal of Vision*, 13(5), 3–3. <https://doi.org/10.1167/13.5.3>
- Aubry, F., Guillaume, N., Morigato, G., Bergeret, L. & Celsis, P. (2008). Stimulus complexity and prospective timing: Clues for a parallel process model of time perception. *Acta Psychologica*, 128(1), 63–74. <https://doi.org/10.1016/J.ACTPSY.2007.09.011>
- Aufschnaiter, S., Kiesel, A. & Thomaschke, R. (2020). Humans derive task expectancies from sub-second and supra-second interval durations. *Psychological Research*, 84(5), 1333–1345. <https://doi.org/10.1007/s00426-019-01155-9>
- Bak, K., Chan, G., Schutz, M. & Campos, J. (2021). Perceptions of Audio-Visual Impact Events in Younger and Older Adults. *Multisensory Research*, 34(8), 839–868. <https://doi.org/10.1163/22134808-bja10056>
- Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>

- Block, R. (2003). *Psychological timing without a timer: The roles of attention and memory*.  
<https://psycnet.apa.org/record/2004-00315-003>
- Block, R. A., Hancock, P. A. & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, 3, 330–343.  
<https://doi.org/https://doi.org/10.1016/j.actpsy.2010.03.006>
- Bolger, D., Trost, W. & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, 142(2), 238–244.  
<https://doi.org/10.1016/J.ACTPSY.2012.11.012>
- Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities. *Perception and Psychophysics*, 67(8), 1362–1375.  
<https://doi.org/10.3758/BF03193641>
- Bratzke, D., Seifried, T. & Ulrich, R. (2012). Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research*, 221(2), 205–210. <https://doi.org/10.1007/S00221-012-3162-0/FIGURES/2>
- Buhusi, C. V. & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. In *Nature Reviews Neuroscience* (Vol. 6, Issue 10, pp. 755–765). Nature Publishing Group. <https://doi.org/10.1038/nrn1764>
- Buonomano, D. V. & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2), 113–125.  
<https://doi.org/https://doi.org/10.1038/nrn2558>
- Burger, B., Saarikallio, S., Luck, G., Thompson, M. R. & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception*, 30(5), 517–533. <https://doi.org/10.1525/MP.2013.30.5.517>
- Burger, B. & Toiviainen, P. (2013). MoCap Toolbox-A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the Sound and Music Computing Conference 2013* (pp. 172–178). Logos Verlag Berlin.  
<http://urn.fi/URN:NBN:fi:jyu-201401211091>
- Burr, D., Banks, M. S. & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, 198(1), 49–57.  
<https://doi.org/10.1007/s00221-009-1933-z>



- Chen, L., Zhou, X., Müller, H. J. & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration. *Journal of Experimental Psychology: General*, *147*(12), 1851–1864. <https://doi.org/10.1037/xge0000487>
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. <https://doi.org/10.3758/BF03203962>
- de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, *13*(4), 247–268. <https://doi.org/10.1007/BF00990296>
- Droit-Volet, S., Ramos, D., Bueno, J. L. O. & Bigand, E. (2013). Music, emotion, and time perception: The influence of subjective emotional valence and arousal? *Frontiers in Psychology*, *4*(JUL), 417. <https://doi.org/10.3389/fpsyg.2013.00417>
- Droit-Volet, S., Trahanias, P. & Maniadakis, M. (2017). Passage of time judgments in everyday life are not related to duration judgments except for long durations of several minutes. *Acta Psychologica*, *173*, 116–121. <https://doi.org/10.1016/j.actpsy.2016.12.010>
- Droit-Volet, Sylvie, Bigand, E., Ramos, D. & Bueno, J. L. O. (2010). Time flies with music whatever its emotional valence. *Acta Psychologica*, *135*(2), 226–232. <https://doi.org/10.1016/j.actpsy.2010.07.003>
- Eerola, T., Luck, G. & Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. *Proceedings of the 9th International Conference on Music Perception & Cognition, Bologna*, 472–476. <http://www.marcocosta.it/icmpc2006/pdfs/235.pdf>
- Escoffier, N., Sheng, D. & Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica*, *135*(1), 12–16. <https://doi.org/https://doi.org/10.1016/j.actpsy.2010.04.005>
- Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Perceptual Coding* (pp. 203–254). Academic Press.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. <https://doi.org/10.1037/0033-295X.84.3.279>
- Gil, S., Rousset, S. & Droit-Volet, S. (2009). How liked and disliked foods affect time perception. *Emotion*, *9*(4), 457. <https://doi.org/10.1037/a0015751>

- Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research*, 220(1), 51–61. <https://doi.org/10.1007/s00221-012-3114-8>
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, and Psychophysics*, 72(3), 561–582. <https://doi.org/https://doi.org/10.3758/APP.72.3.561>
- Guttman, S. E., Gilroy, L. A. & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science*, 16(3), 228–235. <https://doi.org/10.1111/j.0956-7976.2005.00808.x>
- Hammerschmidt, D. & Wöllner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time. *Music Perception*, 37(4), 263–277. <https://doi.org/https://doi.org/10.1525/mp.2020.37.4.263>
- Hammerschmidt, D., Wöllner, C., London, J. & Burger, B. (2021). Disco time: The relationship between perceived duration and tempo in music. *Music & Science*, 4, 2059204320986384. <https://doi.org/10.1177/2059204320986384>
- Hartcher-O’Brien, J., Luca, M. Di & Ernst, M. O. (2014). The Duration of Uncertain Times: Audiovisual Information about Intervals Is Integrated in a Statistically Optimal Fashion. *Journals.Plos.Org*, 9(3). <https://doi.org/10.1371/journal.pone.0089339>
- Hove, M. J., Iversen, J. R., Zhang, A. & Repp, B. H. (2013). Synchronization with competing visual and auditory rhythms: Bouncing ball meets metronome. *Psychological Research*, 77(4), 388–398. <https://doi.org/10.1007/s00426-012-0441-0>
- Hove, M. J., Spivey, M. J. & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1525–1534. <https://doi.org/10.1037/a0019059>
- Huang, Y., Gu, L., Yang, J., Zhong, S. & Wu, X. (2018). Relative contributions of the speed characteristic and other possible ecological factors in synchronization to a visual beat consisting of periodically moving stimuli. *Frontiers in Psychology*, 9(JUL), 1226. <https://doi.org/10.3389/FPSYG.2018.01226/BIBTEX>
- Ivry, R. B. & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12(7), 273–280. <https://doi.org/10.1016/j.tics.2008.04.002>

- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. <https://doi.org/10.3758/BF03212378>
- Jones, M. R. & Boltz, M. (1989). Dynamic Attending and Responses to Time. *Psychological Review*, *96*(3), 459–491. <https://doi.org/10.1037/0033-295X.96.3.459>
- Karmarkar, U. R. & Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron*, *53*(3), 427–438. <https://doi.org/10.1016/j.neuron.2007.01.006>
- Lewkowicz, D. J. (2001). The Concept of Ecological Validity: What Are Its Limitations and Is It Bad to Be Invalid? *Infancy*, *2*(4), 437–450. [https://doi.org/10.1207/S15327078IN0204\\_03](https://doi.org/10.1207/S15327078IN0204_03)
- London, J. (2011). Tactus ≠ tempo: Some dissociations between attentional focus, motor behavior, and tempo judgment. *Empirical Musicology Review*, *6*, 43–55.
- Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. In *Behavior Research Methods* (Vol. 44, Issue 2, pp. 314–324). Springer. <https://doi.org/10.3758/s13428-011-0168-7>
- Matthews, W. J. & Meck, W. H. (2014). Time perception: The bad news and the good. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(4), 429–446. <https://doi.org/10.1002/wcs.1298>
- Mayer, K. M., Di Luca, M. & Ernst, M. O. (2014). Duration perception in crossmodally-defined intervals. *Acta Psychologica*, *147*, 2–9. <https://doi.org/10.1016/J.ACTPSY.2013.07.009>
- Motala, A., Heron, J., McGraw, P. V., Roach, N. W. & Whitaker, D. (2018). Rate after-effects fail to transfer cross-modally: Evidence for distributed sensory timing mechanisms. *Scientific Reports*, *8*(1), 1–10. <https://doi.org/10.1038/s41598-018-19218-z>
- Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Ortega, L., Guzman-Martinez, E., Grabowecky, M. & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Attention, Perception, and Psychophysics*, *76*(5), 1485–1502.

<https://doi.org/10.3758/s13414-014-0663-x>

- Repp, B. & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085. <https://doi.org/https://doi.org/10.1037/0096-1523.28.5.1085>
- Rose, D., Müllensiefen, D., Lovatt, P. & Orgs, G. (2020). *The Goldsmiths Dance Sophistication Index (Gold-DSI): a new psychometric tool to assess individual differences in dance experience*. <https://doi.org/10.31234/osf.io/wapkx>
- Schutz, M. & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36(6), 888–897. <https://doi.org/10.1068/p5635>
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145(3638), 1328–1330. <https://doi.org/10.1126/science.145.3638.1328>
- Su, Y. H. (2014). Visual enhancement of auditory beat perception across auditory interference levels. *Brain and Cognition*, 90, 19–31. <https://doi.org/10.1016/j.bandc.2014.05.003>
- Su, Y. H. & Jonikaitis, D. (2011). Hearing the speed: Visual motion biases the perception of auditory tempo. *Experimental Brain Research*, 214(3), 357–371. <https://doi.org/10.1007/s00221-011-2835-4>
- Thomaschke, R., Kunchulia, M. & Dreisbach, G. (2015). Time-based event expectations employ relative, not absolute, representations of time. *Psychonomic Bulletin and Review*, 22(3), 890–895. <https://doi.org/10.3758/S13423-014-0710-6/FIGURES/1>
- Tobin, S., Bisson, N. & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLoS ONE*, 5(2), 16–18. <https://doi.org/10.1371/journal.pone.0009271>
- Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the “internal clock”. *Psychological Monographs*, 77(13), 1–31. <https://doi.org/10.1037/h0093864>
- Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C. & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, 55(3), 1208–1218. <https://doi.org/10.1016/J.NEUROIMAGE.2010.12.068>

- Van Wassenhove, V., Grant, K. W. & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598–607.  
<https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2006.01.001>
- Vergheze, P. & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, 381(6578), 161–163. <https://doi.org/10.1038/381161a0>
- Wang, X. & Wöllner, C. (2019). Time as the ink that music is written with: A review of the internal clock models and their explanatory power in audiovisual perception. *DGM Jahrbuch*, 29(2019), 1–22. <https://doi.org/https://doi.org/10.5964/jbdgm.2019v29.67>
- Wang, X., Wöllner, C. & Shi, Z. (2021). Perceiving Tempo in Incongruent Audiovisual Presentations of Human Motion: Evidence for a Visual Driving Effect. *Timing & Time Perception*, 1–21. <https://doi.org/10.1163/22134468-bja10036>
- Wapnick, J., Campbell, L., Siddell-Strebel, J. & Darrow, A. A. (2009). Effects of non-musical attributes and excerpt duration on ratings of high-level piano performances. *Musicae Scientiae*, 13(1), 35–54.  
<https://doi.org/https://doi.org/10.1177/1029864909013001002>
- Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, 48(2), 242–259.  
<https://doi.org/10.1037/1196-1961.48.2.242>
- Warm, J. S., Stutz, R. M. & Vassolo, P. A. (1975). Intermodal transfer in temporal discrimination. *Perception & Psychophysics*, 18(4), 281–286.  
<https://doi.org/10.3758/BF03199375>
- Wittmann, M. (2013). The inner sense of time: How the brain creates a representation of duration. *Nature Reviews Neuroscience*, 14(3), 217–223.  
<https://doi.org/10.1038/nrn3452>
- Wöllner, C. & Hammerschmidt, D. (2021). Tapping to hip-hop: Effects of cognitive load, arousal, and musical meter on time experiences. *Attention, Perception, & Psychophysics*, 83(4), 1552–1561. <https://doi.org/10.3758/S13414-020-02227-4>
- Wöllner, C., Hammerschmidt, D. & Albrecht, H. (2018). Slow motion in films and video clips: Music influences perceived duration and emotion, autonomic physiological

activation and pupillary responses. *PLOS ONE*, 13(6), e0199161.

<https://doi.org/10.1371/journal.pone.0199161>

Zakay, D. & Block, R. A. (1995). An attentional gate model of prospective time estimation. *Time and the Dynamic Control of Behavior*, 5, 167–178.

## Summary (English)

Time cannot be perceived directly via a dedicated sensory organ but is rather construed from inputs of multiple sensory modalities. The debate of whether the internal clock model is central or intrinsic (distributed) has been a topic of interest in time perception studies. This dissertation aims to find theoretical and empirical evidence for the optimal internal clock model in the seconds range timescale. In addition, human timing performances on different levels of ecological validity were also examined to further verify the sensitivity of the internal clock model towards the environmental influences as a less explored attribute. Four studies were conducted accordingly: Study 1 explored the literature backgrounds of the two clock models and identified key points of differences. Study 2 followed the thread of modality specificity in temporal processing and confirmed the discrepancy between visual and auditory inputs with tempo judgments. Study 3 further compared the duration judgments with visual and audiovisual stimuli using the same movements as in Study 2, investigating whether bimodal stimuli change time perception compared to unimodal inputs. Finally, Study 4 examined the participants' passage of time (PoT) ratings and duration estimation in the context of audiovisual drumming performances, which entailed higher ecological validity than the lab-designed motions in Study 2 and 3.

In Study 1, a narrative literature review was conducted to investigate the main internal clock models, the supportive mechanisms, and the key factors that could affect the internal clock speed. The study identified the intrinsic and central clock as the optimal leading models in the sub-seconds and supra-seconds timescales, respectively. Evidence that supports either model diverged on the presence of a modality effect: The intrinsic clock is in line with modality-specific timing with an emphasis on sub-second, sensory-based timing. In contrast, the central clock hypothesizes a universal timer across modalities where cognitive mechanisms such as attention contribute to the seconds' range and beyond. However, empirical evidence suggested that the modality-specific effect in time-related judgments was also present in the seconds-to-minute time span. The finding invites further exploration to confirm the possibility of an intrinsic clock model in the time range of what used to be believed to be the applicable range for the central clock.

To this end, Study 2 investigated tempo judgments in a tempo-incongruent context with the bisection paradigm. Participants judged the overall tempo of a stimulus to be closer to the slow or fast reference tempi. In contrast, the auditory and visual tempi of the stimulus did not necessarily align. The findings suggested that participants not only preferred to rely on the visual tempo, but it also elicited greater change per unit in the proportion of "fast" judgments than the auditory tempo overall. In this case, the results support the intrinsic clock model in the supra-second timescale. Apart from the modality specificity, Study 2 also revealed a possibility of visual dominance in temporal processing by increasing the ecological validity of the stimulus. Instead of employing artificial stimuli as previous studies did, this study used human movements that may have provided higher plausibility that was in line with the participants' perceptual experience.

Study 3 continued investigating modality-relevant effects for the intrinsic clock model by comparing perceived durations and emotions between visual (unimodal) and audiovisual (bimodal) presentations. In this study, participants judged the durations, emotional valence, and arousal of tempo-shifted movements (same basis as in Study 2). No significant differences were found between the modalities, suggesting that the bimodal advantage in temporal processing that has been found in some studies earlier may not be present when ecologically valid movements were evaluated. This might be due to the high salience of the

visual stimuli, again supporting the visual dominance effect found in Study 2. In addition, the findings showed that emotional arousal and valence were sensitive to differences in movement complexity and fluidity when stimuli of different original tempi were shifted to the same speed. This indicates a possibility that tempo, as an important factor in the internal clock speed, is not merely event frequency but a multiple-faceted concept that calls for future exploration.

Extending on the findings of ecological validity, Study 4 adopted a musical context of high ecological validity where a drummer performed rhythms varying in tempo and complexity as stimuli. The study compared timing performances when participants observed the stimuli only and when they freely and isochronously tapped to the stimuli. The findings indicated significantly faster PoT and shorter duration (at slow and medium tempi) perceived when tapping than not tapping. In this case, motor involvement appeared to distract attention from time judgments, suggesting that attention is a key component in supra-second timing. Though not present in all conditions, tempo effects on both PoT and duration estimation in Study 4 compared to not finding an effect in Study 3 might suggest that naturalistic performance scenes that were more ecologically valid than lab-designed movements potentially also received more attention, leading to a greater extent of changes in the internal clock speed according to the pacemaker-counter theory. It is speculated that ecological validity as an important environmental factor influences the allocation of attention in temporal processing. However, the existing intrinsic clock lacks the power to explain the role of attention in supra-second timing as with the central clock model. Therefore, the discrepancy demands future investigations of a timing mechanism compatible with modality specificity and cognitive processes indispensable to seconds-range temporal judgments.

In general, the findings of the current dissertation have contributed to the understanding of the internal clock that is at the center of time perception in an audiovisual context. Study 1 identified modality specificity that essentially distinguished the central vs. intrinsic clock model. The Modality effects found in Study 2 further provided evidence for an intrinsic clock in the supra-second timescale. Using the same movement stimuli, Study 3 revealed the absence of a bimodal advantage in temporal processing compared to unimodal stimuli, potentially corresponding to a visual dominance effect with ecologically valid inputs. Study 4 found evidence for the involvement of attention in supra-second PoT and duration judgments, in addition to a possible effect of ecological validity on the internal clock speed. The observations provide implications for the evolution of internal clock models, along with passive and proactive strategies that could influence time perception in day-to-day auditory and visual experiences.



## Zusammenfassung (Deutsch)

Zeit kann nicht direkt über ein bestimmtes Sinnesorgan wahrgenommen werden, sondern wird vielmehr aus den Informationen mehrerer Sinnesmodalitäten konstruiert. Die Diskussion darüber, ob der menschliche Zeitmessungsmechanismus, die sogenannte innere Uhr, ein zentrales oder ein intrinsisches (verteilt)es Uhrenmodell darstellt, ist ein wichtiges Thema in der Zeitwahrnehmungsforschung. Ziel dieser Dissertation war es, theoretische und empirische Belege für das optimale interne Uhrenmodell im Sekundenbereich zu finden. Dazu wurden menschliches Zeitempfinden auf verschiedenen Ebenen der ökologischen Validität untersucht, um die Sensitivität des internen Uhrenmodells gegenüber Umwelteinflüssen als einen weniger erforschten Gegenstand weiter zu verifizieren. Hierfür wurden vier Studien durchgeführt: Studie 1 untersuchte die existierende Forschungsliteratur zu den beiden Uhrenmodellen und identifizierte die wichtigsten Unterschiede. Studie 2 vertiefte den Aspekt der Modalitätsspezifität bei der zeitlichen Verarbeitung und bestätigte die Diskrepanz zwischen visuellen und auditiven Inputs bei Tempobeurteilungen. In Studie 3 wurde die Beurteilung der Dauer bei visuellen und audiovisuellen Reizen mit denselben Stimulus-Bewegungen wie in Studie 2 verglichen in Erwartung eines additiven Effekts bei bimodalen im Gegensatz zu unimodalen Inputs. Studie 4 untersuchte schließlich die Zeitspannen- (PoT) und Zeitdauerinschätzung von Versuchsteilnehmenden im Kontext audiovisueller Perkussionsperformances, die eine höhere ökologische Validität als die im Labor entworfenen Bewegungen in Studie 2 und 3 mit sich brachten.

Studie 1 beschäftigt sich mit der vorhandenen Literatur, um die wichtigsten Modelle der inneren Uhr, die unterstützenden Mechanismen sowie die Schlüsselfaktoren zu untersuchen. In der Studie wurden die intrinsische Uhr und die zentrale Uhr als die führenden Modelle identifiziert, die im Sub-Sekunden- bzw. Supra-Sekunden-Zeitbereich optimal waren. Die Beweise, die die beiden Modelle unterstützen, unterscheiden sich hinsichtlich des Vorhandenseins eines Modalitätseffekts: Die intrinsische Uhr steht im Einklang mit der modalitätsspezifischen Zeitmessung mit Schwerpunkt auf der sensorischen Zeitmessung im Subsekundenbereich, während die zentrale Uhr von einem universellen Zeitgeber über alle Modalitäten hinweg ausgeht, bei dem die kognitiven Mechanismen, wie z. B. die Aufmerksamkeit, einen Beitrag im Sekundenbereich und darüber hinaus leisten. In Studie 1 wurden jedoch empirische Belege gefunden, die den Modalitätseffekt auch auf der supra-sekundlichen Zeitskala hervorheben. Dieses Ergebnis lädt zu weiteren Untersuchungen ein, um die Möglichkeit eines intrinsischen Uhrenmodells in dem Zeitbereich zu bestätigen, der früher als der für die zentrale Uhr geltende Bereich angesehen wurde.

Zu diesem Zweck wurden in Studie 2 die Tempobeurteilungen in einem tempo-inkongruenten Kontext mit dem Bisektionsparadigma untersucht. Die Versuchsteilnehmenden bewerteten das Gesamttempo eines Reizes als näher am langsamen oder schnellen Referenztempo, während das auditive und visuelle Tempo des Reizes nicht unbedingt übereinstimmte. Die Ergebnisse deuten darauf hin, dass nicht nur mehr Teilnehmende es vorzogen sich auf das visuelle Tempo zu verlassen, sondern dass dieses auch eine größere Veränderung pro Einheit im Anteil der "schnellen" Beurteilungen hervorrief als das auditive Tempo insgesamt. In diesem Fall spricht die Evidenz für das intrinsische Uhrenmodell auf der Supra-Sekunden-Zeitskala. Abgesehen von der Modalitätsspezifität zeigte Studie 2 auch eine mögliche visuelle Dominanz bei der zeitlichen Verarbeitung, indem die ökologische Validität des Stimulus erhöht wurde. Statt künstliche Stimuli wie vorangehende Studien zu verwenden, nutze diese Studie menschliche Bewegungen, welche möglicherweise eine höhere Plausibilität bieten, mit der Wahrnehmungserfahrung der Teilnehmende übereinzustimmen.

Studie 3 setzte den Weg der modalitätsrelevanten Effekte für das intrinsische Uhrenmodell fort, indem sie die wahrgenommenen Dauern und Emotionen zwischen visuellen (unimodalen) und audiovisuellen (bimodalen) Präsentationen verglich. In dieser Studie beurteilten die Teilnehmenden die Dauer, die emotionale Valenz und die Erregung von tempoverschobenen Bewegungen (dieselbe Basis wie in Studie 2). In Studie 3 wurden keine signifikanten Unterschiede in der Modalität festgestellt, was darauf hindeutet, dass der bimodale Vorteil bei der zeitlichen Verarbeitung, der in einigen früheren Studien festgestellt wurde, bei der Bewertung ökologisch valider Bewegungen möglicherweise nicht gegeben ist. Dies könnte auf die hohe Salienz der visuellen Reize zurückzuführen sein, was wiederum den in Studie 2 gefundenen Effekt der visuellen Dominanz unterstützt. Darüber hinaus zeigten die Ergebnisse, dass emotionale Erregung und Valenz von Unterschieden in der Bewegungskomplexität und -fluidität abhängig waren, wenn Stimuli mit unterschiedlichem Originaltempo in derselben Geschwindigkeit präsentiert wurden. Dies deutet darauf hin, dass das Tempo als wichtiger Faktor für die interne Taktrate nicht nur die bloße Frequenz von Ereignissen ist, sondern ein vielschichtiges Konzept, das noch weiter erforscht werden sollte.

In Studie 4 wurde ein musikalischer Kontext mit hoher ökologischer Validität untersucht, in dem ein Schlagzeuger Rhythmen mit unterschiedlichem Tempo und unterschiedlicher Komplexität als Stimuli einspielte. Verglichen wurden die Zeitleistungen, als die Teilnehmer die Stimuli nur beobachteten und als sie frei und isochron zu den Stimuli klopfen. Die Ergebnisse zeigten eine signifikant schnellere PoT und eine kürzere Dauer (bei langsamen und mittleren Tempi) beim Klopfen im Vergleich zu keinem Klopfen. Hier schien die motorische Beteiligung die Aufmerksamkeit von der Zeitbeurteilung abzulenken, was darauf hindeutet, dass die Aufmerksamkeit eine Schlüsselkomponente beim Supersekunden-Timing ist. Das Auftreten von Tempo-Effekten in Studie 4 bei der PoT als auch bei der Dauereinschätzung, im Vergleich zum fehlenden Effekt in Studie 3, könnte bedeuten, dass naturalistische Aufführungsszenen, die ökologisch valider waren als im Labor entworfene Bewegungen, möglicherweise auch mehr Aufmerksamkeit erhielten, was gemäß der Schrittmacher-Zähler-Theorie zu einem größeren Ausmaß an Veränderungen in der internen Uhrgeschwindigkeit führte. Es wird vermutet, dass ökologische Validität als wichtiger Umweltfaktor die Aufmerksamkeitszuweisung bei der zeitlichen Verarbeitung beeinflusst. Allerdings fehlt der existierenden intrinsischen Uhr die Erklärungskraft hinsichtlich der Rolle der Aufmerksamkeit bei der Supra-Sekunden-Taktung wie beim zentralen Uhrenmodell. Die Diskrepanz erfordert daher zukünftige Untersuchungen eines Timing-Mechanismus, der mit der Modalitätsspezifität und den kognitiven Prozessen, die für zeitliche Urteile im Sekundenbereich unerlässlich sind, kompatibel ist.

Insgesamt haben die Ergebnisse der vorliegenden Dissertation zum weiteren Verständnis der internen Uhr beigetragen, die im Zentrum der Zeitwahrnehmung in einem audiovisuellen Kontext steht. In Studie 1 wurde eine Modalitätsspezifität identifiziert, die im Wesentlichen zwischen dem zentralen und dem intrinsischen Uhrenmodell unterscheidet.

Modalitätseffekte, die in Studie 2 gefunden wurden, lieferten weitere Belege für eine intrinsische Uhr in der Supersekunden-Zeitskala. Unter Verwendung der gleichen Bewegungsreize zeigte Studie 3 das Fehlen eines bimodalen Vorteils bei der zeitlichen Verarbeitung im Vergleich zu unimodalen Reizen, was möglicherweise einem visuellen Dominanzeffekt mit ökologisch validen Stimuli entspricht. In Studie 4 wurden Belege für die Beteiligung der Aufmerksamkeit an der Beurteilung von PoT und Dauer im Sekundenbereich gefunden, zusätzlich zu einem möglichen Effekt der ökologischen Validität auf die Geschwindigkeit der internen Uhr. Die Beobachtungen haben Auswirkungen auf die Entwicklung von Modellen der inneren Uhr sowie auf passive und proaktive Strategien, die

die subjektive Zeitwahrnehmung bei alltäglichen Hör- und Seherlebnissen beeinflussen könnten.

## List of Publications and Author Contributions

The following scientific articles are included in this dissertation. The first author for all papers listed below was the primary contributor in study design, stimuli formulation, experiment implementation, data collection, analyses, results interpretation, tables and plots preparation, and manuscript writing. For all 4 studies, Prof. Dr. Clemens Wöllner provided the initial ideas and guidance in every research stage. For Study 2, guidance on statistical analyses was provided by Prof. Dr. Zhuanghua Shi from Ludwig-Maximilians-Universität München, while David Hammerschmidt provided advice on stimuli creation and experiment implementation. For Study 3 and 4, Dr. Birgitta Burger assisted in stimulus creation by designing the rhythms, finding the drummer, and producing the videos via the MoCap system and MATLAB. She also provided guidance on statistical analyses and writing. Sebastian Schwarz as research assistant and Dominik Leiner from the SoSci Survey provided help in implementing the experiments for Study 3 and 4. All studies were funded by a Consolidator Grant from the European Research Council to the third author. The research is part of the five-year project: “Slow motion: Transformations of musical time in perception and performance” (SloMo; Grant No. 725319). The German translation of the short summary of this dissertation was checked and revised by Dr. Birgitta Burger, Frithjof Faasch, and Adrian Schneiders.

1. Wang, X., & Wöllner, C. (2020). Time as the ink that music is written with: A review of internal clock models and their explanatory power in audiovisual perception. *Jahrbuch Musikpsychologie*, 29, e67. <https://doi.org/10.5964/jbdgm.2019v29.67>
2. Wang, X., Wöllner, C., & Shi, Z. (2021). Perceiving Tempo in Incongruent Audiovisual Presentations of Human Motion: Evidence for a Visual Driving Effect. *Timing & Time Perception*, 10(1), 75-95. <https://10.1163/22134468-bja10036>
3. Wang, X., Burger, B., & Wöllner, C. (2024). Body movement and emotion: Investigating the impact of audiovisual tempo manipulations on emotional arousal and valence. *Jahrbuch Musikpsychologie*, 32, e191. <https://doi.org/10.5964/jbdgm.191>
4. Wang, X., Burger, B., & Wöllner, C. (2023). Tapping to drumbeats in an online experiment changes our perception of time and expressiveness. *Psychological Research*, 88, 127-140. <https://doi.org/10.1007/s00426-023-01835-7>

## **Statutory Declaration/ Eidesstattliche Erklärung**

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare on oath that I have written the present dissertation independently and have not used further resources and aids than those stated.

Hamburg, 23.03.2022

Stadt, den | City, date

A handwritten signature in black ink, appearing to read 'Wang Xinyu', written in a cursive style.

Unterschrift | Signature

## Study 1

**Time as the ink that music is written with:  
A review of internal clock models and their explanatory power in  
audiovisual perception**

Xinyue Wang and Clemens Wöllner

*Jahrbuch Musikpsychologie, 29, e67*

<https://doi.org/10.5964/jbdgm.2019v29.67>

Reproduced with kind permission from Jahrbuch Musikpsychologie

## Forschungsberichte zum Themenschwerpunkt

### **Time as the Ink That Music Is Written With: A Review of Internal Clock Models and Their Explanatory Power in Audiovisual Perception**

Zeit als Grundlage der Musik: Ein Überblick zu Modellen innerer Uhren und deren Erklärungswert für die audiovisuelle Wahrnehmung

Xinyue Wang\*<sup>a</sup>, Clemens Wöllner<sup>a</sup>

[a] Institut für Systematische Musikwissenschaft, Universität Hamburg, Hamburg, Germany.

#### **Abstract**

The current review addresses two internal clock models that have dominated discussions in timing research for the last decades. More specifically, it discusses whether the central or the intrinsic clock model better describes the fluctuations in subjective time. Identifying the timing mechanism is critical to explain and predict timing behaviours in various audiovisual contexts. Music stands out for its prominence in real life scenarios along with its great potential to alter subjective time. An emphasis on how music as a complex dynamic auditory signal affects timing accuracy led us to examine the behavioural and neuropsychological evidence that supports either clock model. In addition to the timing mechanisms, an overview of internal and external variables, such as attention and emotions as well as the classic experimental paradigms is provided, in order to examine how the mechanisms function in response to changes occurring particularly during music experiences. Neither model can explain the effects of music on subjective timing entirely: The intrinsic model applies primarily to subsecond timing, whereas the central model applies to the suprasecond range. In order to explain time experiences in music, one has to consider the target intervals as well as the contextual factors mentioned above. Further research is needed to reconcile the gap between theories, and suggestions for future empirical studies are outlined.

*Keywords:* internal clock models, Dynamic Attending Theory, Scalar Expectancy Theory, music perception, audiovisual timing

#### **Zusammenfassung**

Dieser Überblick befasst sich mit zwei Modellen der inneren Uhr, die in den letzten Jahrzehnten die Diskussion in der Forschung zur Zeitwahrnehmung und -gestaltung bestimmt haben. Insbesondere wird diskutiert, ob das zentrale oder das intrinsische Uhrenmodell Schwankungen der subjektiven Zeit besser erklärt. Dabei ist das Erkennen des zugrundeliegenden Mechanismus entscheidend, um das Zeiterleben im Nachhinein zu erklären oder in verschiedenen audiovisuellen Kontexten vorherzusagen. Musik zeichnet sich durch ihre Bedeutung in realen Szenarien sowie durch ihr großes Potenzial zur Veränderung des subjektiven Zeiterlebens aus. Musik kann als komplexes dynamisches Audiosignal die zeitliche Genauigkeit beeinflussen. Dies ist der Hintergrund, verhaltensbezogene und neuropsychologische Belege zu diskutieren, die eines der Uhrenmodelle oder beide unterstützen. Neben den Zeitmechanismen wird ein Überblick auf interne und externe Variablen wie Aufmerksamkeit und Emotion, sowie auf klassische experimentelle Paradigmen gegeben. Dadurch wird dargelegt, welche Rolle den Mechanismen zukommt hinsichtlich der Reaktion auf Änderungen im Stimulusmaterial, insbesondere beim Erleben von Musik. Im Ergebnis kann kein Modell die Auswirkungen von Musik auf das subjektive Zeiterleben vollständig erklären. Während das intrinsische Modell in erster Linie das Zeiterleben für sehr kurze Dauern unterhalb einer Sekunde zu erklären vermag, bietet das zentrale Modell einen höheren Erklärungswert für den Suprasekundenbereich, das heißt für das Timing von Sekunden bis Minuten. Um Zeiterfahrungen in der Musik zu erklären, müssen die Zielintervalle sowie die oben genannten Kontextfaktoren berücksichtigt werden. Weitere Forschungen sind erforderlich, um die Kluft zwischen den Theorien zu schließen, wobei Vorschläge für künftige empirische Studien skizziert werden.

*Schlüsselwörter:* Innere Uhrenmodelle, Dynamic Attending Theory, Scalar Expectancy Theory, Musikwahrnehmung, audiovisuelles Zeiterleben

Jahrbuch Musikpsychologie, 2020, Vol. 29: Musikpsychologie — Musik im audiovisuellen Kontext, Artikel e67,  
<https://doi.org/10.5964/jbdgm.2019v29.67>

Eingereicht: 2019-09-30. Akzeptiert: 2020-05-08. Publiziert (VoR): 2020-07-01.

Begutachtet von: Wolfgang Auhagen; Günther Rötter.

\*Korrespondenzanschrift: Institut für Systematische Musikwissenschaft, Universität Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany. E-Mail: [xinyue.wang@uni-hamburg.de](mailto:xinyue.wang@uni-hamburg.de)



Dieser Open-Access-Artikel steht unter den Bedingungen einer Creative Commons Namensnennung 4.0 International Lizenz, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.de>). Diese erlaubt für beliebige Zwecke (auch kommerzielle) den Artikel zu verbreiten, in jedwedem Medium zu vervielfältigen, Abwandlungen und Bearbeitungen anzufertigen, unter der Voraussetzung, dass der Originalartikel angemessen zitiert wird.

Properties of time have attracted the interest of researchers since long. From a Newtonian perspective, time is seen as an arrow flying eternally forward, whereas for classical thermodynamics, the passage of time shares similarity with the irreversible increase of entropy, or the degree of disorder in the universe (Lieb & Yngvason, 1999). The psychological study of time perception, in comparison, held different opinions. One of the earliest efforts in capturing an internal timing system stemmed from doctors' skills in making accurate estimates of time based on heartbeats and breathing (Goudriaan, 1921). Composers, as experts of time and timing, have discovered the link between the tempo in which their works are played and the impression of durations among the audience. Ravel once complained to Furtwängler that his Boléro, when played too fast, would feel unjustifiably long (Nichols, 2011). Ravel's somewhat paradoxical observation comes close to that of cognitive scientists, such that music played at various tempi could induce corresponding time distortions (e.g., Droit-Volet et al., 2010).

Much as Ravel perceived, music is a form of art closely intertwined with time. A rich vein of literature has pointed out that rhythmic patterns, or beats, are fundamentally embedded in all genres of music, leading to perceptual periodicities (London, 2004; Nozaradan, 2014). The perceived periodicities provide a sequence of external events, which could subsequently be internalized as representations of time (Droit-Volet et al., 2013; Gibbon et al., 1984; Jones & Boltz, 1989). Music- or beat-induced movements are noticed across a wide range of ages as the substantiation of an individuals' anticipation of the rhythmic patterns – among infants (Zentner & Eerola, 2010), pre-schoolers (Eerola et al., 2006), and adults (Burger et al., 2018). By either proactively or passively synchronising to the musical beats, an individual's perceptual time is subject to modifications. The general tendency indicates that fast music leads to duration overestimation, whereas slow music to underestimations (Droit-Volet et al., 2013; Wang & Shi, 2019).

In this review, timing refers to the active process of monitoring temporal order by explicit (e.g., tapping) or implicit (e.g., silent counting) actions, with an emphasis on the efforts involved in the task. Meanwhile, duration estimation, the action of gauging past time, composes the second part of time perception. It encompasses the concepts of both prospective (knowing before that the duration of an event should be judged) and retrospective timing (judging duration afterwards) after the target duration has elapsed (Zakay & Block, 1995). In this sense, one could passively experience the passage of time with or without attending to the passage of time itself, which affects subsequent judgments.



In relation to duration estimation, it is equally important to underline the role of beat perception, or inner timing as the ability to perceive and predict the temporal location of events. As a supporter of the cerebral clock model, Pöppel (1989) hypothesized that maintaining a constant tempo in music production, especially in classical music, has to do with a time keeping mechanism that functions mainly by tracking the temporal order such as synchrony and succession in addition to measuring durations. In the same vein, a “3-second window of temporal integration” (Pöppel, 1989, p. 86) was assumed to constitute the psychological present. This has consequences for perceiving musical tempo and the integration of beats, and hence subjective experiences of time. Similar attempts for developing clock models were based on beat perception (Langner, 2002; Schulze, 1978). Schulze, in particular, pivoted the *Dynamic Attending Theory* (Jones & Boltz, 1989) that emerged later by emphasizing the variability of internal clock speed under the influence of environmental cues (accelerating and decelerating beat patterns).

While some studies investigated musical tempo in order to formulate hypotheses about clock models, other research found that the variables embedded in tempo, such as isochrony, salience, or complexity, directly evoked changes in the functioning of the internal clock. Povel and Essens (1985) observed in their experiments that different grouping of rhythmic beats led to various temporal reproductions, giving rise to a best fit for the internal clock. Explanations lie in the coupling of beat accents and the clock ‘tick’: the stronger the beat pattern is (in this case, higher metrical level), and the less complex the metrical structure is, the more likely the beat would activate the internal time recording system and be represented in temporal processing. Recognizing beat perception not only helps understanding human timing better, but also particularly with music listening and performing, which can also be understood in terms of proactive and active timing. In fact, frequent exposure to musical beat production appears to enhance one’s temporal sensitivity, and this effect may transcend to other sensory modalities (from audition to vision; Cicchini et al., 2012). Apart from external training, the stability of intrinsic rhythm was also positively correlated with tempo reproduction performances (McPherson et al., 2018). It is therefore essential to look into the complexity in musical tempo itself.

Musical tempo is subject to ambiguity. The complexity of tempo structures in music has long been recognized (e.g., Pressnitzer et al., 2011). It is not only marked by the number of note events in a melody (Behne, 1976), nor only by the patterns of percussion instruments, but rather by changes in pitch, timbre, or loudness (Brochard et al., 2003), as well as phrasing and articulation (Auhagen & Busch, 1998). Multiple sound sources of the instruments in a symphony orchestra vary tremendously across different sections and therefore constitute auditory streams that are hard to disentangle (Shamma & Micheyl, 2010), especially for non-musicians. Note that the difficulty of correctly identifying temporal structures in music is not equal to that of correctly identifying the tempo of music, considering the latter has more to do with detecting the absolute ‘speed’ and tempo changes. Attempts have been made to examine the thresholds of detecting musical tempo acceleration and deceleration, for instance, among musically trained and untrained groups (e.g., Ellis, 1991). There are several assumptions of how we cope with “noisy” auditory signals in terms of time and tempo perception. Some argued that the process of tempo extraction depends mainly on periodic regularities (McDermott et al., 2011), while others emphasized the importance of learning, regardless of tempo structure complexities (Agus et al., 2010).

A small number of studies aimed at the disentanglement of auditory rhythmic features and revealed how tempo salience affects perceptual time. One study investigated different metrical levels and found effects on listeners’ sense of time (Hammerschmidt & Wöllner, 2020). More specifically, the lower the metrical level individuals at-

tended to by tapping (e.g., eight notes versus half notes), the longer a music excerpt was perceived, providing some evidence for the impact of event density (cf. Behne, 1976). In this case, the count of time was affected by the number of beats registered in memory.

Apart from music, inputs from other sensory modalities may also affect temporal processing. Indeed, psychological research has often used visual stimuli such as flashes or flickering lights to investigate time. For instance, studies of the entrainment effect for independent modalities showed that the presence of either visual flickers or pure tones led to higher entrainment (e.g., Ortega & López, 2008; Treisman & Brogan, 1992; Treisman et al., 1990). The effect, nevertheless, is not limited to one modality. Past research suggests that auditory signals of various complexities could enhance the entrainment effect for visual sequences and, in some cases, were transferable to the attention acuity of the other modality (Bolger et al., 2013; Escoffier et al., 2010). In Bolger et al.'s study, participants were able to perform equally well in a target detection task regardless of the target modality (auditory or visual) when entrained with tone sequences. Another case in point is the cross-modal transfer in tempo discrimination between auditory and tactile domains, where training with rhythmic sounds led to enhanced performance in that of the latter (Nagarajan et al., 1998). These studies provide evidence that the cognitive processes involved in timing and time perception should function at a domain-general level.

In this review, an overview is provided of internal clock models that were established or further developed in recent years. In particular, the aim was to show how each model accounts for the experience of musical time in auditory and audiovisual contexts. We will tackle questions such as: How does music facilitate temporal processing? What are the timing mechanisms and models, and how do they explain the inference between music and perceptual time, respectively? What are the implications of studying music and time perception?

## The Internal Clock

Comparable to an actual clock, the internal clock has been an analogy for the timing mechanism in human and animals (Eagleman et al., 2005; Ivry & Schlerf, 2008). The temporal order of events is recorded by multiple sensory modalities and processed in, according to different theories, a variety of pathways before becoming representations of time, that is, the occurrence of "clock ticks". Early in the discussion, hypotheses stated that time perception was a form of information processing that highly depended on the recording capacity (Ornstein, 1969). Researchers such as Barry (1990) and Schulze (1978) both emphasized the importance of music as an environmental construct of attention that shaped both the perceived time (in terms of its duration) and the passage of time (the perceived speed).

In the past years, two major theories were on the forefront in discussions as to how the temporal units are recorded, both postulating the presence of a specific cognitive module dedicated to timing. The 'no clock' hypothesis, or state-dependent network, and the 'central clock' hypothesis have both received increasing attention in research (Grondin, 2010b). The latter, in particular, encompasses two theories: The Dynamic Attending Theory, based on a non-linear cumulation of temporal units, as well as the Scalar Expectancy Theory, which assumed that the emission of temporal pulses follows a linear approach (Ivry & Schlerf, 2008). Such as Stern (1897) had already pointed out for "Präsenzzeit" (the experienced present moment) and the time range for other cognitive processes, it appears that different theories function best at specific time ranges (Figure 1).

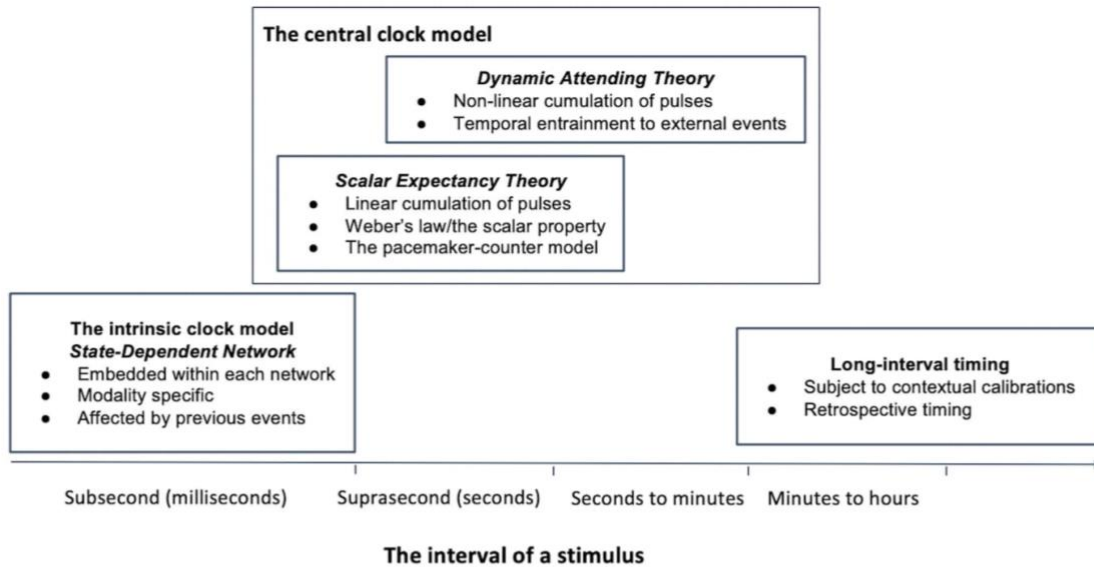


Figure 1. An overview of the internal clock models specified by interval ranges (subsecond, suprasecond, seconds to minutes, and minutes to hours) as well as by the division of central vs. intrinsic model.

Note. The most important features described in the overview were discussed in detail in the following sections. For a review of the research methods adopted in timing studies, see Grondin (2010b).

### The Intrinsic Clock Model

Unlike the traditional view of a clock, some researchers believe that there might be no clock at all. Such a 'no-clock' model is known as a state-dependent timing system or intrinsic model (Ivry & Schlerf, 2008). The 'state' here describes the specific circumstances generated by a neural network in response to external changes. Timing is seen as an implicit function of each neural network that is activated for a given sensory modality and is sensitive to pre- and post-interval changes. It is postulated that activity-elicited changes in neural networks directly reflect the inherent temporal structures and therefore serve as references for timing in sub-second intervals (e.g., Karmarkar & Buonomano, 2007). The process is also referred to as "a temporal-to-spatial transformation" (Karmarkar & Buonomano, 2007, p. 3), or the "intrinsic model", as researchers hypothesize that timing is an integral function of neural activity (Ivry & Schlerf, 2008).

The model essentially suggests that timing as a function is distributed to a variety of neural structures in which oscillatory patterns stay consistent, known as the recurrent neural network (Buonomano & Laje, 2011). Ramping, or climbing activities in neural oscillations ranging from primarily low frequency gamma band to higher frequency such as beta and alpha band (e.g., Wittmann, 2013) have been revealed as the physiological basis for the model, in addition to neural spikes across a wide range of brain regions such as the striatum (Gu et al., 2015). The time stamps, or accumulated states, are hypothesized to be expressed on both micro (individual neurons) as well as macro (populatory neuron excitation/inhibition) levels (Buonomano & Laje, 2011).

The intrinsic model proposes the possibility that timing is an inherent function of multiple dynamic neural networks. The flexibility allows the network to take into account calibrations towards previous durations and to judge the duration of the current event on this basis. Furthermore, by hypothesizing the implicitness of timing, there is no need for external triggers even when an event is absent. However, the state-dependent network also has its shortcomings. Studies suggested that the model is only applicable to the subsecond range. That is to say, the cumulative effects of previous events, either enhancing or reducing one's temporal sensitivity, diminished within up to 300 milliseconds (Buonomano et al., 2009). On the other hand, it does not offer a clear explanation of cross-modal temporal information integration. This is where the central clock model provides a useful alternative perspective.

### The Central Clock Model

The central timing mechanism, also known as the dedicated clock model (e.g., Allman et al., 2014), stemmed from Treisman's (1963) work. Decades of research into the human timing mechanism were based on this model and have assumed that timing is a specific cognitive module, hypothetically located across the global neural network (e.g., Allman & Meck, 2012).

Where is the 'clock' in our brain? Neurological studies supporting the timing mechanism as an independent cognitive module that is dedicated solely to this function, however, do not necessarily assume one single structure in the brain for it. Research rather supports the roles of a wide range of brain regions working collaboratively in order to process time (e.g., Buhusi & Meck, 2005). The cerebellum, for example, is involved in short duration judgments, arguably from a few hundred milliseconds to 30 seconds (Allman & Meck, 2012). Disruptions to other cortical and subcortical structures including the basal ganglia can lead to timing deficits also in larger time frames. In Schwartz and colleagues' (2011) study, participants with basal ganglia lesion failed to detect tempo acceleration and deceleration of tones and could not entrain tapping movements with the signals. The evidence implicates a global network where multiple brain regions are involved; impairments at any part of the chain could possibly lead to malfunctions of the timing mechanism. There are two prominent theories with consequences for time processing in music that will be explained below.

### The Dynamic Attending Theory (DAT)

DAT, also known as the oscillator model, hypothesizes that the ability to estimate the duration of past events depends on the coupling between attentional pulses and the occurrences of external events (Jones & Boltz, 1989). The theory supports the presence of a central clock, in the sense that the allocation of attention as a limited resource is based on the expectation of the next event on the timeline (Large & Jones, 1999). An exogenous stimulus, when aligned with the peak of attention, is best retained in working memory and transformed into representations of time (Barnes & Jones, 2000). Essential to this theory is that, like the unstable periodicities of external events, the emission of attentional pulses or oscillations is a non-linear process (Large, 2008).

Holding the central clock premise, DAT suggests that attention plays a critical role in regulating the frequency of the pacemaker pulses according to the Attentional Gate Model (Block & Gruber, 2014; Zakay & Block, 1995). More specifically, the "gate" through which temporal units pass before registering with the counter device opens wider when more attention is assigned to the specific time point. When one's attention is shifted elsewhere irrelevant to temporal cues, fewer pulses are recorded, leading to duration underestimation. Some argue that

DAT applies exclusively to prospective timing, while in retrospective timing, it is subject to contextual influences and memory retrieval (Block & Gruber, 2014; Gu et al., 2015).

Note that DAT is hypothesized to function mostly within the suprasecond range, because prospective timing recedes with time due to limited capacities of working memory (WM) (for a review, see Gu et al., 2015). Concurrent tasks that require extra attentional resources could reduce timing accuracy (Brown & Boltz, 2002). Polti et al. (2018) attempted to explore the interval boundary of attention in prospective timing and found that the magnitude of WM interference on time estimation tasks increased proportionally with interval lengths (30 to 90s). In a more naturalistic setting, gamers were asked to estimate the elapsed time (12, 35, or 58 minutes) either knowingly (prospective timing) or not (retrospective; Tobin et al., 2010). The 12-minute session was estimated significantly longer in the prospective than the retrospective paradigm, while estimation differences were less pronounced in 35- and 58-minute conditions, suggesting that DAT's predictive power may be reduced in longer intervals. However, no evidence so far has made a clear cut of the interval ranges where each model fits best. This question clearly deserves further exploration and will be discussed in the conclusion.

### The Scalar Expectancy Theory (SET)

Another well-known model for the internal timing mechanism argues that the perceived amount of time is composed of regularly emitted pulses from a pacemaker, as an analogy of an internal clock, and accumulated by a counter device, therefore also known as the pacemaker-counter model (Gibbon, 1977; Gibbon et al., 1984; Treisman, 1963). This model specifies that the temporal process is accomplished through roughly three different steps: the clock, the memorisation, and the judgment of time. Accordingly, in order to explain the temporal flow, SET proposes that subjective time is composed of (a) the representation of objective durations, and (b) the estimation variance or error rate by Weber's fraction (Allman & Meck, 2012; Grondin, 2010b). The variance is hypothesized to stem from transferring the clock readings to working memory (Meck, 1984; Treisman, 1963). Inaccuracy in duration estimation, according to SET, is also subject to the influence of attention, clock speed error, task switch, decision error and other factors (Allman & Meck, 2012). The longer the duration, the larger the variance. It is, however, observed that SET has a controversial applicability in the sub-second to supra-second range. Grondin (2010a) focused on the violation of scalar property when examining participants' timing performance in a subsecond range, and found a tendency for Weber's fraction increasing as the interval approached 1s. This suggests that SET does not provide a powerful explanation of timing behaviour in the millisecond range. The applicability of SET in the millisecond to second range was further supported by the audiovisual evidence for this theory. Evidence is still needed for a clear boundary of 'time ranges of the best fit' for SET.

From a neurobiological perspective, the striatal beat frequency (SBF) theory offers an explanation for the original pacemaker-counter model (Miall, 1989; van Rijn et al., 2014). Unlike the latter, SBF theory instantiated the biological structure of the clock pulses as the oscillations of striatal medium spiny neurons (MSN), locating at the suprachiasmatic nucleus at the anterior hypothalamus. Researchers proposed that the neurons at different oscillatory frequencies reset when the timing begins, receiving inputs from cortical neurons firing as the consequences of dopaminergic releases (Merchant et al., 2013). Detection of the synchronous neural oscillation is known as the coincidental detection (Buhusi & Meck, 2005). The MSNs are capable of detecting coincidental oscillations from the cortical neurons that fire at similar frequencies, also known as the input, then translate to temporal units as the output. To justify the scalar property, that is variance in accumulative timing, it's been proposed that neural oscillations phase out and disperse into the inherent frequency of each neuron

after the initial alignment. As a result, the discrepancy increases proportionally until the neurons that were firing together completely desynchronize in the end. The MSNs, however, retain the robust ability to detect temporal patterns up to minutes despite the complexity of the inputs thanks to ironically the large number of cortical neurons (e.g., [Matell et al., 2003](#)), making the theory viable for a wider range of durations.

## Factors Overarching Both Clock Models

### Sensory Modalities

Multisensory inputs often interact with one another in our daily life. A vase dropping to the ground is usually followed by a shattering sound. A knock on the door leads to a knocking sound. To a broader extent, signals from vision, hearing, touch, smell and taste constitute the intangible framework of timing references together. Hence it is critical to understand the specificity of each sensory modality and their joint effects in temporal perception.

The dominant role of audition in temporal processing has been evidenced by a series of studies (e.g., [Boltz, 2017](#); [Chen et al., 2018](#); [Repp & Penel, 2002](#)). A number of studies supported higher precision in temporal discrimination in audition compared to vision (e.g., [Large, 2008](#); [Phillips & Hall, 2002](#)). Furthermore, auditory temporal processing is capable of interfering with visual timing. In this case, participants' performances in identifying the correct rhythmic visual patterns were most heavily compromised when the task was accompanied by a new string of isochronous sounds rather than visual display ([Guttman et al., 2005](#)). One may assume that temporal information derived from auditory events weighs more than that of visual inputs. The auditory dominance view is, however, not without dispute. [van Wassenhove and colleagues \(2008\)](#) found that incongruent visual displays could distort temporal perception of auditory information in both directions. A recent finding, in addition, suggests that temporal perception was biased towards the visually perceived tempo of natural human movements rather than that of the drumbeats when the two sensory modalities were incongruent ([Wang et al., 2019](#)).

Research showed that auditory stimuli can effectively distort visual perception (e.g., [Burr et al., 2013](#)). The auditory driving effect emphasizes the perceived coupling of fluttering sounds to visual flicker rates, if the temporal gap between the flutter and the flicker does not exceed a certain range ([Shipley, 1964](#)). In other words, perceptual integration is accomplished by averaging auditory and visual input frequencies while endowing more weights on the former. A robust auditory driving effect could be observed when the sounds were presented as a brief distractor ([Burr et al., 2013](#)). Furthermore, [Chen and colleagues' \(2018\)](#) study suggested that, in addition to traditional regular flutters, irregular auditory inputs accompanying the visual flickers could also lead to distortions in perceiving the latter. Similar observations of the audiovisual bias were reported as fission and fusion illusions ([Shams et al., 2002](#)). In this case, the former specifies two visual events perceived as one when presented simultaneously with a beep, while in the latter, one flash is perceived as two when accompanied by two beeps.

Therefore, it is inevitable to take into account the arguably dominant position of the auditory modality when exploring the role of music in temporal processing. It should be noted that music encompasses not only complex acoustic signals, but a rich source of emotions that alter subjective time. Films are an example of how music shapes experiences of time. In a study investigating slow motion film scenes as compared to the

same scenes played back in real time, participants were significantly influenced in their temporal judgments of the scenes' duration when music was present (Wöllner et al., 2018). While slow motion scenes led to an underestimation of time, the same scenes in real time seemed to last relatively longer, and music yielded more accurate time estimations. Furthermore, music led to higher physiological arousal and larger pupil diameters in observers, suggesting that music modulates emotional responses and experiences of time in audiovisual scenes.

### Working Memory and Attention

Central to the SET is the memory stage, in which working memory is retained, and the judgment stage, in which the current count of temporal units is compared to references retrieved from long-term memory (Gibbon et al., 1984; van Rijn et al., 2014). Individual differences in short-term memory capacity and discrepancies in timing performances bring attention to the role of working memory in temporal processing (e.g., Broadway & Engle, 2011). More specifically, higher working memory capacities imply higher potential to hold more time units at the second and third stage of timing, thus leading to more precision (Teki & Griffiths, 2014).

Working memory is positively related to other executive functions such as selective and divided attention (Colflesh & Conway, 2007), for both auditory and visual modalities (Wöllner & Halpern, 2016). Both shift in weights in various timing scenarios. This is particularly relevant for understanding the different mechanisms behind prospective and retrospective timing, as mentioned before (Block & Zakay, 1997). In the oscillator model (DAT), attentional pulses are emitted in order to track external beats. These pulses are recorded and transferred to working memory before entering the stage of comparison with a reference duration in long-term memory (Block & Zakay, 1997; Gibbon et al., 1984). Attention diverted from the timing task results in fewer temporal units taken into the count and consequently underestimations of time, while attention directed to timing led to overestimations regardless of test durations (Polti et al., 2018). Despite a lack of evidence, we hypothesize a similar result with music listening. When instructed to time a piece of music before it commences, a listener processes the passage of time differently than when asked to estimate the time elapsed at the end of the excerpt.

The interpretation of the roles of WM and attention also depends on the theories. DAT, compared to SET, highlights the role of attention rather than working memory (e.g., Jones, 2010; Jones & Boltz, 1989). It postulates that attention, when quantified as regular emitted pulses, could synchronize with external periodicity and therefore serve as a reference for time. The periodicities of external events, that is regular or irregular patterns, do affect the strength of their synchronisation with attentional pulses. The more predictable an exogenous pattern is, the better the effect, known widely as the temporal entrainment effect (Barnes & Jones, 2000; Schroeder & Lakatos, 2009). This has been evidenced by a number of visual (Cravo et al., 2013), auditory (Barnes & Jones, 2000; Jones, 2010), and movement (Burger et al., 2018) studies. Jones (1981, 1990) proposed that the characteristics of the information, in this case musical expressions, could distort the perception of time. Empirical studies supporting her claims found that, for instance, music was perceived to be slower when there were more pitch variations and inconsistent metrical accents (Boltz, 1998). We may predict that music genres with more predictable rhythms such as pop and rock, compared to those with less predictability such as Jazz, are associated with higher duration sensitivity and better timing accuracy.

## Emotions

“Time flies when you are having fun”. Understanding the nature of emotions in time perception is important to comprehend how music distorts subjective time, as it essentially conveys a wide spectrum of emotions. The relatively small number of studies that have directly looked into the effects of musical emotions on subjective time show that information of strongly emotional contents were more engaging and were subsequently better processed and stored in WM, leading to time overestimation (for a review, see Schäfer et al., 2013). Music as a powerful tool to induce emotions was found to induce a sense of timelessness (duration overestimation) as well as faster passage of time when an individual is completely submerged in the experience (Herbert, 2012). Apart from the aesthetic pleasure, other types and intensity of emotions may also have an impact on how music could distort the perception of time. The reasons may lie in the psycho-physiological arousal levels. Higher arousal level is believed to cause time overestimation (e.g., Droit-Volet et al., 2013). A group of participants, for instance, were presented with emotional film excerpts to induce corresponding emotions in them (Droit-Volet et al., 2011). Results indicate that, compared to baseline temporal judgments, participants tended to overestimate the durations after watching scary films. There are nevertheless findings implying the opposite, that is, higher emotional arousal leads to duration underestimation especially from a retrospective point of view (Herbert, 2012).

Another line of studies investigates the impact of emotional valences on temporal processing. Positive emotions, substantiated by happy music, led to duration underestimation, while negative emotions in sad music to duration overestimation with retrospective paradigms (Bisson et al., 2008). It was speculated that the positive emotions gave rise to less contextual changes than did the negative, therefore registering fewer events in the memory. Some evidence, on the contrary, implies that valence does not matter. Further investigations showed that highly arousing emotional pictures accelerated the internal clock speed and caused a leftward shift in the reaction time compared with pictures of low emotional arousal, regardless of its valence (Droit-Volet & Berthon, 2017).

The seemingly puzzling observations may be explained by the mechanism by which emotions take effect on time perception. One approach is rooted in the emission rates of attentional pulses, which can be moderated by the affective states, especially the arousal level. According to the pacemaker-counter model (Treisman, 1963), more attentional pulses are emitted when the arousal level is high, and subsequently be recorded as the sum of clock ticks, that is, the perceived duration. Attention could either facilitate or hinder the interaction between emotions and temporal processing. More specifically, when attention is allocated to sustaining the temporal units, the effect would lead to duration overestimation. In contrast, when attention is shifted from temporal information to the emotionally charged event, fewer ticks are accumulated, resulting in duration underestimation.

## Modality-Specific Evidence for the Internal Clock Models

### Audiovisual Evidence for the Intrinsic Clock Model

Time-dependent neural oscillations are specific to sensory modalities. Studies have revealed that neuron excitation and inhibition could be elicited according to a specific type of sensory input, such as sound (Schnupp et al., 2006) and visual flicker (Burr et al., 2007). Researchers found that the time-dependent decodability of



visual objects with MEG in a window of 1000ms varied significantly, suggesting that time might be an inherent feature in the local visual network (Carlson et al., 2013). Furthermore, transcranial magnetic stimulation studies revealed that auditory timing could be dissociated with that in other sensory modalities (Buetti et al., 2008), as participants performed worse in duration discrimination task (pure tones, 10 to 40ms) when receiving disruptions in the auditory cortex. We might as well propose that, when listening to complex auditory signals such as music, particular groups of neurons in the human auditory cortex generate time-dependent responses, which simultaneously serve as time codes. However, relatively few studies with humans have directly confirmed the time-dependent variability of the local auditory network (Toiviainen et al., 2019).

The disassociation in timing abilities among different sensory modalities also showed that time is processed as a local flow of information. Early findings entail significantly higher timing precision with hearing than with vision (e.g., Penney et al., 2000), indicating a superiority of audition over vision in providing temporal cues. Timing is a highly selective, localized process even within one modality. Burr and colleagues (2007) successfully modulated the perceived durations of the target visual stimuli by manipulating the apparent rate of flickers in a confined retinal region. Their finding is among one of the first to empirically support (a) the spatial-temporal connection in neural representations, and (b) the modality specificity in temporal processing, particularly the superiority of audition (e.g., Repp & Penel, 2002). In Lustig and Meck's (2011) study, the modality effect was stronger for participants at both ends of the age spectrum. One potential cause was that older adults were more susceptible to varying allocation of attention under different experimental conditions, whereas children might be influenced by developing sensory functions. That is not to say that SDN is a 'one modality, one clock' system, but rather a large network that also covers the interactions between multiple networks.

Taken together, from an intrinsic model's perspective, time is a consequence of cumulative states in a recurrent neural network that represent the amount of changes induced by external stimuli. In this sense, when listening to a piece of music repetitively, the perceived duration of both music and video (as a further stimulus) will be altered if presented again later on.

### Audiovisual Evidence for the Dynamic Attending Theory

DAT is endowed with a particular emphasis on attention, given that the count of temporal units depends on how well attentional pulses synchronize with the external event, also known as the temporal entrainment effect. The term specifies the coupling of the tempo of extrinsic temporal cues and that of pacemaker pulses (Jones, 2010). The emphasis on external entrainment like music began in the early days of the formulation of the clock model (Barry, 1990; Pöppel, 1989). Neurobiological evidence suggests that the just noticeable differences for auditory gaps can be modulated when neural activities were entrained with specific frequency bands and amplitudes (Henry et al., 2014). Regarding music, the synchronization between neural oscillations and musical beats was substantiated as the steady-state event potential (SS-EP) evoked by periodicity in musical beats (Nozaradan, 2014).

Behavioural evidence provided similar findings. Fast tempo was found to lead to overestimation, or "time dilation", and slow tempo to underestimation, or "time contraction" in both auditory (Wang & Shi, 2019) and visual perception (Ortega & López, 2008). In addition, behavioural entrainment to external beats were found across age ranges and stimulus types, including auditory sequences and music excerpts (for a review, see Repp & Su, 2013). The experimental paradigms usually provide participants with a rhythmic beat that ceases

(or not) after a short period of entrainment and require them to continue tapping or moving along with the beats. [Boasson and Granot \(2012\)](#) adopted a paradigm of tapping to pitch rises and drops in multiple melodic sequences, in order to examine the entrainment effect. In their study, however, musicians and non-musicians uniformly exhibited faster-paced tapping behavior with rising pitch. This is consistent with other findings which revealed no difference in predictive timing between musically trained and untrained groups (e.g., [Repp, 2010](#)), whereas other studies indicated that musicians (percussionists) exhibited better entrainment performance when exposed to intense beat production activities ([Cicchini et al., 2012](#)). These studies suggest that individuals actively entrain with external rhythms and perceive past durations accordingly, and may provide evidence of the wide applicability of DAT.

Building upon simple click paradigms as previously discussed ([Treisman et al., 1990](#)), research in recent years used naturalistic stimuli, since DAT is most applicable in music and speech. Periodic tone entrainment studies yielded new results: [Wearden et al. \(2017\)](#) found the residual effect of the classic click train paradigm, that is, the higher the preceding click frequency, the longer the following duration would be perceived. They have also observed similar effects with irregular tones as well as white noise. This study revealed multiple approaches to activate and to speed up the internal clock. Periodic and aperiodic clicks, as well as rhythmic visual flickers and even white noise influenced results. In addition, the entrainment effect was also verified to transcend as long as 8s after hearing high-frequency clicks, indicating that the emission of attentional pulses has a latency between activation and cessation.

More complex stimuli such as music are processed similarly. Fast music compared to slow one was perceived to be longer due to the accumulation of more temporal units. A study adopted Mozart's Sonata for two pianos (K.448), where participants tended to overestimate the duration when the excerpt was at the "fast" (120BPM) end of the spectrum ([Wang & Shi, 2019](#)). The effect, nevertheless, is subject to the allocation of attention. [Keller and Burnham \(2005\)](#) emphasized the flexibility of attention when listening to musical meter, which could be composed of multiple metrical layers. Therefore, tracking high and low metrical structures is expected to have its corresponding effects on psychological time (cf. [Hammerschmidt & Wöllner, 2020](#)), as the former should hypothetically lead to fewer mental counts and thus time compression. Neurological evidence also indicated that focusing on different temporal structures led to alignments in steady state event potential (SS-EP) frequencies, deciphered from EEG recordings ([Nozaradan et al., 2012](#)). In this case, neural entrainment reflects that attending to local features in complex auditory signals could form mental representations of time by modulating the original neural oscillations.

When more attention is allocated to the temporal features of music, [Cocenas-Silva et al. \(2011\)](#) observed a time dilation effect. When participants were asked to group excerpts of various arousal levels based on their estimated lengths, those which were highly arousing tended to be overestimated. The finding is consistent with [Droit-Volet et al.'s \(2013\)](#) observation that faster music, which was thought to be more arousing, was judged to be longer than the slow, less arousing ones. We might reason that, when individuals attend to temporal features of the auditory signals, the temporal entrainment effect is stronger compared to situations when they attend to other features such as key chords and pleasantness.

## Audiovisual Evidence for the Scalar Expectancy Theory

The following examines the evidence for multiple sensory modalities that either support or disagree with SET. To establish a solid ground for SET, researchers tried to find evidence for Weber's fraction, or a constant variance to subjective timing, across different sensory modalities, durations, populations, and other conditions. Wearden and Jones (2007) probed the scalar property of subjective timing using two variations of the duration comparison task with auditory tones ranging from 600ms to 10s. They found a linear increase in subjective timing that conforms to Weber's law. This effect is consistent also in the visual domain. In a duration discrimination study, Grondin (2001) found that participants exhibited similar sensitivity towards intervals marked by visual flickers between 600 to 900 ms, in accordance with Weber's law. However, the ratio changed when the inter-stimulus interval went beyond 900ms. The violation of Weber's law might be due to potentially explicit counting.

Similarly, mixed findings have been reported in multi-modalities studies. Hypothetically, if the scalar property holds across modalities, one should expect a consistent linear increase in different modalities. This was indeed the case when participants performed predictive saccades, or eye-movement timing, when intervals from 500 to 1000ms were presented either as visual flashes or auditory tone flutters (Joiner et al., 2007). However, comparing Weber's ratios between the two modalities revealed that auditory timing had greater variability than visual timing, as shown in participants' reactive eye movements when tracking the periodic cues. Hence, one might deduct that the scalar property holds but is also subject to stimulus modality. Block and Gruber (2014) argued that the obstacles of finding a cross-modal transfer effect was restricted to below the 3 to 5s window, beyond which the automatic processing should diminish due to the limited capacity of working memory.

On the other hand, evidence against the scalar property has been presented in auditory studies. Grondin (2012) adopted three approaches to measure Weber's ratio: duration discrimination, reproduction and categorization tasks on a spectrum from 1 to 1.9 seconds using pure tones. In all three tasks, Weber's ratio appeared to be higher when the intervals were longer regardless of the number of interval repetitions, in this case either 1, 3, or 5 times. These results indicate the inconsistency in Weber's ratio or temporal sensitivity despite different emphases of each paradigm on the timing process. Grondin (2010a) pointed out that the failure of conforming to the pacemaker-counter model, which SET is built upon, was because this model no longer applied to this duration range (see Figure 1). More specifically, a cut-off point at 1.2 to 1.3s was observed. This aligns with observations from other studies (for a review, see Matthews & Meck, 2014). The question is, how is time processed beyond that point? Some researchers proposed that a learning effect might have altered the variance, as the brain was influenced by multiple exposures to the same interval (Matthews & Grondin, 2012). Findings across timing tasks and sensory modalities, nevertheless, support the presence of a unitary clock system.

Despite the controversial evidence, reports investigating timing precision on multiple sensory modalities align with what the striatal beat frequency theory proposed: a familiarity effect that is reflected by enhanced synaptic communication between neurons. This might lead to higher processing efficiency and smaller variability compared to unfamiliar intervals. Grondin's (2012) experiments revealed that participants performed better in 3- and 5-interval discrimination than when only one interval was presented. Frequent exposure to timing tasks, as a part of music training, may also implicate the benefits of enhanced neural connection. In Rammsayer and Altenmüller's (2006) study, musicians outperformed non-musicians in a perceptual timing task in terms

of showing less variance and thus higher temporal sensitivity for instance in duration discrimination tasks. Musicians, however, did not exhibit significant superiority in a temporal generalization task, where participants compared the duration of an excerpt to the reference at the beginning, hypothetically stored in one's working memory. The authors believed that this was due to the fact that the intervals exceeded working memory capacities. This explanation is equally applicable to Grondin and Killeen's (2009) results, where participants in a reproduction-by-tapping task performed significantly better if they adopted counting or singing, compared to doing nothing. Thus it might be concluded that the SET indeed predicts the timing performance only within short intervals of no more than 2s (for a review, see Ivry & Schlerf, 2008). Nevertheless, it is equally important to understand timing within a few notes as well as in larger musical structures such as phrases.

## Conclusions

This review has discussed two internal clock models: the intrinsic and the central clock models. The intrinsic model emphasizes automatic processing of temporal information in the subsecond range, while the central clock model explains the suprasecond (seconds to minutes) range of timing, which demands higher levels of cognitive control. Controversially, the Scalar Expectancy Theory, which can be seen as a specific account of the central clock model, applies to timing in the seconds range only, while the Dynamic Attending Theory works for timing intervals from seconds to minutes. According to SET and DAT, short intervals are represented linearly through the accumulation of pacemaker pulses, while longer intervals are represented nonlinearly, as pulse emission is calibrated to align with external periodicity. As for intervals of hours and longer, the timing process is subject to contextual changes and memory segmentation, and relevant research is scarce.

Audition, among all modalities, shows superiority in temporal processing by entailing higher sensitivity to detect changes and to estimate interval lengths compared to vision and other sensory modalities. In this sense, the modality specificity supports a distributed timing mechanism. Yet more evidence is needed to explain the cross-modal transfer of training effects in, for example, duration discrimination. Despite years of debate on the superiority of one clock model, there is no conclusive evidence to the best of our knowledge. We come to the observation that each model has its best fit at a different time duration scale, and as to whether discrete events (SET) or complex streams (DAT) such as in music are at the core of the investigation.

Regarding the explanatory power of the internal clock models for the perception of musical time, it is therefore necessary to consider an interval-specific approach. Short interval timing within the milliseconds range plays a crucial role in music production such as expressive microtiming, whereas long interval timing is more strongly modified by attention, emotion, and working memory, consequently adding more variables to the equation. In this regard, the timing paradigm adopted in an ecologically plausible environment such as music concerts, movies, or sports should receive more attention. Ways of applying clock models to longer-interval timing and time estimation are yet to be investigated.

## Funding

This research was supported by a grant from the European Research Council to the second author (grant agreement: 725319) for the project "Slow motion: Transformations of musical time in perception and performance" (SloMo).

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

The authors have no support to report.

## References

- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66(4), 610-618. <https://doi.org/10.1016/j.neuron.2010.04.014>
- Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain*, 135(3), 656-677. <https://doi.org/10.1093/brain/awr210>
- Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: First-and second-order principles of subjective time. *Annual Review of Psychology*, 65, 743-771. <https://doi.org/10.1146/annurev-psych-010213-115117>
- Auhagen, W., & Busch, V. (1998). The influence of articulation on listeners' regulation of performed tempo. In R. Kopiez & W. Auhagen (Eds.), *Controlling creative processes in music* (pp. 69–92). Bern, Switzerland: Peter Lang.
- Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, 41(3), 254-311. <https://doi.org/10.1006/cogp.2000.0738>
- Barry, B. R. (1990). *Musical time: the sense of order*. Hillsdale, NY, USA: Pendragon Press.
- Behne, K. E. (1976). „Zeitmaße“: Zur Psychologie des musikalischen Tempoempfindens. *Die Musikforschung*, 29(2), 155-164.
- Bisson, N., Tobin, S., & Grondin, S. (2008). Remembering the duration of joyful and sad musical excerpts: Assessment with three estimation methods. *NeuroQuantology*, 7(1), 46-57. <https://doi.org/10.14704/nq.2009.7.1.206>
- Block, R. A., & Gruber, R. P. (2014). Time perception, attention, and memory: A selective review. *Acta Psychologica*, 149, 129-133. <https://doi.org/10.1016/j.actpsy.2013.11.003>
- Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4, 184-197. <https://doi.org/10.3758/BF03209393>
- Boasson, A. D., & Granot, R. (2012). Melodic direction's effect on tapping. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pasteriadis (Eds.), *Proceedings of 12th international conference on music perception and cognition, and the 8th triennial conference of the European society for the cognitive sciences of music* (pp.110-119). The joint conference ICMPC – ESCOM 2012, Thessaloniki, Greece. Retrieved from [http://icmpe-escom2012.web.auth.gr/files/papers/110\\_Proc.pdf](http://icmpe-escom2012.web.auth.gr/files/papers/110_Proc.pdf)
- Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, 142, 238-244. <https://doi.org/10.1016/j.actpsy.2012.11.012>

- Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, 60(8), 1357-1373. <https://doi.org/10.3758/BF03207998>
- Boltz, M. G. (2017). Auditory driving in cinematic art. *Music Perception*, 35(1), 77-93. <https://doi.org/10.1525/mp.2017.35.1.77>
- Broadway, J. M., & Engle, R. W. (2011). Individual differences in working memory capacity and temporal discrimination. *PLOS ONE*, 6(10), Article e25422. <https://doi.org/10.1371/journal.pone.0025422>
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14(4), 362-366. <https://doi.org/10.1111/1467-9280.24441>
- Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, 28(3), 600-615. <https://doi.org/10.1037/0096-1523.28.3.600>
- Bueti, D., van Dongen, E. V., & Walsh, V. (2008). The role of superior temporal cortex in auditory timing. *PLOS ONE*, 3(6), Article e2481. <https://doi.org/10.1371/journal.pone.0002481>
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews: Neuroscience*, 6(10), 755-765. <https://doi.org/10.1038/nrn1764>
- Buonomano, D. V., Bramen, J., & Khodadadifar, M. (2009). Influence of the interstimulus interval on temporal processing and learning: Testing the state-dependent network model. *Philosophical Transactions of the Royal Society of London. Series B*, 364(1525), 1865-1873. <https://doi.org/10.1098/rstb.2009.0019>
- Buonomano, D. V., & Laje, R. (2011). Population clocks: Motor timing with neural dynamics. In S. Dehaene, & E. Brannon (Eds.), *Space, time and number in the brain* (pp. 71-85). Cambridge, MA, USA: Academic Press. <https://doi.org/10.1016/B978-0-12-385948-8.00006-2>
- Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research*, 82(6), 1195-1211. <https://doi.org/10.1007/s00426-017-0894-2>
- Burr, D., Della Rocca, E., & Morrone, M. C. (2013). Contextual effects in interval-duration judgements in vision, audition and touch. *Experimental Brain Research*, 230(1), 87-98. <https://doi.org/10.1007/s00221-013-3632-z>
- Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience*, 10(4), 423-425. <https://doi.org/10.1038/nn1874>
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, 13(10), 1-19. <https://doi.org/10.1167/13.10.1>
- Chen, L., Zhou, X., Müller, H. J., & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration. *Journal of Experimental Psychology: General*, 147(12), 1851-1864. <https://doi.org/10.1037/xge0000487>

- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience*, 32(3), 1056-1060. <https://doi.org/10.1523/JNEUROSCI.3411-11.2012>
- Cocenas-Silva, R., Bueno, J. L. O., Molin, P., & Bigand, E. (2011). Multidimensional scaling of musical time estimations. *Perceptual and Motor Skills*, 112(3), 737-748. <https://doi.org/10.2466/11.24.PMS.112.3.737-748>
- Colflesh, G. J., & Conway, A. R. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, 14(4), 699-703. <https://doi.org/10.3758/BF03196824>
- Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience*, 33(9), 4002-4010. <https://doi.org/10.1523/JNEUROSCI.4675-12.2013>
- Droit-Volet, S., & Berthon, M. (2017). Emotion and implicit timing: The arousal effect. *Frontiers in Psychology*, 8, Article 176. <https://doi.org/10.3389/fpsyg.2017.00176>
- Droit-Volet, S., Bigand, E., Ramos, D., & Bueno, J. L. O. (2010). Time flies with music whatever its emotional valence. *Acta Psychologica*, 135, 226-232. <https://doi.org/10.1016/j.actpsy.2010.07.003>
- Droit-Volet, S., Fayolle, S. L., & Gil, S. (2011). Emotion and time perception: Effects of film-induced mood. *Frontiers in Integrative Neuroscience*, 5, Article 33. <https://doi.org/10.3389/fnint.2011.00033>
- Droit-Volet, S., Fayolle, S., Lamotte, M., & Gil, S. (2013). Time, emotion and the embodiment of timing. *Timing & Time Perception*, 1, 99-126. <https://doi.org/10.1163/22134468-00002004>
- Eagleman, D. M., Peter, U. T., Buonomano, D., Janssen, P., Nobre, A. C., & Holcombe, A. O. (2005). Time and the brain: How subjective time relates to neural time. *The Journal of Neuroscience*, 25(45), 10369-10371. <https://doi.org/10.1523/JNEUROSCI.3487-05.2005>
- Eerola, T., Luck, G., & Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception and cognition* (pp. 472-476). The Society for Music Perception and Cognition and European Society for the Cognitive Sciences of Music Bologna, Bologna, Italy.
- Ellis, M. C. (1991). Research note: Thresholds for detecting tempo change. *Psychology of Music*, 19(2), 164-169. <https://doi.org/10.1177/0305735691192007>
- Escoffier, N., Sheng, D. Y. J., & Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica*, 135, 12-16. <https://doi.org/10.1016/j.actpsy.2010.04.005>
- Gibbon, J. (1977). Scalar Expectancy Theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279-325. <https://doi.org/10.1037/0033-295X.84.3.279>
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52-77. <https://doi.org/10.1111/j.1749-6632.1984.tb23417.x>
- Goudriaan, J. C. (1921). Le rythme psychique dans ses rapports avec les fréquences cardiaque et respiratoire. *Archives Néerlandaises de Physiologie*, 6, 77-110.

- Grondin, S. (2001). Discriminating time intervals presented in sequences marked by visual signals. *Perception & Psychophysics*, 63(7), 1214-1228. <https://doi.org/10.3758/BF03194535>
- Grondin, S. (2010a). Unequal Weber fractions for the categorization of brief temporal intervals. *Attention, Perception & Psychophysics*, 72(5), 1422-1430. <https://doi.org/10.3758/APP.72.5.1422>
- Grondin, S. (2010b). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, 72(3), 561-582. <https://doi.org/10.3758/APP.72.3.561>
- Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 880-890. <https://doi.org/10.1037/a0027188>
- Grondin, S., & Killeen, P. R. (2009). Tracking time with song and count: Different Weber functions for musicians and nonmusicians. *Attention, Perception & Psychophysics*, 71(7), 1649-1654. <https://doi.org/10.3758/APP.71.7.1649>
- Gu, B. M., van Rijn, H., & Meck, W. H. (2015). Oscillatory multiplexing of neural population codes for interval timing and working memory. *Neuroscience and Biobehavioral Reviews*, 48, 160-185. <https://doi.org/10.1016/j.neubiorev.2014.10.008>
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science*, 16(3), 228-235. <https://doi.org/10.1111/j.0956-7976.2005.00808.x>
- Hammerschmidt, D., & Wöllner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time. *Music Perception*, 37(4), 263-277. <https://doi.org/10.1525/mp.2020.37.4.263>
- Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 111(41), 14935-14940. <https://doi.org/10.1073/pnas.1408741111>
- Herbert, R. (2012). *Everyday music listening: Absorption, dissociation and trancing*. Abingdon, United Kingdom: Routledge.
- Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12(7), 273-280. <https://doi.org/10.1016/j.tics.2008.04.002>
- Joiner, W. M., Lee, J. E., Lasker, A., & Shelhamer, M. (2007). An internal clock for predictive saccades is established identically by auditory or visual information. *Vision Research*, 47(12), 1645-1654. <https://doi.org/10.1016/j.visres.2007.02.013>
- Jones, M. R. (1981). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology: Music, Mind, and Brain*, 1(2), 34-51. <https://doi.org/10.1037/h0094282>
- Jones, M. R. (1990). Musical events and models of musical time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 207-240). NJ, USA: Lawrence Erlbaum Associates.
- Jones, M. R. (2010). Attending to sound patterns and the role of entrainment. In A. C., Nobre, & J. T., Coull (Eds.), *Attention and time* (pp. 317-330). Oxford, United Kingdom: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199563456.003.0023>



- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459-491. <https://doi.org/10.1037/0033-295X.96.3.459>
- Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron*, 53(3), 427-438. <https://doi.org/10.1016/j.neuron.2007.01.006>
- Keller, P. E., & Burnham, D. K. (2005). Musical meter in attention to multipart rhythm. *Music Perception*, 22(4), 629-661. <https://doi.org/10.1525/mp.2005.22.4.629>
- Langner, J. (2002). *Musikalischer Rhythmus und Oszillation* (Vol. 13). Bern, Switzerland: Peter Lang Publishing.
- Large, E. W. (2008). *Resonating to musical rhythm: Theory and experiment*. In S. Grondin (Ed.), *Psychology of time* (pp. 189-232). Bingley, United Kingdom: Emerald Group Publishing. <https://doi.org/10.1016/B978-0-08046-977-5.00006-5>
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119-159. <https://doi.org/10.1037/0033-295X.106.1.119>
- Lieb, E. H., & Yngvason, J. (1999). The physics and mathematics of the second law of thermodynamics. *Physics Reports*, 310, 1-96. [https://doi.org/10.1016/S0370-1573\(98\)00082-9](https://doi.org/10.1016/S0370-1573(98)00082-9)
- London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford, United Kingdom: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195160819.001.0001>
- Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition*, 77(2), 298-303. <https://doi.org/10.1016/j.bandc.2011.07.007>
- Matell, M. S., Meck, W. H., & Nicolelis, M. A. (2003). Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behavioral Neuroscience*, 117(4), 760-773. <https://doi.org/10.1037/0735-7044.117.4.760>
- Matthews, W. J., & Grondin, S. (2012). On the replication of Kristofferson's (1980) quantal timing for duration discrimination: Some learning but no quanta and not much of a Weber constant. *Attention, Perception & Psychophysics*, 74(5), 1056-1072. <https://doi.org/10.3758/s13414-012-0282-3>
- Matthews, W. J., & Meck, W. H. (2014). Time perception: The bad news and the good. *Cognitive Science*, 5(4), 429-446. <https://doi.org/10.1002/wcs.1298>
- McDermott, J. H., Wroblewski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences of the United States of America*, 108(3), 1188-1193. <https://doi.org/10.1073/pnas.1004765108>
- McPherson, T., Berger, D., Alagapan, S., & Fröhlich, F. (2018). Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. *Scientific Reports*, 8(1), Article 11782. <https://doi.org/10.1038/s41598-018-29267-z>
- Meck, W. H. (1984). Attentional bias between modalities: Effect on the internal clock, memory, and decision stages used in animal time discrimination. *Annals of the New York Academy of Sciences*, 423(1), 528-541. <https://doi.org/10.1111/j.1749-6632.1984.tb23457.x>
- Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience*, 36, 313-336. <https://doi.org/10.1146/annurev-neuro-062012-170349>

- Miall, C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation*, 1, 359-371. <https://doi.org/10.1162/neco.1989.1.3.359>
- Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 18(4), 1559-1570. <https://doi.org/10.1523/JNEUROSCI.18-04-01559.1998>
- Nichols, R. (2011). *Ravel*. London, United Kingdom: Yale University Press. <https://doi.org/10.1093/ml/gcs082>
- Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society of London. Series B*, 369, Article 20130393. <https://doi.org/10.1098/rstb.2013.0393>
- Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, 32(49), 17572-17581. <https://doi.org/10.1523/JNEUROSCI.3203-12.2012>
- Ornstein, R. E. (1969). *On the experience of time*. London, United Kingdom: Penguin Publisher.
- Ortega, L., & López, F. (2008). Effects of visual flicker on subjective time in a temporal bisection task. *Behavioural Processes*, 78(3), 380-386. <https://doi.org/10.1016/j.beproc.2008.02.004>
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1770-1787. <https://doi.org/10.1037/0096-1523.26.6.1770>
- Phillips, D. P., & Hall, S. E. (2002). Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra. *Hearing Research*, 174(1-2), 133-141. [https://doi.org/10.1016/S0378-5955\(02\)00647-0](https://doi.org/10.1016/S0378-5955(02)00647-0)
- Polti, I., Martin, B., & van Wassenhove, V. (2018). The effect of attention and working memory on the estimation of elapsed time. *Scientific Reports*, 8(1), Article 6690. <https://doi.org/10.1038/s41598-018-25119-y>
- Pöppel, E. (1989). The measurement of music and the cerebral clock: A new theory. *Leonardo*, 22, 83-89. <https://doi.org/10.2307/1575145>
- Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4), 411-440. <https://doi.org/10.2307/40285311>
- Pressnitzer, D., Suied, C., & Shamma, S. (2011). Auditory scene analysis: The sweet music of ambiguity. *Frontiers in Human Neuroscience*, 5, Article 158. <https://doi.org/10.3389/fnhum.2011.00158>
- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24(1), 37-48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science*, 29(2), 200-213. <https://doi.org/10.1016/j.humov.2009.08.002>
- Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085-1099. <https://doi.org/10.1037/0096-1523.28.5.1085>

- Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3), 403-452. <https://doi.org/10.3758/s13423-012-0371-2>
- Schäfer, T., Fachner, J., & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology*, 4, Article 508. <https://doi.org/10.3389/fpsyg.2013.00508>
- Schnupp, J. W., Hall, T. M., Kokelaar, R. F., & Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *The Journal of Neuroscience*, 26(18), 4785-4795. <https://doi.org/10.1523/JNEUROSCI.4330-05.2006>
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9-18. <https://doi.org/10.1016/j.tins.2008.09.012>
- Schulze, H. H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 40(2), 173-181. <https://doi.org/10.1007/BF00308412>
- Schwartz, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research*, 216(2), 685-691. <https://doi.org/10.1016/j.bbr.2010.09.015>
- Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20(3), 361-366. <https://doi.org/10.1016/j.conb.2010.03.009>
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research: Cognitive Brain Research*, 14(1), 147-152. [https://doi.org/10.1016/S0926-6410\(02\)00069-1](https://doi.org/10.1016/S0926-6410(02)00069-1)
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145(3638), 1328-1330. <https://doi.org/10.1126/science.145.3638.1328>
- Stern, L. W. (1897). Psychische Präsenzzeit. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 13, 325-349.
- Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology*, 5, Article 1329. <https://doi.org/10.3389/fpsyg.2014.01329>
- Tobin, S., Bisson, N., & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLOS ONE*, 5(2), Article e9271. <https://doi.org/10.1371/journal.pone.0009271>
- Toiviainen, P., Burunat, I., Brattico, E., Vuust, P., & Alluri, V. (2019). The chronnectome of musical beat. *NeuroImage*, Article 116191. Advance online publication. <https://doi.org/10.1016/j.neuroimage.2019.116191>
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs*, 77(13), 1-31. <https://doi.org/10.1037/h0093864>
- Treisman, M., & Brogan, D. (1992). Time perception and the internal clock: Effects of visual flicker on the temporal oscillator. *The European Journal of Cognitive Psychology*, 4(1), 41-70. <https://doi.org/10.1080/09541449208406242>
- Treisman, M., Faulkner, A., Naish, P. L., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, 19(6), 705-742. <https://doi.org/10.1068/p190705>

- van Rijn, H., Gu, B. M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. In H. Merchant & V. D. Lafuente (Eds.), *Neurobiology of interval timing* (pp. 75-99). Berlin, Germany: Springer. [https://doi.org/10.1007/978-1-4939-1782-2\\_5](https://doi.org/10.1007/978-1-4939-1782-2_5)
- van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLOS ONE*, 3(1), Article e1437. <https://doi.org/10.1371/journal.pone.0001437>
- Wang, X., & Shi, Z. (2019, September 10-12). *Temporal entrainment effect: Can music enhance our attention resolution in time?* [Poster presentation]. The 12th International Conference of Students of Systematic Musicology. SysMus, Berlin, Germany.
- Wang, X., Wöllner, C., & Shi, Z. (2019, September 6-8). *Perceiving tempo in incongruent audiovisual contexts: An exploratory study with a temporal bisection paradigm* [Poster presentation]. Jahrestagung der Deutsche Gesellschaft für Musikpsychologie, Eichstätt, Germany.
- Wearden, J. H., & Jones, L. A. (2007). Is the growth of subjective time in humans a linear or nonlinear function of real time? *Quarterly Journal of Experimental Psychology*, 60(9), 1289-1302. <https://doi.org/10.1080/17470210600971576>
- Wearden, J. H., Williams, E. A., & Jones, L. A. (2017). What speeds up the internal clock? Effects of clicks and flicker on duration judgements and reaction time. *Quarterly Journal of Experimental Psychology*, 70(3), 488-503. <https://doi.org/10.1080/17470218.2015.1135971>
- Wittmann, M. (2013). The inner sense of time: How the brain creates a representation of duration. *Nature Reviews. Neuroscience*, 14(3), 217-223. <https://doi.org/10.1038/nrn3452>
- Wöllner, C., & Halpern, A. R. (2016). Attentional flexibility and memory capacity in conductors and pianists. *Attention, Perception & Psychophysics*, 78(1), 198-208. <https://doi.org/10.3758/s13414-015-0989-z>
- Wöllner, C., Hammerschmidt, D., & Albrecht, H. (2018). Slow motion in films and video clips: Music influences perceived duration and emotion, autonomic physiological activation and pupillary responses. *PLOS ONE*, 13(6), Article e0199161. <https://doi.org/10.1371/journal.pone.0199161>
- Zakay, D., & Block, R. A. (1995). An attentional-gate model of prospective time estimation. *Time and the Dynamic Control of Behavior*, 167-178.
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), 5768-5773. <https://doi.org/10.1073/pnas.1000121107>

## Study 2

### **Perceiving tempo in incongruent audiovisual presentations of human motion: Evidence for a visual driving effect**

Xinyue Wang, Clemens Wöllner, and Zhuanghua Shi

*Timing & Time Perception. 10(1), 75-95*

<https://doi.org/10.1163/22134468-bja10036>

Reproduced with kind permission from Brill



## Perceiving Tempo in Incongruent Audiovisual Presentations of Human Motion: Evidence for a Visual Driving Effect

Xinyue Wang<sup>\*,\*\*</sup>, Clemens Wöllner<sup>\*\*\*</sup> and Zhuanghua Shi<sup>\*\*\*\*</sup>

<sup>1</sup>Institute for Systematic Musicology, Universität Hamburg, Hamburg, Germany

<sup>2</sup>Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

Received 31 August 2020; accepted 30 March 2021

### Abstract

Compared to vision, audition has been considered to be the dominant sensory modality for temporal processing. Nevertheless, recent research suggests the opposite, such that the apparent inferiority of visual information in tempo judgements might be due to the lack of ecological validity of experimental stimuli, and reliable visual movements may have the potential to alter the temporal location of perceived auditory inputs. To explore the role of audition and vision in overall time perception, audiovisual stimuli with various degrees of temporal congruence were developed in the current study. We investigated which sensory modality weighs more in holistic tempo judgements with conflicting audiovisual information, and whether biological motion (point-light displays of dancers) rather than auditory cues (rhythmic beats) dominate judgements of tempo. A bisection experiment found that participants relied more on visual tempo compared to auditory tempo in overall tempo judgements. For fast tempi (150 to 180 BPM), participants judged 'fast' significantly more often with visual cues regardless of the auditory tempo, whereas for slow tempi (60 to 90 BPM), they did so significantly less often. Our results support the notion that visual stimuli with higher ecological validity have the potential to drive up or down the holistic perception of tempo.

### Keywords

Tempo judgement, audiovisual timing, visual driving, bisection paradigm

\* To whom correspondence should be addressed. E-mail: xinyue.wang@uni-hamburg.de

\*\* ORCID: 0000-0003-2986-7735

\*\*\* ORCID: 0000-0002-8508-3508

\*\*\*\* ORCID: 0000-0003-2388-6695

## 1. Introduction

Perceiving inconsistent audiovisual information is common in daily life. In many cases, conflicting inputs of one modality are able to alter the percept of another. The McGurk effect (McGurk & MacDonald, 1976), for example, is a famous example that lip movements not corresponding to the speech alternate the perceived sounds. Different pianists' performances coupled with the same soundtrack have been perceived to be different in a number of dimensions (Behne & Wöllner, 2011). It is of interests whether similar observations can be extended to the perception of timing and tempo. The question of how audiovisual asynchrony affects temporal processing has also attracted much attention. The dominant role of audition has been long recognised in temporal processing in the sense that it provides higher accuracy and precision than vision (Grondin, 2010). However, this view is challenged by emerging evidence of the superiority of meaningful visual movements in time perception with both abstract (Grahn, 2012) and real-life stimuli (Hove & Keller, 2010; London et al., 2016). Thus, it remains controversial whether audition or vision dominates our perception of tempo.

There has been a long debate in research about whether timing is based on a central or on a distributed system (Occelli et al., 2011; Penney, 2003; Van Wassenhove et al., 2008; for an overview, see Wang & Wöllner, 2019). Some argue that the timing mechanism is distributed, which can explain the discrepancies in timing performance between, for instance, vision and audition (Grondin et al., 2008); while others favour the notion that the modality difference in time perception comes from the interaction between different sensory modalities in the central timing (e.g., Levitan et al., 2015). The distributed account is supported by findings that audition has an advantage over vision and other sensory modalities in terms of duration discrimination (Grondin et al., 2008), reproduction (Gamache & Grondin, 2010), and estimation (Kanai et al., 2011). However, it should be noted that the domination of audition in temporal processing does not apply to all cases, particularly with biological trajectories (Hove et al., 2010) or movements (Allingham et al., 2020).

Evidence of temporal entrainment, which specifies the synchronisation between two rhythms, has been observed with musical (Hammerschmidt & Wöllner, 2020), visual (Iversen et al., 2015) as well as tactile stimuli (Occelli et al., 2011). For tempo judgements, the temporal ventriloquism effect (Burr et al., 2009) and the auditory driving effect (Shipley, 1964) have both suggested the dominance of audition over visual displays by 'dragging' the temporal location of the latter to that of the former. Even when the auditory stimuli were not attended to, or reduced in salience, duration judgements clearly leaned towards that of the perceived tone rather than of the visual circle in this case, suggesting that the processing of auditory temporal cues was possibly autonomous and occupied minimal cognitive resources (Ortega et al., 2014). Alternatively, the reliability

of the sensory inputs was crucial when perceiving durations: whichever channel provided the least noise was assigned the most weight in duration estimation (Hartcher-O'Brien et al., 2014; Shi et al., 2010). It is therefore likely that auditory temporal inputs were more reliable in studies where audition outweighed vision, given that multisensory temporal information is integrated in an optimal fashion (Shi et al., 2013).

Relatively few studies have explored tempo judgement in the context of audiovisual stimuli with high ecological validity. Among those who adopted naturalistic visual stimuli, a strong influence of the visual over the auditory inputs has been found. For example, videos of musicians playing long notes on the marimba, coupled with long and short corresponding sounds, shifted the perceived note length towards 'long' (Schutz & Lipscomb, 2007). The effect of human point-light displays on unimodal (auditory or visual) and bimodal (audiovisual) stimuli of varying musical tempi indicated that movements of high energy led to faster perceived tempo of auditory stimuli (London et al., 2016). However, it remains unclear how tempo is holistically perceived when participants are asked to judge it based on both sensory information channels. Furthermore, to our knowledge, no study investigated a combination of audition and vision in a controlled variation of audiovisually inconsistent stimuli.

Evidence has suggested that vision might not always be less accurate than audition in duration and tempo judgements. Past research believes that vision dominates spatial rather than temporal localisation (e.g., Burr et al., 2009; Repp, 2003), and has a lower temporal sensitivity than audition in low-level information processing (Ortega et al., 2014). However, it should be noted that the evidence for high precision of audition comes from simple and controlled laboratory stimuli, and the arrangements of their presence in the task, such as simple visual flickers (Shipley, 1964; Treisman et al., 1990), coloured squares (Grahn et al., 2011), looming or receding discs or dots (Van Wassenhove et al., 2008) rather than naturalistic stimuli. Naturalistic stimuli, such as biological motion derived from human behaviours and social activities (Boltz, 2005), often yield movements with high compatibility (Hove et al., 2010). The point-light technique for human motion has first been used in experimental research by Johansson (1973). Point-light displays (PLD) of biological motion are derived from natural human (or animal) movements such that visual details including facial expressions or clothes are not shown, yet the naturalness in terms of movement kinematics is preserved. Abstract visual stimuli such as flashes, on the other hand, provide fewer cues, no biological movement trajectories, and thus a less rich temporal structure. An earlier attempt by Boltz (2005) compared the effects of naturalistic scenes when they were presented in either auditory, visual, or audiovisual channels. Results suggested that modality differences were not observed in terms of duration reproduction or estimation accuracy. This led to a string of studies that adopted the naturalistic approach, in other words using visual movements that were common



in everyday life to achieve the same salience and temporal discriminability as of the auditory stimuli presented in past research (e.g., Grahn, 2012; Hove et al., 2010; Hove et al., 2013; London et al., 2016). Extraction of rhythmic patterns from visual movements was not only viable, but also independent from auditory interferences, indicating a robust temporal representation in the visual module (Su & Salazar-López, 2016).

The above evidence calls for further exploration in multisensory timing. The current study intends to advance research as follows: (1) Provide a scenario where auditory and visual information is equally important to tempo judgements. (2) Explore the potential interaction of audition and vision in tempo judgement. In light of this, the current study examined the effects of competing auditory and visual information on tempo judgement with biological motion and drumbeats, taking into account the ecological validity of both. We hypothesised that, with meaningful visual information such as point-light displays (PLDs), the visual tempo relative to auditory tempo would contribute more to the overall tempo judgement. A given unit change in visual tempo should thus lead to larger changes in the tempo judgement ratio (fast/slow). Accordingly, more participants should rely on visual rather than auditory information when judging tempo. Finally, we hypothesised that perceived naturalness would decrease as the audiovisual tempo discrepancy increases.

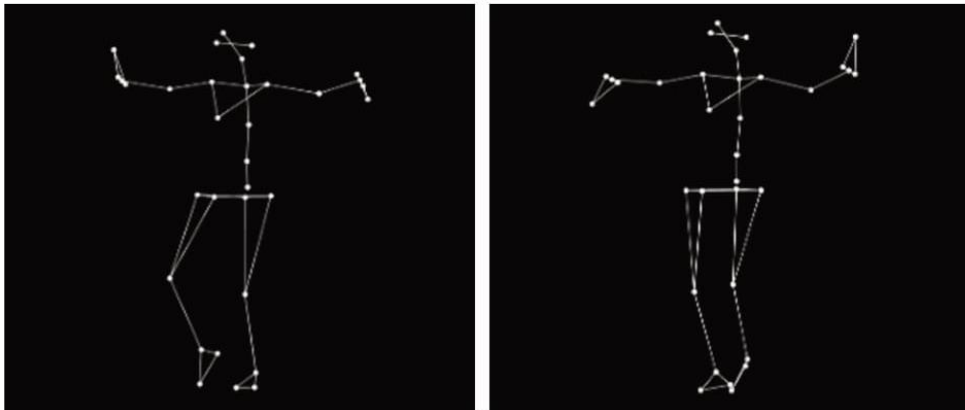
## 2. Methods

### 2.1. Participants

Twenty-four participants were recruited for the study (12 female; aged  $M = 24.21$  years,  $SD = 4.68$ ). Participants had a mean of 10.04 years ( $SD = 7.09$ ) of regular practice with musical instruments (including voice), and a mean of 7.65 years ( $SD = 7.18$ ) of lessons on their instrument. Thus, the current sample represents a population that has moderate to advanced musical training. The sample size had been calculated a priori for a  $3 \times 3$  design ( $\alpha = 0.05$ , Cohen's  $f = 0.25$ , power = 0.8), requiring at least 15 participants (using G\*Power; Faul et al., 2009). For a conservative approach, we recruited 24 participants. We also followed the guidelines of the Ethics Committee of the Faculty of Humanities, Universität Hamburg, and each participant was compensated 10 Euro for taking part.

### 2.2. Material

Participants were presented with audiovisual stimuli synthesised from isochronous drumbeats of nine tempi (60, 75, 90, 105, 120, 135, 150, 165, 180 beats per minute [BPM]), and visuals of the same tempo spectrum. The visual stimulus showed a PLD of a person jumping from left to right with the hands moving up and down (Fig. 1). This movement pattern was recorded with an eleven-camera motion-capture system (Qualisys Oqus, Qualisys AB, Göteborg, Sweden) at a



**Figure 1.** Stills of the PLD stimulus material. The movements entail flexion and extension of the hands, as well as lifting the left and right foot in turn. The stimuli can be found on Zenodo: [https://zenodo.org/record/4449683#.YAa\\_ROhKhaQ](https://zenodo.org/record/4449683#.YAa_ROhKhaQ).

framerate of 200 frames per second. Thirty-one markers were attached to the performer. The movement pattern was intended to be neither towards an action-based nor to a habitual (highly automatised) outcome, in order to avoid familiarity with the movement (Calvo-Merino et al., 2006). The movement was originally recorded at a speed of 120 BPM. The motion was presented from a 30-degree angle, in frontal view. The MATLAB Motion Capture (MoCap) Toolbox (Burger & Toiviainen, 2013) was adopted to speed up and slow down the PLD to the eight further tempi as specified above, while ensuring that the visual resolution and the number of data points per second were unchanged. The auditory stimuli of nine tempi (60 to 180 BPM, 15 BPM per step), on the other hand, were directly synthesized from a bass drum on the online drumbeat generator Drumbit (<https://drumbit.app>). Drum beats can be found in real life scenarios such as listening to techno music, thus providing higher naturalness than abstract auditory stimuli adopted in past studies such as sine waves. The PLDs and the drumbeat soundtracks of all tempi were then combined in Adobe Premiere Pro CC 2017 (Adobe Systems, San Jose, CA, USA) to create a total of 81 stimuli with all audiovisual tempo combinations. That is, all stimuli were bimodal videos varying in tempi.

The experiment was conducted in the SloMo laboratory at Universität Hamburg on a Dell U2414Hb monitor (Dell Technologies Inc., Round Rock, TX, USA), controlled by the software OpenSesame (Mathôt et al., 2012). A Sennheiser HD600 headphone set (Sennheiser GmbH, Hanover, Germany) was provided for the soundtrack. Participants responded to the experimental task by pressing the leftward or rightward button on the keyboard.

### 2.3. Design and Procedure

The current study introduced a  $3 \times 3$  design where three auditory tempi ranges (slow: 60, 75, 90 BPM; medium: 105, 120, 135 BPM; fast: 150, 165, 180 BPM) and

three visual tempi of the same spectrum acted as the independent variables, while taking the corresponding tempo judgement as the dependent variable. We chose the temporal bisection, a two-alternative forced-choice task (2AFC), to examine the ratio of 'fast' judgement at different tempo and modality conditions. The 2AFC method has been used in various studies of audiovisual integration (Chen et al., 2018; Gori et al., 2012; Shi et al., 2010). The bisection task has been widely used in cue combination research for both spatial stimuli and time in audiovisual integration processes (Gori et al., 2012). In an audiovisual Ternus apparent motion study, Shi et al. (2010) used the bisection task to measure the audiovisual duration integration. Roach et al. (2006) adopted the 2AFC temporal discrimination (higher or lower than 10 Hz) to estimate the threshold for the audiovisual temporal integration. Similar to other direct measures of duration or tempi, such as reproduction tasks, the bisection task is able to probe the audiovisual temporal integration as well as decisions in the tempo judgements. One benefit of using the bisection task, compared to the direct tempo reproduction or other motor-related tasks, is that the task is not influenced by motor noise. In a similar manner, here we applied the temporal bisection point to measure the holistic tempo judgements, that is whether observers shifted their judgements towards a fast or a slow tempo.

In the current study, participants were first presented with two audiovisual anchors (a fast tempo and a slow tempo) at the beginning of the experiment and were asked to judge the tempo of a given stimulus as close to the fast or the slow tempo holistically. In other words, they should focus on both auditory and visual information in the video stimuli. The slow anchor was a bimodal video with an audiovisually consistent tempo at 60 BPM, and a fast anchor at 180 BPM. They were then shown randomised trials of 81 bimodal videos generated from nine auditory stimuli (60, 75, 90, 105, 120, 135, 150, 165, 180 BPM) and nine visual stimuli of the same tempo spectrum, repeated three times. Each auditory tempo was combined with each visual tempo. The bimodal stimuli include both tempo-consistent and -inconsistent presentations. A total of 243 trials were presented to each participant.

Participants were seated in a quiet room approximately 80 cm in front of the monitor. Instructions were given by an experimenter who was trained to follow fixed protocols to ensure a standardised procedure. Each trial started with a fixation point for 100ms, followed by a PLD presentation of 5 s while drum sounds were simultaneously played through the headphones. After the presentation, a '?' was shown in the centre of the screen, prompting participants to judge if the tempo of the presented stimulus was closer to the slow or the fast anchor tempo. They were asked to press the leftward arrow key for the slow and the rightward arrow key for the fast anchor. To refresh participants' memory of the two anchors, a text reminder 'anchors' popped up after every nine trials, and the anchors were played once each time. Participants pressed any key to proceed and to watch the

fast and slow videos, with no time limit imposed. They were not required to make any response. In addition, an optional short break was offered every 40 or 41 trials.

After completing the bisection task, participants were asked to rate the naturalness ('how natural does this video feel?') of all 81 conditions in randomised order. A trial started with a fixation point for 100 ms, followed by a 5-s bimodal video stimulus. A visual instruction of the naturalness question 'Please rate how natural the video feels' was presented. On a horizontal gauge bar from 1 (marked as 'least natural') to 100 (marked as 'most natural'), participants placed the cursor in a relative position to give a response.

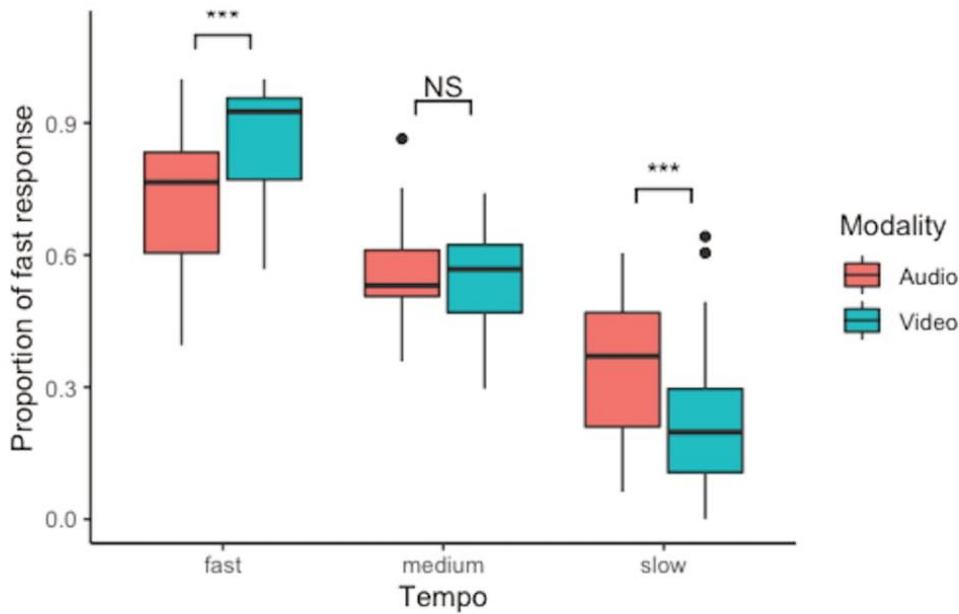
#### 2.4. Data Analyses

All statistical analyses were conducted with R (Version 3.5.3; R Core Team, 2019). Given that the distributions of individual participants' tempo judgements were heavily skewed, we used nonparametric analyses, more specifically a series of chi-square tests, to compare differences between the numbers of 'fast' versus 'slow' judgements for auditory or visual tempo conditions. In addition, we fitted the response ('fast' versus 'slow') as a logistic function of the auditory and visual tempi to obtain a 2D psychometric function, such that we can obtain the points of subjective equality (PSEs). A separation boundary (auditory and visual tempi) by the PSE, yielded by the logistic model when the likelihood of 'fast' judgement was 0.5, was then estimated on the individual and group levels. Pearson's correlations were conducted to explore the relationship between perceived naturalness and audiovisual discrepancy.

### 3. Results

#### 3.1. Visual Versus Auditory Tempo

We first examined the effect of modality on tempo judgement at each tempo condition. First, we grouped the tempo of the presentation either by (a) tempo of drumbeats or (b) tempo of visual PLD, by slow (60, 75, and 90 BPM), medium (105, 120, and 135 BPM), and fast (150, 165, and 180 BPM) tempo ranges. Figure 2 shows the mean proportion of 'fast' responses as a function of tempo, separated by modality. A chi-square test shows that participants were more likely to judge 'fast' with visual rather than auditory cues,  $\chi^2(1, N = 3878) = 81.96, p < 0.001$ , with a small to medium effect size ( $\phi = 0.15$ ). Correspondingly, for slow stimuli the proportion of 'fast' responses was significantly higher when participants relied on auditory cues than visual ones,  $\chi^2(1, N = 3907) = 59.16, p < 0.001$ . The effect size was also small to medium ( $\phi = 0.12$ ), according to Cohen (1988). There was no significant difference between modalities when the stimuli were presented at intermediate tempo,  $\chi^2(1, N = 3879) = 0.07, p = 0.80, \phi = 0.004$ . These results suggest that visual information plays a more important role than the auditory



**Figure 2.** Median proportions of ‘fast’ responses by modality and tempo ranges. Bottom and top of the boxes represent the first and third quartiles, with a line at the median. \*\*\*,  $p < 0.001$ ; NS stands for a non-significant  $p$  value.

information in tempo judgement at both ends of the tempo spectrum: When the PLDs were shown at a fast (150, 165, 180 BPM) or a slow (60, 75, 90 BPM) tempo, participants judged stimuli overall to be fast or slow, regardless of the auditory tempo.

To get a detailed picture of individual contributions of auditory and visual tempo in temporal judgements, we plotted the average response heatmap in Fig. 3. Both auditory and visual tempo contribute to judgements. In general, the faster the tempo, the more likely a participant would judge ‘fast’. Consistent with the analysis shown above, the change of response is more sensitive in the ‘vision’ direction than in the ‘audition’ direction, as evinced by the response contour changes along the visual rather than the auditory modality. To further quantify this, we applied a two-dimensional logistic regression, which is an extension of the one-dimensional psychometric function. We assume participants’ bisection response proportion  $p$  depends on both auditory and visual tempi ( $T_A, T_V$ ) with a logistic relation:

$$\log \frac{p}{1-p} = \alpha + \beta_A T_A + \beta_V T_V$$

where  $\alpha, \beta_A, \beta_V$  are coefficients of the logistic function. When  $p = 0.5$ , the above equation indicates the boundary separation between the fast and slow responses (see the red curve in Fig. 3).

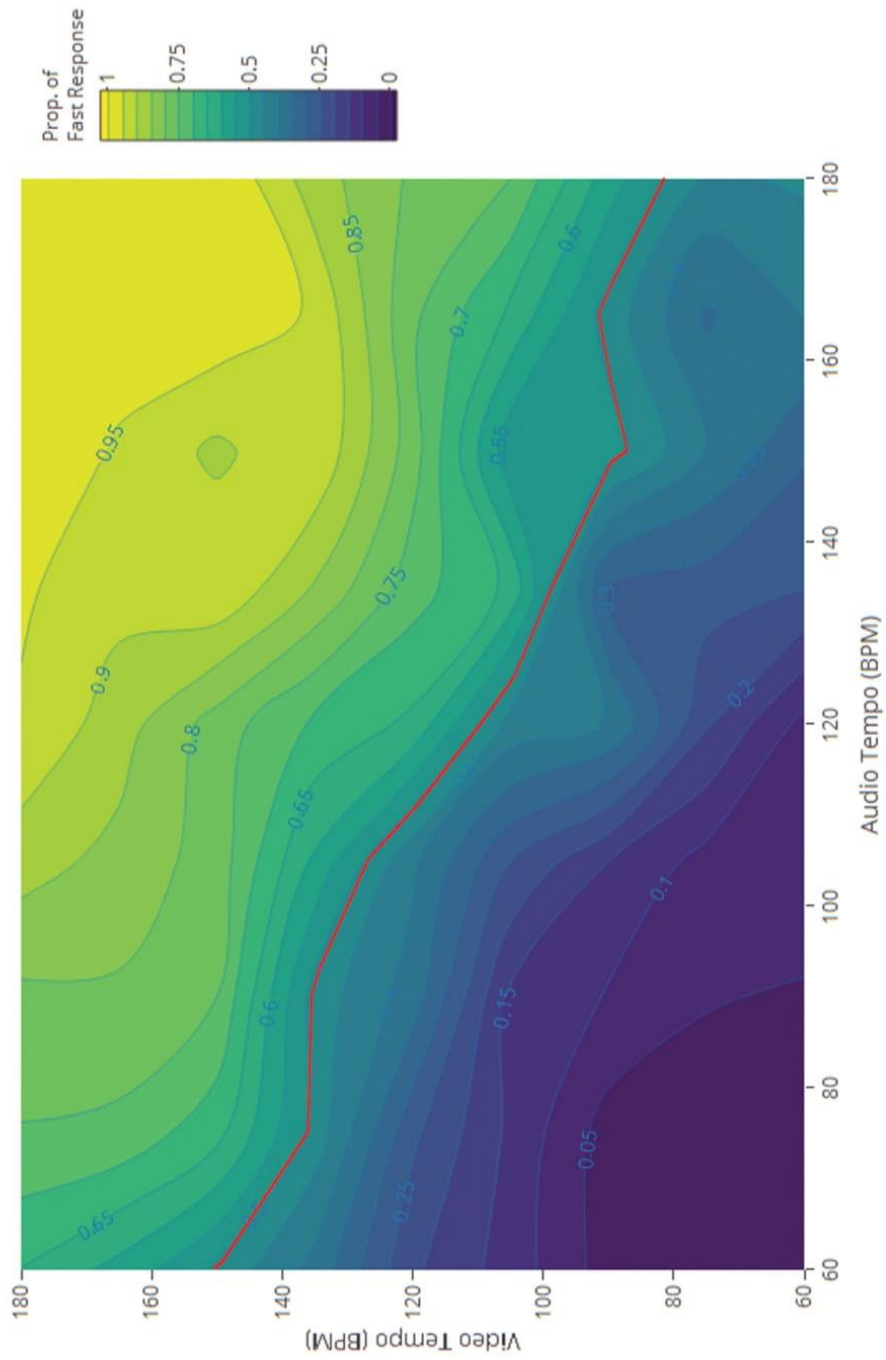


Figure 3. Heat map of the proportion of 'fast' responses distributed over the audio and video tempo spectrum. Yellow stands for high proportions, and blue for low proportions. Note that the red line marks the corresponding audio and video conditions at which the average ratio of fast response versus all responses equals 0.5, or the line of subjective equality.

The tempi in both conditions were first standardised by dividing each value by the median (120 BPM). The logistic model suggested a significant relationship among tempo judgement and the auditory as well as visual tempo,  $\chi^2(5827) = 2757.21$ ,  $p < 0.001$ . McFadden's  $R^2 = 0.34$ , which fell between 0.2 and 0.4, indicating a good fit. The estimated coefficients are shown in Table 1. The coefficients for the visual and auditory tempi were  $\beta_V = 5.04$  and  $\beta_A = 3.53$ , respectively. These reflect the degree of sensitivity in change of responses (according to the model, a unit change in the relative tempo contributes a change of log likelihood of two responses). Furthermore,  $\beta_V$  was significantly larger than  $\beta_A$  (based on the non-overlapping 95% confidence interval, see Table 1), which confirms that in general the visual tempo contributed more than the auditory tempo.

### 3.2. Individual Modality Reliance

Based on the separation boundary, we then categorized participants into one of the following types: vision-, audition-, or bimodal-reliant types. We used the log-ratio between the auditory and visual coefficients ( $|\beta_A / \beta_V|$ ) as an indicator of modality reliance. A  $\log \left| \frac{\beta_A}{\beta_V} \right|$  of 0 is an ideal case of equal reliance for the visual and auditory modality. Considering random fluctuation,  $\log \left| \frac{\beta_A}{\beta_V} \right|$  between  $-0.05$  and  $0.05$  was regarded as equal reliance. A ratio higher than  $0.05$  suggests auditory reliance, while a ratio lower than  $-0.05$  indicates visual reliance. Figure 4 shows examples of participants for the three types of modality reliance.

According to the categorisation, 16 participants were vision-reliant, seven audition-reliant, and one bimodal-reliant. A chi-square test of independence indicated a significant difference among the three groups,  $\chi^2(2, N = 24) = 14.25$ ,  $p < 0.001$ . That is to say, a larger proportion of the sample favoured visual information when it came to tempo judgement, regardless of the auditory tempo (Fig. 5). This finding again supports our hypothesis that the visual tempo, when presented as natural human movements, has higher priority than auditory tempo.

### 3.3. Naturalness and Audiovisual Discrepancy

A Pearson's correlation between the overall naturalness rating, ranging from 0 (least natural) to 100 (very natural), and the absolute values of the audiovisual tempo discrepancy suggested that the smaller the discrepancy between the audio and video tempo, the more natural a stimulus was perceived  $r(81) = -0.56$ ,  $p < 0.01$  (see Fig. 6).

A two-way ANOVA was conducted to examine the effect of auditory and visual tempo on perceived naturalness. Again, tempo ranges were categorised into slow (60 to 90 BPM), medium (105 to 135 BPM), and fast (150 to 180 BPM) for the analysis. Simple main effects suggested that fast visual tempo led to significantly

**Table 1.**  
Summary of the logistic regression analysis for the coefficients.

Variable	Model				
	<i>B</i>	SE <i>B</i>	Standardized	<i>z</i>	<i>p</i>
$\alpha$	-8.198	0.453	0.255	-18.10	< 0.001
$\beta_A$	3.533	0.407	1.105	8.683	< 0.001
$\beta_V$	5.039	0.416	1.590	12.120	< 0.001

higher naturalness ( $F_{2,1935} = 100.38, p < 0.001$ ). A statistically significant interaction between auditory and visual tempo on perceived naturalness was found,  $F_{4,1935} = 37.74, p < 0.001$ . Tukey's HSD post-hoc tests revealed that, for auditory tempo, no statistically significant differences were observed between different tempi. However, for visual tempo, fast stimuli were associated with higher naturalness ratings than the medium ( $p < 0.001, d = 0.24$ ) and the slow ones ( $p < 0.001, d = 0.74$ ). The medium-speed visuals were rated more natural than the slow ones ( $p < 0.001, d = 0.50$ ), regardless of the auditory tempo.

#### 4. Discussion

The present study examined the role of audition and vision in tempo judgements of naturalistic stimuli of biological motion, when the tempi of the two modalities are not consistent. First, the tempo of visual information (here the PLD stimuli) affected overall tempo judgements to a greater extent than that of the auditory information (drumbeats). Secondly, a higher proportion of the participants relied on visual rather than on auditory information for tempo judgement. Different modality weightings exhibited by individual participants again support our hypothesis that visual information, when presented as biological motion PLDs, should possess high ecological validity and consequently serves as the dominant tempo reference. Finally, a larger audiovisual tempo discrepancy led to lower perceived naturalness.

The results are consistent with our main hypothesis in the sense that naturalistic visual input dominated overall tempo judgement. Past studies with visual movements of varying complexities have observed similar effects where ecological validity could be derived from the stimuli. For abstract movements, the 'visual driving effect' has been found for both rhythm perception (Su & Jonikaitis, 2011) and duration estimation (Van Wassenhove et al., 2008). Su and Jonikaitis (2011) revealed that changes in the moving speed of dots or in luminance provided



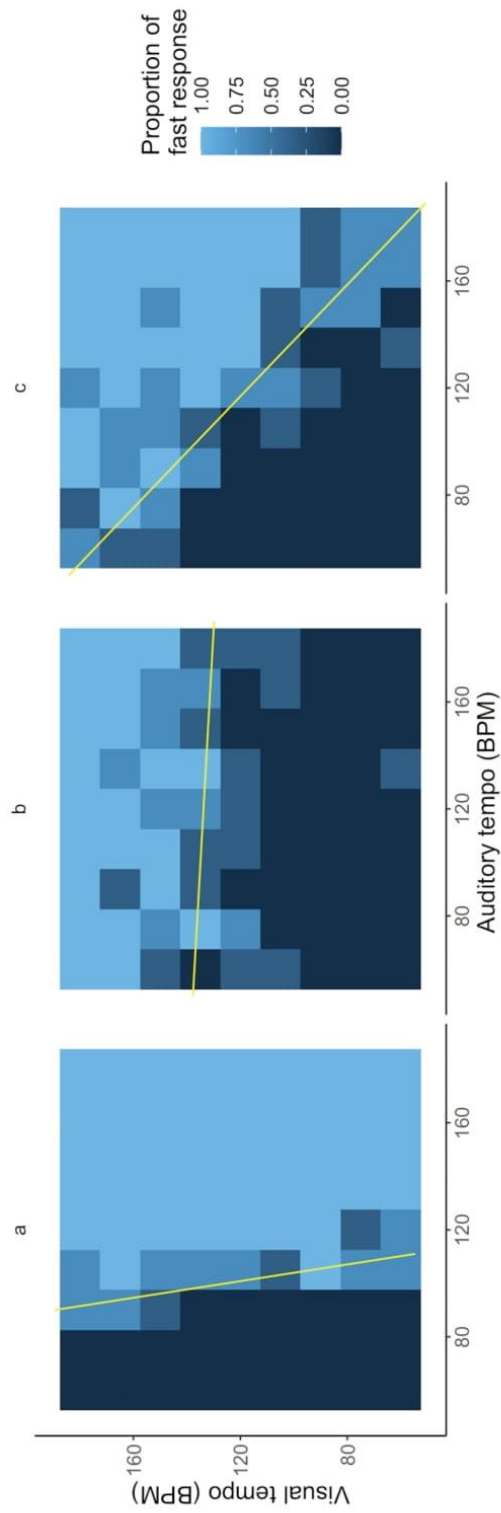


Figure 4. Examples of participants of typical audition-reliant judgement type (a), vision-reliant judgement type (b), and bimodal-reliant judgement type (c). The x-axis stands for auditory tempo, and the y-axis for visual tempo. The heatmap represents the proportion of 'fast' judgements, with lighter colour for higher proportion of 'fast' judgements. The yellow lines stand for the audiovisual tempi at which participants were equally likely to respond 'fast' or 'slow'.

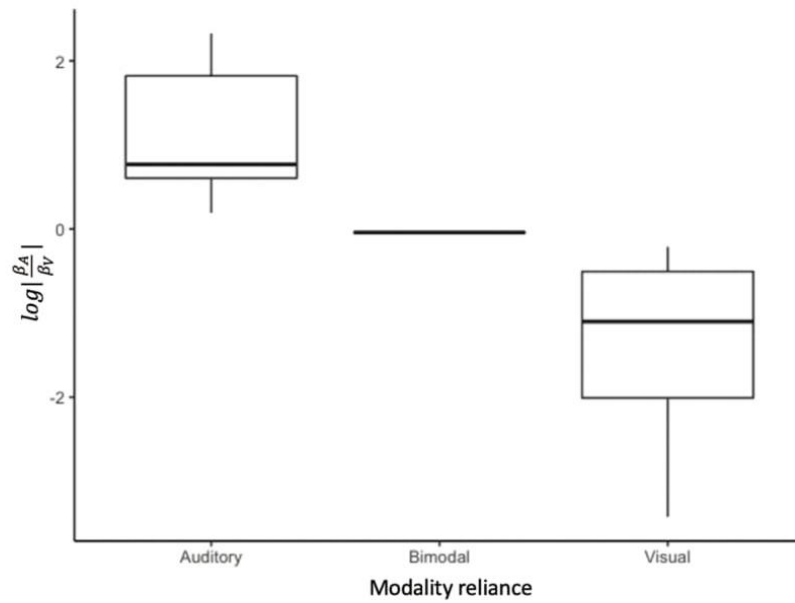


Figure 5. The distribution of  $\log \left| \frac{\beta_A}{\beta_V} \right|$  against the three modality types among participants: visual, auditory, or bimodal reliance.

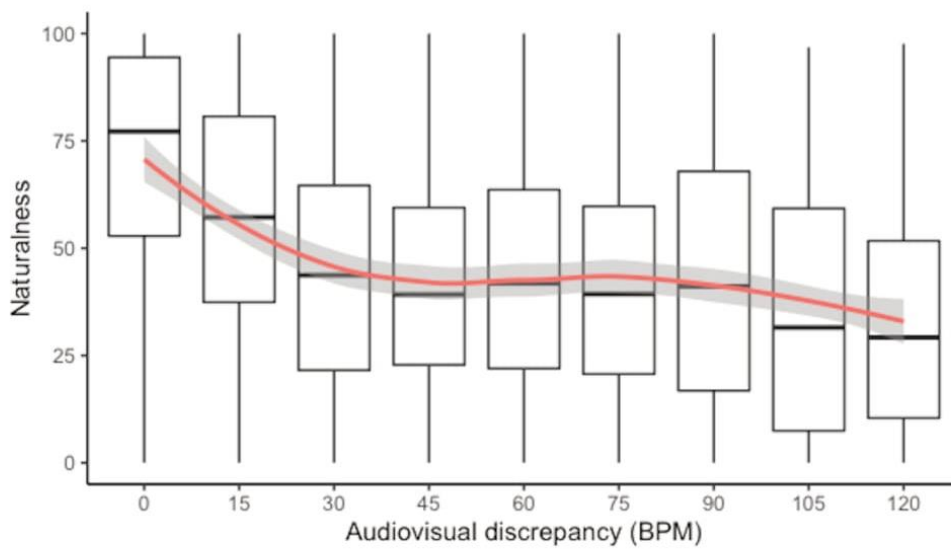


Figure 6. Perceived naturalness and audiovisual tempo discrepancies (absolute values), varying from small to large. Bottom and top of the boxes represent the first and third quartiles, and the centre line the median. The red curve represents the trend for perceived naturalness, as the audiovisual tempo discrepancy increases.

velocity cues that biased the perceived auditory tempo. Other studies (e.g., van Wassenhove et al., 2008) have shown that looming and receding discs alter the perceived duration of pure tones. Those studies have pointed out the importance of vision in audiovisual temporal perception, which was previously believed to be dominated by audition.

Other experiments that have adopted more complex audiovisual stimuli than abstract rhythmic sequences have successfully replicated the visual driving effect too. For example, participants were equally accurate in their discrimination of complex rhythmic patterns for both auditory and visual presentation (Grahn, 2012). Further attempts have taken biological movements into account. When watching vigorous dance movements, the music tempo was perceived as faster compared to the conditions where only music, or music and relaxed dance movements were presented (London et al., 2016). Our study also used point-light dance-like movements, which appeared to entrain the overall perceived tempo toward the visual tempo. This suggests that the preferred modality for tempo judgement may not entirely depend on the precision of the modality (i.e., the modality precision account). Rather, it might depend on how well information of this modality could assist with action prediction (i.e., the modality appropriateness account). As discussed earlier in the *Introduction*, the reliability of the sensory modality determines its contribution weight in the overall judgement (Hartcher-O'Brien et al., 2014; Shi et al., 2013). The more reliable the prediction from the signal, the higher the weight that would be assigned to that modality. Hence, our findings may suggest that predictable biological motions relative to drumbeats may offer reliable cues for temporal judgement.

This in particular is in line with human action prediction. Various studies have suggested that visual attention driven by the action goal was accompanied by higher processing efficiency (Decroix & Kalénine, 2019; Loucks & Pechey, 2016). Loucks and Nagel (2018) found more accurate tempo discrimination performance with human actions compared to non-human actions. In this vein, higher temporal sensitivity with goal-directed biological motions may also be reflected in the current study where repetitive dance-like movements became highly predictable, and thus endowed with more weights in tempo judgement. Compared to drumbeats, the biological movements provide more timing information than discreet bursts of sound. In other words, the continuous nature of the visual information may have provided a more reliable source of tempo information. When both modalities possess information of similar continuities, there might not be an advantageous modality in tempo judgement. In this regard, one of the earlier studies by Boltz (2005) found no modality effect on duration reproduction performances with continuous, natural human behaviours including sports or conversations, presented in either the auditory, visual, or audiovisual channel. However, it is as yet inconclusive whether the continuity or the biological plausibility affects the role of vision in temporal processing. To disentangle the effects of the two features,

Hove et al. (2010) examined the efficiency of facilitating audiomotor synchronisation with continuous and/or direction-compatible motions that were either abstract or biological trajectories. Higher synchronisation rates were observed with continuous, direction-compatible, but not necessarily biological motions. It can be speculated that continuous stimuli contribute more temporal information than discontinuous stimuli, regardless of forms (abstract vs biological) and modalities. In the same vein, the continuity of visual and auditory rhythms has direct effects on participants' timing performances: The sensory modality with continuous inputs was assigned more weight than the discontinuous one (Varlet et al., 2012). Similarly, compared with the discrete bursts of sounds, the continuous biological motions in our study might provide more reliable information in the overall tempo judgements.

It should be noted that the audiovisual source locations might also contribute to the weight in audiovisual judgements. In a study of multisensory simultaneity (Di Luca, Machulla, & Ernst, 2009), it has been shown that in the headphone-based relative to the co-location audiovisual presentation, the auditory estimates are likely to be biased as they are trusted less. However, the contribution of this spatial discrepancy to the visual-dominant temporal judgements, if any, is likely very mild, given that the headphone-based presentation potentially reduced the interference of other external sounds, which would potentially boost the reliability of the auditory modality. In a similar vein, visual and tactile stimuli appearing in different spatial location were associated with less accurate discrimination responses than those in the same location (Spence et al., 2008), indicating that the interference from one modality might pose a threat to the credibility of the other. In the current study, the auditory source that was closer to the participants should have provided a more reliable source of temporal information than the visual displays, yet failed to do so.

The discrepancy between auditory and visual tempi in our study was reflected by the naturalness ratings. The current results suggest that high perceived naturalness could be particularly derived from a fast visual tempo, as well as from high audiovisual temporal congruence. Not surprisingly, the smaller the audiovisual temporal discrepancy, the more natural the stimulus was perceived to be. Stimuli with small or no discrepancies presumably posed the least difficulty in binding multisensory inputs to one (Vatakis & Spence, 2008). Our results align well with past findings in that meaningful visual motion, especially when following an expected movement direction, has a strong impact on timing and, as shown in other research, sensorimotor synchronisation (Hove et al., 2010). This finding was supported by research comparing the effect of biological movements (finger tapping) with abstract visual stimuli (flashes) on timing accuracy, which found higher stability when synchronising with finger movements than with flashes (Hove & Keller, 2010).

There is scarce neurobiological evidence supporting the role of visual input compared to auditory input in temporal processing. Evidence for the latter including beat detection and time estimation, however, can be found mostly in studies using abstract unimodal stimuli (for a review, see Buonomano & Maass, 2009). An fMRI study where participants were asked to discriminate multisensory inputs (visual, auditory, and tactile) revealed that the auditory dorsal pathway was partly specific to beat processing and functioned as a supra-modal network (Araneda et al., 2017). In Kanai and colleagues' (2011) study, Transcranial Magnetic Stimulation (TMS) disrupted the activities in the auditory cortex and consequently impeded the participants' performances in a two-alternative-force choice task where two durations, presented either in pure tones or visual flickers, were compared. By contrast, disrupting the activities in the primary visual cortex only affected the performances of duration judgements with visual stimuli, suggesting the dominant role of the auditory cortex in temporal processing. The evidence above, nevertheless, may not generalise to neural mechanisms of timing with naturalistic stimuli. Biological motion carries spatiotemporal information that helps in forming action predictions, along with the timing of the action. The findings with behavioural data in the current study call for future neurobiological research on the visual dominance with meaningful visual stimuli such as continuous movements.

Furthermore, attentional processes might also contribute the 'visual driving' effect observed in our study. Visual dominance in spatial attention has been supported by ample studies. In an audiovisual context, the visual modality was associated with faster reaction time and less response errors in modality-switching, spatially-incongruent tasks (Lukas, Philipp, & Koch, 2010) and detected with greater sensitivity (Spence et al., 2012). The Colavita effect, more specifically, referred to a phenomenon where visual stimuli were associated with higher salience than auditory stimuli when both appeared simultaneously (Colavita, 1974). The perception of human biological motions in point-light displays led to even higher visual salience than abstract visual displays (Johansson, 1973). Selective attention oriented towards biological movement, particularly motion with a purpose compared to scrambled motion, has been shown to activate the part of the motor cortex associated with action mirroring (Gao et al., 2014). The 'imagined' imitation led to determination of action intention (Knoblich & Sebanz, 2008), in this case prediction of motion trajectories and their spatio-temporal information. The allocation of attention to natural motions (the PLD in the current study) could then explain participants' reliance for visual tempo. According to the Dynamic Attending Theory (Jones & Boltz, 1989), the pace of the internal clock is subject to the environmental rhythm when the limited attentional resources orient towards the rhythm of exogenous stimuli. Synchronising one's attentional 'pulses' with environmental rhythms is also known as the entrainment effect — in this case, participants were predominantly under the

influence of the visual tempo. The biological motion in the PLD was presumably attended to more often than the drumbeats, therefore dominating the perceiver's tempo judgement.

Both naturalistic biological motion and spatial attention towards to biological motion contribute to the visual driving effect as observed in the current study. Yet a few questions remain: firstly, by modifying the ecological validity of auditory stimuli (e.g., beats vs more complex music), it is possible to observe changes in overall tempo judgement. Secondly, instead of multisensory gain, it could be investigated whether there is also a multisensory loss, or more specifically, whether the presence of inconsistent multisensory information could impede temporal processing such as timing or duration judgement. Lastly, studies may explore whether other senses such as tactile perception are capable of multisensory integration and what their weights are in the timing process. A few studies have attempted to explore the potential of tactile-assisted metre judgement when auditory or visual stimuli were presented (Araneda et al., 2017; Huang et al., 2012). In a rhythm pattern identification task, for example, congruent vibrations (tactile) and tones (auditory) raised the correct rhythm discrimination rate to 90%, while incongruent inputs resulted in a decline (Huang et al., 2012). Interestingly, the correct rate was significantly higher when the dominant (correct) pattern was presented with sounds than with vibrations, again confirming the dominant role of audition.

There are a number of limitations that should be addressed in future studies. First, a response bias in the decision process, which might be reflected by participants' preference towards one modality under certain tempo conditions, cannot be fully ruled out. According to the causal inference model (Körding et al., 2007), response bias towards one source of information can be observed when the key feature (tempo) differs between two sources (sensory modalities) to an extreme extent. However, if such a response bias widely existed among the sample population, the stimuli with a large audiovisual gap, regardless of the modality of the fast tempo, should be equally often judged to be 'fast', which would be reflected by a (inverted) U-shaped threshold of equality in Fig. 3. Secondly, considering the essential role of naturalistic stimuli, the perceived predictability of the auditory beats and the PLD has not been quantitatively pre-evaluated. For visual stimuli, the predictability should take into account direction compatibility as in Hove and colleagues' (2010) work. In future studies, a baseline test prior to the experiment collecting familiarity and naturalness ratings, as well as eye fixation concerning the compatibility of motion, should be considered. The imbalance between the two modalities can be minimised by collecting the perceived naturalness of both auditory and visual stimuli respectively from multiple independent raters before the experiment commences. As for auditory stimuli, the predictability is frequently measured by sensorimotor synchronisation accuracy in tapping tasks (e.g., Stupacher et al., 2017). Furthermore, past studies tended to require the participants to judge the temporal information of one modality only (e.g., Klink, et

al., 2011). The advantages and disadvantages of an experimental paradigm that allows participants the liberty to exhibit their modality reliance should be further examined. In addition, the current study did not systematically evaluate differences between auditory and visual attention. Future studies should seek to disentangle the effects of multisensory attention, especially visual attention, from the effect of naturalistic stimuli on temporal judgements. To verify the link between a visual driving effect and direction compatibility, future studies should also consider a control condition in which inverted biological movements are presented.

Taken together, the current study provides evidence for a visual driving effect in multisensory tempo judgements with meaningful movements. On the group level, visual tempo contributed more to the overall tempo judgement than the auditory tempo. On the individual level, in addition, when presented with tempo-inconsistent audiovisual stimuli, more participants relied on the visual tempo to make the overall tempo judgements. The modality reliance provided insights into tempo judgement strategies adopted by different individuals. Future studies should further investigate the apparent dominance of visual information in timing with real-life audiovisual scenes as well as the factors influencing individuals' modality reliance in temporal judgements.

### Acknowledgements

This research was supported by a Consolidator Grant from the European Research Council to the second author. The research is part of the five-year project: "Slow motion: Transformations of musical time in perception and performance" (SloMo; Grant No. 725319). We are grateful to Emma Allingham for helpful comments on a previous version of the manuscript.

### References

- Allingham, E., Hammerschmidt, D., & Wöllner, C. (2020). Time perception in human movement: Effects of speed and agency on duration estimation, *Q. J. Exp. Psychol.* *74*, 559–572. doi:10.1177/1747021820979518.
- Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., & De Volder, A. G. (2017). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream, *Eur. J. Neurosci.* *45*, 1439–1450. doi: 0.1111/ejn.13349.
- Behne, K.-E., & Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audio-visual perception experiment and a replication, *Musicae Sci.* *15*, 324–342. doi:10.1177/1029864911410955.
- Boltz, M. G. (2005). Duration judgments of naturalistic events in the auditory and visual modalities, *Percept. Psychophys.* *67*, 1362–1375. doi:10.3758/BF03193641.
- Buonomano, D. V., & Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks, *Nat. Rev. Neurosci.* *10*, 113–125. doi:10.1038/nrn2558.

- Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration, *Exp. Brain Res.* 198, 49. doi:10.1007/s00221-009-1933-z.
- Burger, B., & Toivainen, P. (2013). MoCap Toolbox — A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the sound and music computing conference 2013* (pp. 172–178). Berlin, Germany: Logos Verlag.
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of Visual and motor familiarity in action observation, *Curr. Biol.* 16, 1905–1910. doi:10.1016/j.cub.2006.07.065.
- Chen, L., Zhou, X., Müller, H. J., & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration, *J. Exp. Psychol. Gen.* 147, 1851–1864. doi:10.1037/xge0000487.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.
- Colavita, F. B. (1974). Human sensory dominance, *Percept. Psychophys.* 16, 409–412. doi:10.3758/BF03203962.
- Decroix, J., & Kalénine, S. (2019). What first drives visual attention during the recognition of object-directed actions? The role of kinematics and goal information, *Atten. Percept. Psychophys.* 81, 2400–2409. doi:10.3758/s13414-019-01784-7.
- Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency, *J. Vis.* 9, 7. doi:10.1167/9.12.7.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses, *Behav. Res. Methods* 41, 1149–1160. doi:10.3758/BRM.41.4.1149.
- Gamache, P.-L., & Grondin, S. (2010). Sensory-specific clock components and memory mechanisms: investigation with parallel timing, *Eur. J. Neurosci.* 31, 1908–1914. doi:10.1111/j.1460-9568.2010.07197.x.
- Gao, Z., Lu, X., Shen, M., Shui, R., & Chen, S. (2014). Rehearsing biological motion in working memory: an fMRI study, *J. Vis.* 14, 1009. doi:10.1167/14.10.1009.
- Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time, *Front. Integr. Neurosci.* 6, 77. doi:10.3389/fnint.2012.00077.
- Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms, *Exp. Brain Res.* 220, 51–61. doi:10.1007/s00221-012-3114-8.
- Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). fMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa, *NeuroImage*, 54, 1231–1243. doi:10.1016/j.neuroimage.2010.09.033.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions, *Atten. Percept. Psychophys.* 72, 561–582. doi:10.3758/APP.72.3.561.
- Grondin, S., Gamache, P.-L., Tobin, S., Bisson, N., & Hawke, L. (2008). Categorization of brief temporal intervals: An auditory processing context may impair visual performances, *Acoust. Sci. Technol.* 29, 338–340. doi:10.1250/ast.29.338.
- Hammerschmidt, D., & Wöllner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time, *Music Percept.* 37, 263–277. doi:10.1525/mp.2020.37.4.263.



- Hartcher-O'Brien, J., Luca, M. Di, & Ernst, M. O. (2014). The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion, *PLoS ONE* 9, e89339. doi:10.1371/journal.pone.0089339.
- Hove, M. J., & Keller, P. E. (2010). Spatiotemporal relations and movement trajectories in visuomotor synchronization, *Music Percept.* 28, 15–26. doi:10.1525/mp.2010.28.1.15.
- Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization, *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1525–1534. doi:10.1037/a0019059.
- Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. H. (2013). Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome, *Psychol. Res.* 77, 388–398. doi:10.1007/s00426-012-0441-0.
- Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., & Hsiao, S. (2012). Feeling music: integration of auditory and tactile inputs in musical meter perception, *PLoS ONE* 7, e48496. doi:10.1371/journal.pone.0048496.
- Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals, *Cognition* 134, 232–244. doi:10.1016/j.cognition.2014.10.018.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis, *Percept. Psychophys.* 14, 201–211. doi:10.3758/BF03212378.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time, *Psychol. Rev.* 96, 459–491. doi:10.1037/0033-295X.96.3.459.
- Kanai, R., Lloyd, H., Buetti, D., & Walsh, V. (2011). Modality-independent role of the primary auditory cortex in time estimation, *Exp. Brain Res.* 209, 465–471. doi:10.1007/s00221-011-2577-3.
- Klink, P. C., Montijn, J. S., & van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates, *Atten. Percept. Psychophys.* 73, 219–236. doi:10.3758/s13414-010-0010-9.
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action, *Phil. Trans. R. Soc. B Biol. Sci.* 363, 2021–2031. doi:10.1098/rstb.2008.0006.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception, *PLoS ONE* 2, e943. doi:10.1371/journal.pone.0000943.
- Levitani, C. A., Ban, Y.-H. A., Stiles, N. R. B., & Shimojo, S. (2015). Rate perception adapts across the senses: evidence for a unified timing mechanism, *Sci. Rep.* 5, 8857. doi:10.1038/srep08857.
- London, J., Burger, B., Thompson, M., & Toiviainen, P. (2016). Speed on the dance floor: Auditory and visual cues for musical tempo, *Acta Psychol.* 164, 70–80. doi:10.1016/j.actpsy.2015.12.005.
- Loucks, J., & Nagel, N. (2018). Temporal perception is enhanced for goal-directed biological actions, *Vis. Cogn.* 26, 530–544. doi:10.1080/13506285.2018.1516708.
- Loucks, J., & Pechey, M. (2016). Human action perception is consistent, flexible, and orientation dependent, *Perception*, 45, 1222–1239. doi:10.1177/0301006616652054.
- Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities: further evidence for visual dominance, *Psychol. Res.* 74, 255–267. doi:10.1007/s00426-009-0246-y.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences, *Behav. Res. Methods* 44, 314–324. doi:10.3758/s13428-011-0168-7.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* 264, 746–748. doi:10.1038/264746a0.
- Ocelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in temporal perception, *Psychon. Bull. Rev.* 18, 429–454. doi:10.3758/s13423-011-0070-4.

- Ortega, L., Guzman-Martinez, E., Grabowecy, M., & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability, *Atten. Percept. Psychophys.* 76, 1485–1502. doi:10.3758/s13414-014-0663-x.
- Penney, T. (2003). Modality differences in interval timing: Attention, clock speed, and memory. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 209–233). Boca Raton, FL, USA: CRC Press/Routledge/Taylor & Francis Group. doi:10.1201/9780203009574.
- Repp, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision, *J. Mot. Behav.* 35, 355–370. doi: 10.1080/00222890309603156.
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration, *Proc. R. Soc. B Biol. Sci.* 273, 2159–2168. doi:10.1098/rspb.2006.3578.
- Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration, *Perception* 36, 888–897. doi:10.1068/p5635.
- Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval, *Exp. Brain Res.* 203, 723–735. doi:10.1007/s00221-010-2286-3.
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception, *Trends Cogn. Sci.* 17, 556–564. doi:10.1016/j.tics.2013.09.009.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker, *Science* 145, 1328–1330. doi:10.1126/science.145.3638.1328.
- Spence, C., F. Pavani, A. Maravita & N.P. Holmes. (2008). Multi-sensory interactions. In M.C. Lin & M.A. Otaduy (Eds), *Haptic rendering: foundations, algorithms, and applications* (pp. 21–52). Wellesley, MA, USA: AK Peters/CRC Press. doi:10.1201/b10636.
- Spence, C., Parise, C., & Chen, Y. C. (2012). The Colavita visual dominance effect. In M. M. Murray & M. T. Wallace (Eds), *The neural bases of multisensory process* (pp. 523–550). Taylor & Francis Group. doi:10.1201/b11092.
- Stupacher, J., Witte, M., & Wood, G. (2017). Go with the flow: Subjective fluency of performance is associated with sensorimotor synchronization accuracy and stability, *Proc. 25th Anniv. Conf. Eur. Soc. Cogn. Sci. Mus.*, Ghent, Belgium, 163–166.
- Su, Y.-H., & Jonikaitis, D. (2011). Hearing the speed: Visual motion biases the perception of auditory tempo, *Exp. Brain Res.* 214, 357–371. doi:10.1007/s00221-011-2835-4.
- Su, Y. H., & Salazar-López, E. (2016). Visual timing of structured dance movements resembles auditory rhythm perception, *Neural Plast.* 2016, 1678390. doi:10.1155/2016/1678390.
- Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency, *Perception* 19, 705–743. doi:10.1068/p190705.
- Van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses, *PLoS ONE* 3, e1437. doi:10.1371/journal.pone.0001437.
- Varlet, M., Marin, L., Issartel, J., Schmidt, R. C., & Bardy, B. G. (2012). Continuity of visual and auditory rhythms influences sensorimotor coordination, *PLoS ONE* 7, e44082. doi:10.1371/journal.pone.0044082.
- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli, *Acta Psychol.* 127, 12–23. doi:10.1016/j.actpsy.2006.12.002.
- Wang, X., & Wöllner, C. (2019). Time as the ink that music is written with: a review of internal clock models and their explanatory power in audiovisual perception, *Jahrb. Musikpsychol.* 29, e67. doi:10.5964/jbdgm.2019v29.67.

## Study 3

### **Body Movement and Emotion: Investigating the Impact of Audiovisual Tempo Manipulations on Emotional Arousal and Valence**

Xinyue Wang, Birgitta Burger, and Clemens Wöllner

*Jahrbuch Musikpsychologie*, 32, e191




<https://doi.org/10.5964/jbdgm.191>

Reproduced with kind permission from Jahrbuch Musikpsychologie

## Research Reports

### **Body Movement and Emotion: Investigating the Impact of Audiovisual Tempo Manipulations on Emotional Arousal and Valence**

Körperbewegung und Emotion: Untersuchung der Auswirkungen von audiovisueller Tempomanipulationen auf emotionale Erregung und Valenz

Xinyue Wang\*<sup>1</sup> , Birgitta Burger<sup>1</sup> , Clemens Wöllner<sup>2</sup> 

[1] Institute for Systematic Musicology, Universität Hamburg, Hamburg, Germany. [2] University of Music Freiburg, Freiburg, Germany.

#### **Abstract**

The perception of emotions in humans moving can be influenced by several movement features such as fluidity, complexity, and tempo. Manipulations of movement tempo by acceleration or deceleration are widely employed in media, yet there has been limited research on how these affect perceived emotions. The current study examined how tempo-manipulated point-light displays (PLDs) of human dance-like movements, compared to their tempo-original counterparts, influence the perceived emotional arousal and valence by altering the tempo and hence the underlying movement characteristics. In an online perceptual experiment, we presented tempo-original and tempo-manipulated PLDs at three tempi, with and without synchronized drumbeats. Participants were asked to judge the PLDs based on their emotional arousal, valence, and naturalness. Results suggest that movements with higher fluidity were perceived as lower in arousal than those with low fluidity. Stimuli higher in naturalness were perceived to be more positive in valence than those low in naturalness. Audiovisual stimuli, including the drumbeats, received higher arousal but lower valence ratings than visual-only stimuli. Furthermore, decelerated movements were associated with increased fluidity compared to accelerated or tempo-original ones. Tempo deceleration was associated with lower naturalness compared to tempo-original movements. These findings indicate that tempo manipulation can be identified via kinematic feature changes, consequently altering emotional attributes perceived in the movements.

*Keywords:* tempo manipulation, fluidity, kinematic features, Point-Light Displays (PLDs), emotions, motions, naturalness

#### **Zusammenfassung**

Die Wahrnehmung von Emotionen in menschlichen Bewegungen kann durch verschiedene Bewegungsmerkmale wie Fluidität, Komplexität und Tempo beeinflusst werden. Es gibt jedoch nur wenige Untersuchungen darüber, wie sich die Manipulation des Tempos auf die wahrgenommenen Emotionen auswirkt, wenn das Präsentationstempo einer Bewegung beschleunigt oder verlangsamt wird. In der vorliegenden Studie sollte untersucht werden, wie tempomanipulierte Point-Light-Displays (PLDs) tänzerischer menschlicher Bewegungen im Vergleich zum Originaltempo die wahrgenommene emotionale Erregung und Valenz beeinflussen, wenn auch die zugrundeliegenden Bewegungsmerkmale durch das Tempo verändert sind. In einem Online-Wahrnehmungsexperiment präsentierten wir tempo-originale und tempo-manipulierte PLDs in drei Tempi, mit oder ohne synchronisierten Schlagzeugklängen. Die Teilnehmer beurteilten die emotionale Erregung, Valenz und Natürlichkeit der Bewegungen. Die Ergebnisse deuten darauf hin, dass Bewegungen mit höherer Fluidität als weniger erregend, und Bewegungen mit höherer Natürlichkeit als positiver in ihrer Valenz wahrgenommen wurden als diejenigen mit niedriger Natürlichkeit. Audiovisuelle Stimuli mit den Schlagzeugklängen wurden höher in der Erregung, aber niedriger in der Valenz bewertet als visuelle Stimuli. Darüberhinaus zeigen die Analysen, dass verlangsamte Bewegungen höhere Werte in der Fluidität erzielten als beschleunigte oder im ursprünglichen Tempo ausgeführte Bewegungen. Die Verlangsamung des Tempos wurde jedoch auch mit einer geringeren Natürlichkeit in Verbindung gebracht als das ursprüngliche Tempo. Diese Ergebnisse deuten darauf hin, dass die Manipulation des Tempos zu Veränderungen der kinematischen Merkmale führt, wodurch sich die in den Bewegungen wahrgenommenen emotionalen Attribute verändern.

*Schlüsselwörter:* Tempo-Manipulation, Fluidität, Kinematische Merkmale, Point-Light-Displays (PLDs), Emotionen, Bewegungen, Wahrnehmungsexperiment, Natürlichkeit

Jahrbuch Musikpsychologie, 2024, Vol. 32, Article e191, <https://doi.org/10.5964/jbdgm.191>

Received: 2024-03-18. Accepted: 2024-07-15. Published (VoR): 2024-07-31.

Reviewed by: Friedrich Platz; Kathrin Schlemmer.

\*Corresponding author at: Institute for Systematic Musicology, Universität Hamburg, Universität Hamburg, Mittelweg 177, 20148 Hamburg, Germany. E-mail: philippapw@gmail.com



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

Since the advent of moving images and audiovisual media, tempo has been manipulated in various forms. Media consumers, be it for videos on social networks, commercial films, or video clips of musical performances, are increasingly confronted with displays of human motion that deviate from standard movement patterns. Given the high relevance of body movements and visual information on the perception of musical performances (Platz & Kopiez, 2012) and the vital link between tempo, arousal, and emotion (Droit-Volet et al., 2013), this study investigated how tempo manipulation in point-light recordings of dance movements, accompanied by drumbeats, were emotionally perceived in relation to original, un-manipulated movements at the same speeds.

Manipulating time by slowing down or speeding up has been a common technique for elevating emotions in films (Wöllner et al., 2018), commercial video marketing (Yin et al., 2021), or the replay of highlights in sports matches and games (Pan et al., 2001). In the film *Matrix* (Wachowski & Wachowski, 1999), when Neo bent backward to dodge the bullets in slow motion, film critics called this moment “bullet time” and referred to this moment as majestic and “incredibly cool to see, even after so many years since the film’s release in 1999” (Dutta, 2022, para. 1). In contrast, accelerated visual scenes usually speed up the processes that otherwise span over hours and longer in standard time or create an amusing effect such as jittered human movements. Understanding the effects of tempo manipulation on emotions can help leverage such tools by acceleration and deceleration to capture attention (Compton et al., 2003), enhance memories (Levine & Pizarro, 2004), or change the perceived time (Droit-Volet & Berthon, 2017). In music and dance practices, by manipulating tempo, artists can shift the emotional expressions in their performances to create unique interpretations of the same work (Quinn & Watt, 2006). Moreover, tempo manipulation may also be useful in music therapy when assisting treatment (Brownlow, 2017) or regulating emotional experiences.

Relatively few studies have shed light on how tempo manipulations affect emotions. Decelerated slow-motion film scenes, ballet, and sports excerpts were perceived as lower in emotional arousal and more positive in valence than tempo-matched, real-time footage (Wöllner et al., 2018). These results were also present in the physiological responses of participants. Further analyses with tempo-adapted film scenes found smaller pupil dilation and higher fixation ratios with slow-motion than with the real-time videos (Hammerschmidt & Wöllner, 2018), suggesting a lower arousal level and increased attention to detail. The findings are consistent with research that found decreased arousal with slow music (Droit-Volet et al., 2013). Furthermore, manipulation of movement speed is frequently employed in popular culture to enhance the audience’s attention to the scenes’ details and increase aesthetic pleasure (Nebens, 2021). However, to our knowledge, no studies have directly examined the effects of tempo manipulation on the two dimensions of emotions: Arousal and valence.

To examine the link between movement speeds and emotions, it is crucial to define how emotions are theorized and measured. Research has developed several systems to describe emotions (Mauss & Robinson, 2009). Among them, the

circumplex model (Russell, 1980), which entails a two-dimensional system of valence and arousal, has been adopted to evaluate facial expressions (Calder et al., 2001) and human movements (Pollick et al., 2001). Following Pollick et al.'s (2001) research, several studies suggest that emotion inferences were also possible from static body postures (Coulson, 2004), gestures (Castillo & Neff, 2019), gait (Kang & Gross, 2016), social interaction (Clarke et al., 2005), movements in joint musical improvisation (Wöllner, 2020), and dance movements (Burger & Toiviainen, 2020a). Together, these findings indicate a connection between movement features and the two emotion dimensions.

Studies further investigated the link between emotions, kinematic features, and movement angularity, the latter referring to the perspectives and geometrical relations of movements in a three-dimensional space. Castro and Boone (2015) found a connection between the accuracy of emotion perception and movement angularity. In this case, happy and sad body postures, being located respectively in the circumplex model, were recognized more accurately when participants were more sensitive toward the angles of geometric line patterns. Kinematic features may be critical indicators of emotions when viewing full-body movements. Several kinematic features have been identified to relate to specific basic emotions. A further study on the motion-emotion link revealed that music-induced emotions, reflected in spontaneous dance movements, were characterized by a set of kinematic features (Burger et al., 2013). In this regard, movement fluidity provides an index quantifying the smoothness of the motion, whereas complexity relates to how many dimensions or directions of the movement are used, with higher dimensionality being more complex than movements of lower dimensionality. Burger et al.'s (2013) study indicated that high fluidity correlated with music-inducing emotions of low arousal yet moderately positive valence. In contrast, high movement complexity correlated with music that induced emotions of moderately high arousal and positive valence. This finding was in line with other studies that suggested similar features, and movement fluidity and complexity were reliable indicators of basic emotional states such as happiness or sadness when enacting (Van Dyck et al., 2013) and perceiving human motions (Camurri et al., 2003; Montepare et al., 1999).

Tempo changes influence kinematic features. A study with professional drummers performing in various tempi found that movement fluidity was higher with slow compared to fast tempo, while movement complexity was the highest with slow tempo when the amount of drummer's movement peaks (Burger & Wöllner, 2023). The finding indicates the influence of tempo but not of tempo manipulation. It is yet to be found whether acceleration and deceleration of motion affect the kinematic features in visual displays. Motion capture with point-light displays (PLDs) is often used to derive kinematic features. PLDs of human movements, initially developed by Johansson (1973), entails the technique that extracts movement patterns by attaching reflective markers to human bodies and recording the movement of these markers using an optical motion capture system, receiving an accurate 3-dimensional representation of the movement (e.g., Burger & Toiviainen, 2020b). PLDs have been widely used in emotion research. A study investigating participants' discrimination sensitivity towards fluctuations in emotional intensity with PLDs suggested that dynamic PLDs were associated with higher sensitivity than static PLDs on a level comparable to full-light displays (Atkinson et al., 2004). The studies showed the potential of accurate emotion recognition with dynamic PLDs. In addition, PLD does not carry confounding factors such as the age or clothing of the performer, making it ideal for experimental designs.

Parallel to kinematic features, movement tempo influences the perceived emotions. Temporal attributes of PLDs were found to strongly predict the actor's emotional states in motion-emotion research (Pollick & Paterson, 2008). Fast movements were often associated with happiness and anger, while slow movements were associated with sadness and a neutral mood (de Meijer, 1989; Montepare et al., 1999; Pollick et al., 2001; Roether et al., 2009). Taking this a step further, according to the circumplex model (Feldman Barrett & Russell, 1998), a fast tempo was related to a high arousal level and a slow tempo to low arousal. In contrast, slow and fast movements were found at both ends of the

spectrum of the valence scale. While the link between tempo and arousal has often been observed, the correlation between movement tempo and emotional valence is less clear. It cannot be ruled out that a specific tempo represents multiple emotions regarding valence. In a passive viewing task, fast movements were rated angry or happy, while slow movements were rated sad or neutral (Montepare et al., 1999). The ambiguity of the tempo-valence relationship in identifying dance-conveyed emotions was echoed by studies adopting local movements such as knocking (Gross et al., 2010), drinking (Pollick et al., 2001), and walking (Roether et al., 2009). The findings, therefore, call upon further validations, which are part of the goals of the current study.

In addition to tempo, a movement's natural appearance may affect the perceived emotions. The tempo-original and -manipulated movements may be distinguishable from how natural they look. Despite few studies on this topic, naturalness has been frequently used to measure the validity of artificially generated motion in virtual reality (Knopp et al., 2019). Highly natural movements represent a crucial overlap between the temporal patterns of the virtual and realistic movements that match human prediction. Interestingly, Nilsson and colleagues (2015) found that the tempo thresholds for natural treadmill walking in virtual reality are higher than walking in place (WIP), suggesting expectations for the pace of perceptually natural walking depend highly on contexts (treadmill or WIP). The definition of natural walking tempo differs by whether a person is moving forward or not. The finding indicates the possibility that viewers who controlled the avatar for walking in virtual space also considered the temporal attributes of different types of motion. Similarly, when viewing real-time compared to decelerated or accelerated movements, naturalness may be tied to how much they fit with typical, realistic movement features at the corresponding speeds. However, research has yet to explore the connection between tempo and tempo manipulation and perceived naturalness. Although no direct evidence has been found, Chen and colleagues' (2023) study revealed that prototypical (most representative of the motion type) walking was perceived as more natural and aesthetically pleasing than atypical walking. This study suggested that high naturalness mediates the effects of visual attributes on aesthetic pleasure. Such an effect might be extended to positive emotional valence. The current study aims to investigate whether tempo manipulation affects the perceived naturalness of body movements and, if yes, whether increases in naturalness affect emotional valence positively.

Similar to tempo, the sensory modality also affects the perceived emotions. Preferences for multimodal rather than unimodal information were found for an emotion recognition task, in which facial expressions of fear and disgust, with or without vocal sounds of the consistent emotions, were presented to viewers (Collignon et al., 2008). A significant improvement in emotion discrimination performances was observed when PLDs of human movements were integrated with voices that expressed affective states, such as anger and fear, compared to neutral voices (Jessen et al., 2012). Multimodal information also increases perceived arousal. A study investigated how visual kinematic features and auditory information contribute to emotion perception in musical performances (Vuoskoski et al., 2016). The findings suggest that the audiovisual condition was perceived higher in emotional arousal than the visual-only condition, while the latter was also rated less positive in valence. The impact of tempo (e.g., comparing 72 BPM to 184 BPM in Droit-Volet et al., 2013; Wöllner et al., 2018), particularly tempo-manipulated movements, is yet to be investigated. Tempo acceleration and deceleration are typically accompanied by auditory information in movies, commercials, or sports, giving rise to strong emotional responses (e.g., Wöllner et al., 2018; Yin et al., 2021). Understanding the effect of sensory modalities in tempo manipulation will shed light on the realistic use of movements in audiovisual media. Therefore, another goal of the current study is to compare visual and audiovisual presentations of the same movements in tempo-original and -manipulated conditions for their effects on the perceived emotions.

Taken together, the current study aims to investigate how real-time (original), decelerated, and accelerated (audio-) visual stimuli influence perceived emotions. With the results, we hope to provide insights into how the presentations

of music and dance performances affect the viewers' emotional experiences. Additionally, we hope to offer solutions to regulate one's moods by managing the tempo of videos and music playlists in everyday life.

We predict that tempo manipulation affects the movement features, which in turn affects emotional valence and arousal. When movements are presented at decelerated speeds, fluidity should increase, and complexity decrease, leading to lower perceived arousal and higher perceived valence than the tempo-original condition (Burger et al., 2013). Furthermore, we hypothesize that tempo manipulation as an independent variable affects naturalness as a dependent variable: The larger the extent of manipulation (manipulated – original tempo), the less natural body movements are perceived. We further expect that naturalness acts as a mediator to the effect of tempo manipulation on perceived valence and arousal. The larger the extent of manipulation, the less natural the stimulus is perceived, and the more negative valence and the lower arousal should be perceived. Therefore, naturalness will act as an independent variable in this analysis to predict changes in emotional arousal and valence as dependent variables. According to Pollick et al. (2001), we hypothesize that a faster presentation tempo leads to higher perceived arousal. Finally, the presentation modality is expected to moderate the perceived emotions such that the presence of auditory drumbeats, synchronized with the visual movements, should lead to higher arousal and more positive valence.

## Method

### Participants

To determine the minimum sample required to test our hypotheses, a prior power analysis was conducted using G\*Power Version 3.1.9 (Faul et al., 2009), suggesting a minimum of 53 participants for achieving 80% power, a medium effect size ( $f^2 = 0.3$ ) for linear multiple regression (fixed model; significance level  $\alpha$  set at 0.05). An international sample of 62 participants was recruited for an online experiment using the platform SoSci Survey (Leiner, 2019) (29 females, one gender undisclosed;  $M_{age} = 29.23$  years,  $SD_{age} = 8.83$ ). According to the Music Training dimension from the Goldsmith Music Sophistication Index (Müllensiefen et al., 2014), participants have been trained for an average of 3.31 years ( $SD = 4.06$ ). The Dance Training dimension from the Goldsmith Dance Sophistication Index (Rose et al., 2020) suggested that participants, on average, have 0.79 years of active dance training ( $SD = 1.63$ ). This suggests a low professional music and dance training prevalence in the sample population. The majority of the sample had achieved bachelor's ( $N = 22$ ) or master's ( $N = 27$ ) degrees, standing for a generally well-educated group (overall 79%). The study was approved by the Ethics Committee at the Faculty of Humanities, University of Hamburg, and participants provided their informed consent before the study. A lottery of two prizes worth € 30 was carried out at the end of data collection, including those who had opted to leave their email address.

### Apparatus

The stimuli consisted of visual and audiovisual presentations of human movements at three tempi: 86 (slow), 130 (medium), and 195 BPM (fast). The visual stimuli, detailed in Allingham et al. (2021), were human movements recorded by an eleven-camera motion-capture system (Qualysis Oqus 700) at 200 frames per second (framerate). The performer (male, 32-year-old) jumped from one leg to the other while raising the arms parallel to the ground, flexing and extending the wrists ipsilaterally to the leg motions. The movements, each lasting 10 seconds, entailed bilateral hand flaps and left-right jumps (see Figure 1) and were recorded at the three above-mentioned original tempi. The MATLAB Motion Capture (MoCap) Toolbox (Burger & Toiviainen, 2013) was used to time-shift the original data and create



animations that matched each original movement with the other two tempi; for example, the performance at 130 BPM was slowed down to match 86 BPM as well as sped up to 195 BPM (see Figure 2).

**Figure 1**

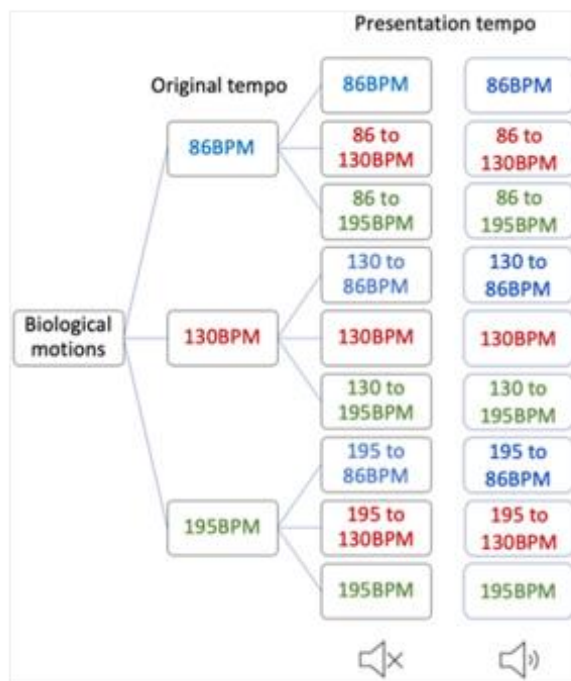
*Depictions of the Biological Motions Used as Visual Stimuli*



Note. The movements entail flexion and extension of the hands and lifting the left and right foot in turn.

**Figure 2**

*Visualization of Tempo Manipulations With the Biological Motions*



Note. The movements, recorded initially at 86 BPM, 130 BPM, and 195 BPM, were accelerated or decelerated to match the other two tempi. The same presentation tempo movements are marked in the same color (86 BPM: Blue, 130 BPM: Red, 195 BPM: Green).

The tempo transformation yielded nine video excerpts: slow-original, slow-to-medium, slow-to-fast, medium-original, medium-to-slow, medium-to-fast, fast-original, fast-to-slow, fast-to-medium. A slow-to-fast stimulus, for example, has an original tempo of 86 BPM and a presentation tempo of 195 BPM. Additionally, nine temporally synchronized audiovisual presentations were produced by aligning the visual material with an auditory drum beat (sequence of isochronous beats at the respective BPM synthesized in an online beat generator, <https://drumbit.app>, in Apple iFilm 10.1.12). In addition, the stimulus set included eight catch trials, which varied in duration to examine if participants paid attention to the displays. Half of the catch trials lasted 5 seconds, while the other half took 15 seconds. All of them were presented with synchronized drumbeats. In total, 18 experimental trials and eight catch trials were created.

## Procedure

Invitations to the online experiment on SoSci Survey (<https://www.sosicisurvey.de>) were distributed through email lists and social media sites. Participants were provided information about the experiment and asked to give their informed consent. Due to the restrictions of online experiments, no control was imposed on screen resolution, the distance to the screen, or the device through which sounds were played. However, at the beginning of the experiment, explicit instructions were given that participation should take place in a quiet environment on a computer with compatible browsers (Chrome or Firefox) and with head- or earphones. A 15-second music excerpt was presented to test the sound volume. Participants were instructed to adjust to a comfortable sound level and to keep the level consistent throughout the experiment. They were then asked to fill in demographic information and short questionnaires about their music training and active dance experience (Factor 3 Musical Training from the Goldsmith Musical Sophistication Index, Müllensiefen et al., 2014, and Factor 4 Dance Training from the Goldsmith Dance Sophistication Index, Rose et al., 2020).

In the experiment, participants were presented with two blocks of randomized stimuli, each consisting of 18 experimental trials, including the complete sets of audiovisual and visual-only presentations and four catch trials balanced in lengths (two for 5 s and two for 15 s). The stimulus (700 x 394-pixel resolution) appeared at the center of the screen. A repeated-measures design was used to control for within-subject variability and increase the experiment's efficiency. Following each stimulus, participants were asked to rate emotional arousal from 1 *calm* to 7 *excited*, emotional valence from 1 *negative* to 7 *positive*, and naturalness from 1 *unnatural* to 7 *natural*. No time restriction was imposed, though the video could be watched only once. A test trial using a different dancer was presented at the beginning to familiarize the participants with the experiment. With regard to the catch trials, all participants differentiated reliably between the different durations of catch trials and experimental trials,  $F(2, 2481) = 68.46, p < .001$ . The effect size ( $\eta_p^2 = 0.05$ ) indicates a medium effect. Therefore, no participant was excluded.

## Analyses

In the first analysis, the model was intended to determine whether tempo manipulation influenced the perceived naturalness, which, in turn, might influence emotional arousal and valence. A mixed linear regression was conducted. A group of independent variables was adopted to identify possible predictors of the perceived naturalness from a group of relevant variables: Movement complexity, fluidity, presentation tempo, tempo manipulation, and stimulus modality. The multilinear regression can be found in Equation I in the Appendix. The model was selected based on the lowest Akaike Information Criterion (AIC) values. To select the model of the highest goodness of fit for each dependent variable, maximum likelihood ratio (MLR) tests were conducted (see Table A4 in the Appendix). In the MLR tests, predictors were added one after another from the baseline model, in which only the random effects were present. The variances of

participants and conditions were considered as random effects. While adding significantly to the previous model, models with the lowest AIC were considered the final models.

The dependent variable for Equation I is perceived naturalness, and the participants rated it on a 7-point scale in response to the question, "Please rate how natural the movement feels." 1 represents the *least natural* movement, and 7 represents the *most natural* movement. The predictor variables were created in the same way as the variables for the second and third models (Equations II and III).

The second and third analyses also used a mixed linear regression to identify possible emotional arousal and valence predictors from relevant variables: Movement complexity, fluidity, perceived naturalness, presentation tempo, tempo manipulation, and stimulus modality. Post-hoc analyses with Bonferroni correction were conducted to follow up on significant main and interaction effects. The independent variables selected include movement complexity and fluidity. The models can be found in Equations II and III in the Appendix.

The dependent variables from the arousal and valence models include the following:

- Perceived arousal: Ratings from 1 (*calm*) to 7 (*excited*).
- Perceived valence: Ratings from 1 (*negative*) to 7 (*positive*) in response to the question, "Please rate whether the emotion of the video is negative or positive."

The independent variables from all models above include the following:

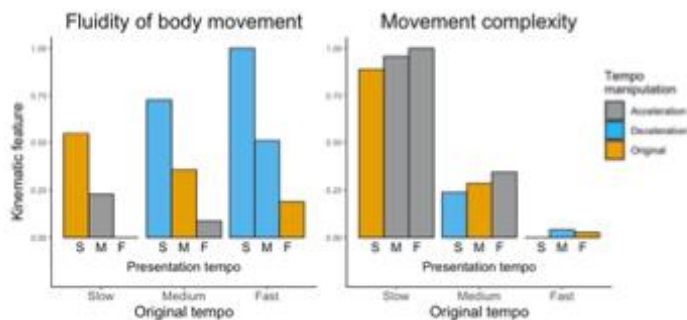
- Fluidity of body movement: The smoothness of the movement (the ratio between velocity and acceleration). This variable had been standardized as follows: 1 (least fluid) to 2 (most fluid) by the equation:  $(x - \min(x)) / (\max(x) - \min(x)) + 1$
- Movement complexity: This refers to the dimensionality of the movements, which is based on a Principal Component Analysis of the movement data. In this case, the movement is considered simple when the first five principal components can explain a large amount of the variance. The movement would be more complex when the first five components explain less variance. See Burger et al. (2013) for more explanation of the features. This variable has also been standardized from 1 (least complex) to 2 (most complex).
- Modality: The modality (visual-only, audiovisual) of each stimulus.
- Presentation tempo: The tempo presented to the participants after manipulation. For tempo-original stimuli, the presentation tempo equals the original tempo.
- Tempo manipulation: The extent and direction of tempo manipulation, where -1 represents the largest deceleration, 0 represents no manipulation, and 1 represents the largest extent of acceleration. The index is calculated as follows: First, calculate the gap between the manipulated and original tempo as  $x$ , set the new anchor to -1 (new min), representing the largest gap in minus). 1 (new max) represents the largest gap above 0. Then run linear standardization of the value  $((x - \min(x)) / (\max(x) - \min(x))) * (\text{new\_max} - \text{new\_min}) + \text{new\_min}$ .
- Perceived naturalness: Ratings from 1 (*least natural*) to 7 (*most natural*) were selected as a control variable to disentangle the contributions of the movement features to changes in emotional valence and arousal.

## Results

Two kinematic features were extracted from the motion capture data of the nine tempo-original and tempo-manipulated performances. Since original *performance tempi* were tempo-manipulated in two directions (acceleration, deceleration) or stayed the same, movements at the same *presentation tempo* may have different fluidity and complexity features (see Figure 3), depending on whether they were accelerated, decelerated, or performed initially at this speed.

**Figure 3**

*Descriptive Plots of Movement Fluidity (Left Pane) and Complexity (Right Pane)*



Note. Movement fluidity and complexity are normalized on a scale of 0 (least fluid/complex) to 1 (most fluid/complex) of tempo-manipulated and original movements in different original tempo conditions. Note that each bar represents one stimulus. For the presentation tempo, S = Slow, M = Medium, and F = Fast. An example of how to read this figure: When the original performance tempo is slow, and the presentation tempo is medium, the stimulus falls into the category "acceleration," color-coded as grey.

Pearson correlation coefficients were computed to examine the linear relationship between presentation tempo and movement features. Negative correlations between tempo and fluidity ( $r = -0.89, p < .001$ ) and between manipulation (acceleration) and fluidity ( $r = -0.92, p < .001$ ) were found. Thus, the faster the presentation tempo and the larger the extent of tempo acceleration, the less fluid the body movements. Positive correlations were found between tempo and complexity ( $r = 0.09, p < .001$ ) and between acceleration and complexity ( $r = 0.70, p < .001$ ), indicating that the faster the presentation tempo, the larger the extent of acceleration, the more complex the movements.

### The Effects of Movement Features on Perceived Naturalness

The linear mixed models revealed significant main effects of tempo manipulation on the perceived naturalness. A significant two-way interaction between the presentation tempo and manipulation was found (see Table 1). Post-hoc comparison with Bonferroni correction suggested that, when the presentation tempo is medium, accelerated movements are perceived more natural ( $M = 4.57, SD = 1.39$ ) than original ( $M = 4.40, SD = 1.39$ ) and decelerated movements ( $M = 4.41, SD = 1.42$ ),  $p < .001$ . The effect size of the model, as measured by Cohen's  $f^2$ , was  $f^2 = 0.04$ , indicating a small effect. Please note that the comparison was not possible with the other tempo conditions as neither have all levels of manipulation.

**Table 1**

Summary of the Significant Results From the Multilinear Model Based on Equation 1 - Dependent Variable Perceived Naturalness

Variable	B	SE B	t	p
$\beta_{\text{tempo} \times \text{modality}}$	0.37	0.09	4.03	< .001***
$\beta_{\text{tempo} \times \text{manipulation}}$	-0.14	0.03	-5.24	< .001***

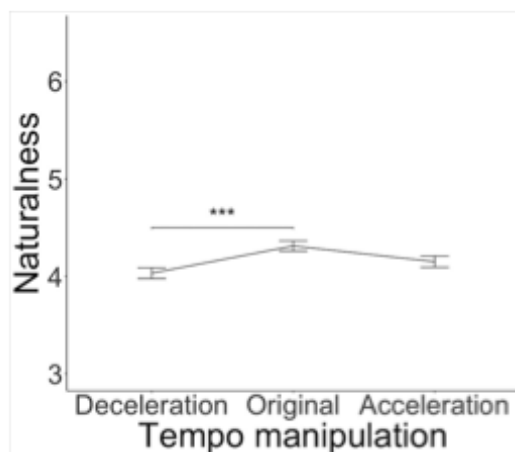
Note. For tempo manipulation, -1 to 0 represents deceleration, 0 stands for no manipulation, and 0-1 represents acceleration. For modality, 1 = Visual-only, 2 = Audiovisual. Values and definitions of the abbreviations are consistent with those in Table 1. For the full table, please refer to A2 in the Appendix.

\*\*\*p < .001.

A one-way ANOVA was run to examine the overall effect of manipulation on the perceived naturalness across all presentation tempi,  $F(2, 2229) = 6.51, p = .002$ , with a small effect ( $\eta_p^2 = 0.006$ ). Significant main effects were found, suggesting that decelerated movements were perceived as significantly less natural than tempo-original ones (see Figure 4). No significant difference between accelerated and original movements was found.

**Figure 4**

Line Plots of Tempo Manipulation by Perceived Naturalness



Note. Mean naturalness with deceleration, original, and acceleration across all tempo levels (original scale from 1 to 7). The whiskers represent the standard errors.

\*\*\*p < .001.

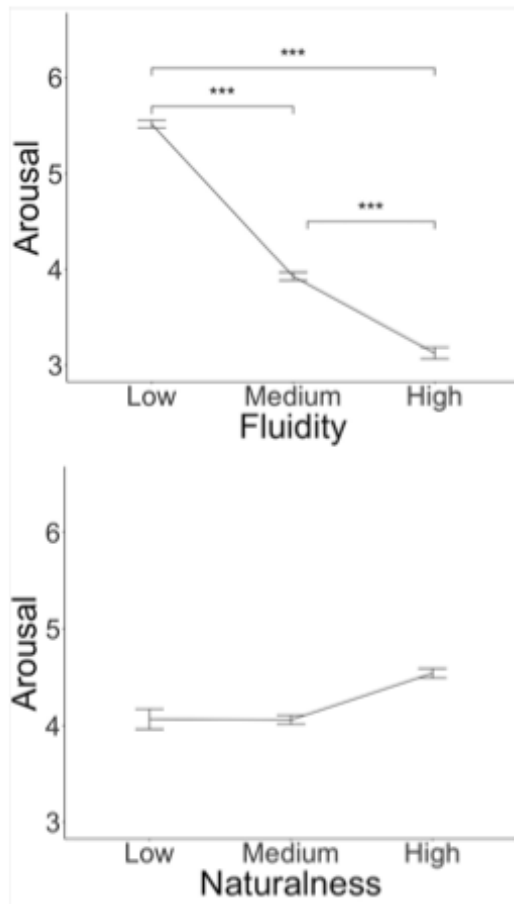
### The Effects of Movement Features on Emotional Arousal

A mixed linear regression revealed significant main effects of presentation tempo, modality, and perceived naturalness on emotional arousal. The effect size of the model, as measured by Cohen's  $f^2$ , was  $f^2 = 0.71$ , indicating a large effect. A faster tempo was associated with higher arousal. Audiovisual stimuli, including the drumbeats, are perceived higher in arousal than the visual-only ones. In addition, higher naturalness is associated with higher arousal (see Figure 5 lower pane, Table 2). A significant interaction between fluidity and presentation tempo was found. Post-hoc comparison with

Bonferroni correction suggested that when tempo is medium, low fluidity (lower than 50 percentiles) was associated with higher arousal ( $M = 4.47$ ,  $SD = 1.21$ ) compared to high fluidity ( $M = 3.98$ ,  $SD = 1.29$ ),  $p < .001$ . Please note that the comparison was impossible with the other tempo conditions as neither has both fluidity levels.

**Figure 5**

*Line Plots of Movement Fluidity and Perceived Naturalness by Arousal*



Note. The upper pane shows mean arousal by low and high fluidity across all tempo levels (original scale from 1 to 7). The lower pane shows mean arousal by low, medium, and high naturalness. The whiskers represent the standard errors.

\*\*\*  $p < .001$ .

**Table 2**

Summary of the Significant Results From the Multilinear Model Based on Equation II - Dependent Variable Emotional Arousal

Variable	B	SE B	t	p
$\alpha$	-2.91	0.99	-2.94	.003**
$\beta_{\text{Natural}}$	0.10	0.02	0.10	< .001***
$\beta_{\text{Modality}}$	0.28	0.05	0.28	< .001***
$\beta_{\text{Tempo}}$	3.60	0.81	3.60	< .001***
$\beta_{\text{fluid*tempo}}$	-1.74	0.58	-3.03	.002**

Note. Tempo = Presentation tempo. Fluid = Fluidity of body movements. Natural = Perceived naturalness. Modality = Sensory modality: 1 = visual-only; 2 = audiovisual. For the full table, please refer to Table A3 in the Appendix.

\*\* $p < .01$ . \*\*\* $p < .001$ .

A one-way ANOVA was run to examine the overall effect of fluidity on emotional arousal across all tempo conditions. Fluidity was split into three levels: Low (lower than 33 quantiles), medium (33 to 67 quantiles), and high (higher than 67 quantiles) fluidity, and resulted in a significant effect on arousal,  $F(2, 2229) = 445.54$ ,  $p < .001$  (see Figure 5, upper pane). The effect size ( $\eta_p^2 = 0.29$ ) indicates a large effect. The significant main effect suggests that high fluidity was perceived to be significantly lower in arousal for all tempo conditions than medium and low fluidity.

### The Effects of Movement Features on Emotional Valence

The mixed linear regression revealed significant main effects of modality and perceived naturalness (Table 3), while movement features, tempo or manipulation showed no significant effects. Increases in naturalness are linked with higher valence (Figure 6). The audiovisual stimuli were perceived to be less positive compared to the visual-only stimuli. The effect size of the model, as measured by Cohen's  $f^2$ , was  $f^2 = 0.27$ , indicating a medium effect.

**Table 3**

Summary of the Significant Results From the Multilinear Model Based on Equation III - Dependent Variable Emotional Valence

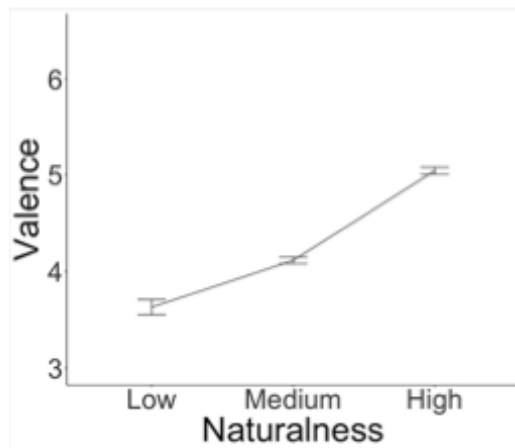
Variable	B	SE B	t	p
$\beta_{\text{Natural}}$	0.37	0.02	23.27	< .001***
$\beta_{\text{Modality}}$	-0.09	0.04	-2.16	.03*

Note. Values and definitions of the abbreviations are consistent with that in Table 1. For the full table, please refer to Table A4 in the Appendix.

\* $p < .05$ . \*\*\* $p < .001$ .

Figure 6

Line Plots of Perceived Naturalness by Emotional Valence



Note. Mean valence by low (lower than 33% percentiles), medium (33% to 67% percentiles), and high (higher than 67% percentiles) naturalness across all tempo levels (original scale from 1 to 7).

## Discussion

The current study investigated the effects of tempo manipulation on perceived emotional arousal and valence. Variables investigated were movement fluidity and complexity, the tempo-manipulated and original movements' perceived naturalness, the presentation tempo, and the sensory modality. Results suggest that: 1) Arousal was influenced by movement fluidity and naturalness. The higher the movement fluidity and the lower the naturalness, the lower the arousal. 2) Naturalness was affected by tempo manipulation. At a medium presentation tempo, accelerated movements were rated more natural than decelerated or even original ones. Overall, decelerated movements were perceived to be the least natural. 3) Emotional valence was most strongly affected by naturalness rather than tempo or kinematic features, such that more natural movements were also rated more positive in valence. Taken together, changing the speed of movements in audiovisual presentations has manifold consequences for the perception of movement qualities and emotions.

The manipulation of presentation tempo, as is often undertaken in video clips and films for various emotional purposes (Wöllner et al., 2018), affected the fluidity of movements. Higher movement fluidity, in turn, was associated with lower emotional arousal, confirming the results of previous studies that found, for example, tender-related motions to be less jerky than anger-elicited ones (Burger et al., 2013). According to Dahl and Friberg (2007), anger was related to low movement smoothness and happiness with high smoothness—the latter was similar to anger in its high emotional arousal but is characterized by higher emotional valence. Similar results were found for dance (Camurri et al., 2003) and music-induced movements (Boone & Cunningham, 2001), such that happiness and anger which are both relatively high in the arousal dimension, induced more frequent tempo changes and higher jerkiness in movements than sadness. Dance movements with high fluidity were also correlated with tenderness and sadness (Burger & Toiviainen, 2020b).



In our study, given that tempo acceleration was strongly correlated with lower fluidity (see Figure 5), participants distinguished between the tempo manipulations in terms of fluidity differences in the tempo-original and tempo-manipulated stimuli. They also perceived the emotional arousal of the two types of movements accordingly. Fluidity might thus be a more salient indicator of tempo manipulation than complexity. It should be stated that movement complexity is inherently related to tempo for stimuli of the same movement types: The faster the tempo, the more movement samples are displayed in a fixed window of time, leading to an increase in complexity. Thus, a potential effect of complexity on arousal can also be partially attributed to changes in tempo. Since the movements in the current study were highly controlled PLD displays, it should be further investigated if such effects can be replicated with naturalistic stimuli such as movie scenes or sports video clips. On the other hand, fluidity affects arousal but not valence, which could be caused by the different salience of the two emotional dimensions in perceived movements. Movements with positive or negative associations may thus exhibit similar fluidity, while valence perception in movements could be more ambiguous (Gross et al., 2010; Pollick et al., 2001).

On the other hand, valence perception was significantly affected by naturalness: The more natural a stimulus is perceived, the more positive the valence (see Figure 6), and the tempo-original and accelerated movements were rated to be more natural than tempo-decelerated ones. As the perceived validity of the stimuli can define the naturalness (Knopp et al., 2019), the effects may be due to an incongruence between the participants' expectations of a given movement and how the PLD movements appeared. The smaller the gap, the more pleasant it was perceived. Movements perceived more positively in valence may also possess a unique combination of features, also known as movement "fingerprints" (Van Vugt et al., 2013), that led to an elevated quality. Van Vugt and colleagues (2013) investigated the movement of "fingerprints" or individuality via professional pianists' movements in unexpressive and muted performances. Their study found that the timing of the pianists' movements is different for each individual. Participants may expect similar movement features with real-life motions that manipulated motions may not have. In this way, complex motions that are less smooth in their trajectories could be most plausible. Similarly, a study found a congruence effect such that audiovisual stimuli consistent in arousal and valence level induced significant psychophysiological responses compared to inconsistent ones (Christensen et al., 2014). However, the link between naturalness and valence should be explored with a more extensive variety of movements.

Not surprisingly, we found that faster presentation tempo was linked to higher arousal. The effect of tempo on emotional arousal has been shown in various studies showing that the faster the sequence, the higher the arousal (Droit-Volet et al., 2013; Sievers et al., 2013; Wöllner et al., 2018), both in music and human movements. Furthermore, the absence of a tempo effect on emotional valence is also consistent with previous findings. Hence, tempo did not predict the emotional valence but rather the arousal level (Khalifa et al., 2008). In addition, our results suggest that emotional arousal was significantly higher in audiovisual than in visual-only presentations, whereas emotional valence was more negative in audiovisual than in visual-only presentations. Audiovisual stimuli evoke stronger emotional arousal than unimodal stimuli (e.g., Vuoskoski et al., 2016; Wöllner et al., 2018). Our finding implies that the multisensory effect persists despite tempo manipulations with the visual inputs, thus shedding light on, for example, the usage of multimedia and slow-motion videos in real-world scenarios.

A limitation of the study lies in the generalizability of the current findings, which could be enhanced with a larger number of PLD movements than that of the current sample. In future research, movements could include more scenarios that cover, for instance, interpersonal interactions and a variety of emotions. Furthermore, a higher number of smaller tempo manipulation steps may allow more detailed comparisons between original and tempo-manipulated conditions. In our experiment, the movements are generated specifically for the experiment in order to match various performance

and presentation tempo conditions in a discrete and recognizable way and could thus have been unusual in comparison with day-to-day activities. In Burger et al.'s (2013) and Burger and Toiviainen's (2020b) studies, the PLD movements are extracted from humans moving or dancing naturally to music.

## Conclusions

In this study, we investigated the impact of tempo manipulation on emotional arousal and valence. Our findings reveal that increased movement fluidity led to decreased perceived emotional arousal, and decreases in perceived naturalness with tempo-decelerated movements resulted in negative emotional valence. Thus, tempo manipulations influence both emotional dimensions, particularly when slowing down from the original tempo. The ratings from our study corroborated earlier research examining the emotional and peripheral physiological responses to slow-motion scenes from movies, sports, and dance (Hammerschmidt & Wöllner, 2018; Wöllner et al., 2018).

The findings also point to the possible mechanism of how tempo manipulations are experienced through the perception of movement features such as fluidity. Identifying the gap between manipulated vs. tempo-original movements may provide insights into how artificial movements could be produced with high plausibility (Chen et al., 2023), creating advertisements with high perceived arousal and positive valence (Yin et al., 2021), or simply understanding the emotional responses that movements could elicit in various scenarios. Our results shed light on the possibilities of shifting the emotional experiences of viewers with tempo-manipulated body movements through music/dance performances or day-to-day media consumption. Future research may investigate the extent to which fluidity affects the perceived emotions, which other movement features apart from tempo affect the perceived emotions, and whether the movement features play the same role in a set of naturalistic and manipulated scenes, for which perceived naturalness should be a key factor.

## Funding

This research was supported by a Consolidator Grant from the European Research Council (Grant No. 725319) to the third author. The research is part of the 5-year project: "Slow motion: Transformations of musical time in perception and performance" (SloMo).

## Acknowledgments

The authors have no additional (i.e., non-financial) support to report.

## Competing Interests

The authors have declared that no competing interests exist.

## Ethics Statement

The present study was in accordance with ethical principles and standards according to the guidelines of the Ethics Committee at the Faculty of Humanities, University of Hamburg. It was approved by the Ethics Committee at the Faculty of Humanities, University of Hamburg.

## Data Availability

The research data, including all perceptual ratings and the movement features for this article, are available on Zenodo (see Wang, 2024).

## Supplementary Materials

For this article, the following supplementary Materials are available:

- Research data, including all perceptual ratings and the movement features (see Wang, 2024)
- Video materials (see Wang et al., 2021)

### Index of Supplementary Materials

Wang, X. (2024). *The dataset of body movement and emotion: Investigating the impact of audiovisual tempo manipulations on emotional arousal and valence* [Data]. Zenodo. <https://doi.org/10.5281/zenodo.12601004>

Wang, X., Burger, B., & Wöllner, C. (2021). *Biological motion and emotion: Investigating the impact of tempo manipulations* [Video materials]. Zenodo. <https://doi.org/10.5281/zenodo.5212232>

## References

- Allingham, E., Hammerschmidt, D., & Wöllner, C. (2021). Time perception in human movement: Effects of speed and agency on duration estimation. *Quarterly Journal of Experimental Psychology*, *74*(3), 559–572. <https://doi.org/10.1177/1747021820979518>
- Atkinson, A. P., Dittrich, W., Germmel, A., & Young, A. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746. <https://doi.org/10.1068/p5096>
- Boone, R. T., & Cunningham, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behavior*, *25*(1), 21–41. <https://doi.org/10.1023/A:1006733123708>
- Brownlow, B. (2017). The effect of music tempo on the psychophysiological measures of stress. *Continuum: The Spelman Undergraduate Research Journal*, *3*(1), 8–16.
- Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception*, *30*(5), 517–533. <https://doi.org/10.1525/mp.2013.30.5.517>
- Burger, B., & Toiviainen, P. (2013). MoCap Toolbox-A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the sound and music computing conference 2013* (pp. 172–178). Logos Verlag Berlin. <http://urn.fi/URN:NBN:fjy-201401211091>
- Burger, B., & Toiviainen, P. (2020a). Embodiment in electronic dance music: Effects of musical content and structure on body movement. *Musicae Scientiae*, *24*(2), 186–205. <https://doi.org/10.1177/1029864918792594>
- Burger, B., & Toiviainen, P. (2020b). See how it feels to move: Relationships between movement characteristics and perception of emotions in dance. *Human Technology*, *16*(3), 233–256. <https://doi.org/10.17011/ht/urn.202011256764>
- Burger, B., & Wöllner, C. (2023). Drumming action and perception: How the movements of a professional drummer influence experiences of tempo, time, and expressivity. *Music & Science*, *6*, 1–17. <https://doi.org/10.1177/20592043231186870>
- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208. [https://doi.org/10.1016/S0042-6989\(01\)00002-5](https://doi.org/10.1016/S0042-6989(01)00002-5)

- Camurri, A., Lagerlöf, L., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2), 213-225. [https://doi.org/10.1016/S1071-5819\(03\)00050-8](https://doi.org/10.1016/S1071-5819(03)00050-8)
- Castillo, G., & Neff, M. (2019). What do we express without knowing? Emotion in gesture. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 702-710). International Foundation for Autonomous Agents and Multiagent Systems. <https://dl.acm.org/doi/abs/10.5555/3306127.3331759>
- Castro, V. L., & Boone, R. T. (2015). Sensitivity to spatiotemporal percepts predicts the perception of emotion. *Journal of Nonverbal Behavior*, 39(3), 215-240. <https://doi.org/10.1007/s10919-015-0208-6>
- Chen, Y.-C., Pollick, F., & Lu, H. (2023). Aesthetic preferences for prototypical movements in human actions. *Cognitive Research: Principles and Implications*, 8(1), Article 55. <https://doi.org/10.1186/s41235-023-00510-0>
- Christensen, J. F., Gaigg, S. B., Gomila, A., Oke, P., & Calvo-Merino, B. (2014). Enhancing emotional experiences to dance through music: The role of valence and arousal in the cross-modal bias. *Frontiers in Human Neuroscience*, 8, Article 757. <https://doi.org/10.3389/fnhum.2014.00757>
- Clarke, T. J., Bradshaw, M. F., Fieldó, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception*, 34(10), 1171-1180. <https://doi.org/10.1068/p5203>
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, 1242, 126-135. <https://doi.org/10.1016/j.brainres.2008.04.023>
- Compton, R. J., Banich, M. T., Mohanty, A., Milham, M. P., Herrington, J., Miller, G. A., Scalf, P. E., Webb, A., & Heller, W. (2003). Paying attention to emotion: An fMRI investigation of cognitive and emotional Stroop tasks. *Cognitive, Affective & Behavioral Neuroscience*, 3(2), 81-96. <https://doi.org/10.3758/CABN.3.2.81>
- Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28(2), 117-139. <https://doi.org/10.1023/B:JONB.0000023655.25550.be>
- Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24(5), 433-454. <https://doi.org/10.1525/mp.2007.24.5.433>
- de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13(4), 247-268. <https://doi.org/10.1007/BF00990296>
- Droit-Volet, S., & Berthon, M. (2017). Emotion and implicit timing: The arousal effect. *Frontiers in Psychology*, 8, Article 176. <https://doi.org/10.3389/fpsyg.2017.00176>
- Droit-Volet, S., Ramos, D., Bueno, J. L. O., & Bigand, E. (2013). Music, emotion, and time perception: The influence of subjective emotional valence and arousal? *Frontiers in Psychology*, 4, Article 417. <https://doi.org/10.3389/fpsyg.2013.00417>
- Dutta, D. (2022, July 5). *The Matrix's original bullet-time method was a little too risky to work*. Slash Film. <https://www.slashfilm.com/917171/the-matrixs-original-bullet-time-method-was-a-little-too-risky-to-work/>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>

- Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*(4), 967–984. <https://doi.org/10.1037/0022-3514.74.4.967>
- Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2010). Methodology for assessing bodily expression of emotion. *Journal of Nonverbal Behavior, 34*(4), 223–248. <https://doi.org/10.1007/s10919-010-0094-x>
- Hammerschmidt, D., & Wöllner, C. (2018). The impact of music and stretched time on pupillary responses and eye movements in slow-motion film scenes. *Journal of Eye Movement Research, 11*(2), 1–17. <https://doi.org/10.16910/jemr.11.2.10>
- Jessen, S., Obleser, J., & Kotz, S. A. (2012). How bodies and voices interact in early emotion perception. *PLoS One, 7*(4), Article e36070. <https://doi.org/10.1371/journal.pone.0036070>
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211. <https://doi.org/10.3758/BF03212378>
- Kang, G. E., & Gross, M. M. (2016). The effect of emotion on movement smoothness during gait in healthy young adults. *Journal of Biomechanics, 49*(16), 4022–4027. <https://doi.org/10.1016/j.jbiomech.2016.10.044>
- Khalifa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology, 68*(1), 17–26. <https://doi.org/10.1016/j.ijpsycho.2007.12.001>
- Knopp, B., Velychko, D., Dreibrodt, J., Endres, D., Knopp, B., Velychko, D., Dreibrodt, J., & Endres, D. (2019). Predicting perceived naturalness of human animations based on generative movement primitive models. *ACM Transactions on Applied Perception, 16*(3), 1–18. <https://doi.org/10.1145/3355401>
- Leiner, D. J. (2019). *SoSci Survey* (Version 3.2.00) [Computer software]. <https://www.sosicisurvey.de>
- Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition, 22*(5), 530–554. <https://doi.org/10.1521/soco.22.5.530.50767>
- Maus, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209–237. <https://doi.org/10.1080/02699930802204677>
- Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior, 23*(2), 133–152. <https://doi.org/10.1023/A:1021435526134>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One, 9*(2), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nebens, R. (2021, April 10). *Zack Snyder's Justice League: Cinematographer responds to all that slow motion*. The Direct. <https://thedirect.com/article/justice-league-slow-motion-zack-snyder>
- Nilsson, N. C., Serafin, S., & Nordahl, R. (2015). The effect of visual display properties and gain presentation mode on the perceived naturalness of virtual walking speeds. *2015 IEEE Virtual Reality Conference, VR 2015 - Proceedings*, 81–88. <https://doi.org/10.1109/VR.2015.7223328>
- Pan, H., Van Beek, P., & Sezan, M. I. (2001). Detection of slow-motion replay segments in sports video for highlights generation. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 3*, 1649–1652. <https://doi.org/10.1109/ICASSP.2001.941253>

- Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, 30(1), 71–83. <https://doi.org/10.1525/mp.2012.30.1.71>
- Pollick, F. E., & Paterson, H. (2008). Movement style, movement features, and the recognition of affect from human movement. In F. Thomas, Shipley, M. Jeffrey (Eds.), *Understanding events: From perception to action* (pp. 286–308). Oxford Scholarship Online. <https://doi.org/10.1093/acprof:oso/9780195188370.001.0001>
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2), B51–B61. [https://doi.org/10.1016/S0010-0277\(01\)00147-0](https://doi.org/10.1016/S0010-0277(01)00147-0)
- Quinn, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, 35(2), 267–280. <https://doi.org/10.1068/p5353>
- Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6), Article 15. <https://doi.org/10.1167/9.6.15>
- Rose, D., Müllensiefen, D., Lovatt, P., & Orgs, G. (2020). *The Goldsmiths Dance Sophistication Index (Gold-DSI): A new psychometric tool to assess individual differences in dance experience*. <https://doi.org/10.31234/osf.io/waplx>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), 70–75. <https://doi.org/10.1073/pnas.1209023110>
- Van Dyck, E., Maes, P. J., Hargreaves, J., Lesaffre, M., & Leman, M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior*, 37(3), 175–190. <https://doi.org/10.1007/s10919-013-0153-1>
- Vuoskoski, J. K., Gatti, E., Spence, C., & Clarke, E. F. (2016). Do visual cues intensify the emotional responses evoked by musical performance? A psychophysiological investigation. *Psychomusicology: Music, Mind, and Brain*, 26(2), 179–188. <https://doi.org/10.1037/pmu0000142>
- Van Vugt, F. T., Jabusch, H. C., & Altenmüller, E. (2013). Individuality that is unheard of: Systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. *Frontiers in Psychology*, 4, Article 134. <https://doi.org/10.3389/fpsyg.2013.00134>
- Wachowski, L., & Wachowski, L. (1999). *The Matrix*. Warner Bros.
- Wöllner, C. (2020). Call and response: Musical and bodily interactions in jazz improvisation duos. *Musicae Scientiae*, 24(1), 44–59. <https://doi.org/10.1177/1029864918772004>
- Wöllner, C., Hammerschmidt, D., & Albrecht, H. (2018). Slow motion in films and video clips: Music influences perceived duration and emotion, autonomic physiological activation and pupillary responses. *PLoS One*, 13(6), Article e0199161. <https://doi.org/10.1371/journal.pone.0199161>
- Yin, Y., Jia, J. S., & Zheng, W. (2021). The effect of slow motion video on consumer inference. *JMR, Journal of Marketing Research*, 58(5), 1007–1024. <https://doi.org/10.1177/00222437211025054>

### Appendix

The Appendix includes the equations on which the models were based (see Analyses in the Methods section), tables of model selection processes, and tables of the full model results. In the following, Equation I refers to the naturalness model, Equation II refers to the arousal model, and Equation III refers to the valence model.

$$\text{Equation I: Perceived naturalness} = \alpha + \beta_{fluid}X_{fluid} + \beta_{complex}X_{complex} + \beta_{tempo}X_{tempo} + \beta_{manipulate}X_{manipulate} + \beta_{fluid*complex}X_{fluid}X_{complex}$$

In Equation I,  $\alpha$  stands for the fixed intercept, while  $\beta_{fluid}$  and  $\beta_{complex}$  represent the betas of the fixed effects of fluidity, body movements, and movement complexity respectively.  $\beta_{tempo}$  and  $\beta_{manipulate}$  represent the betas of the fixed effects of the presentation tempo and tempo manipulation.

$$\begin{aligned} \text{Equation II: Arousal} = & \alpha + \beta_{complex}X_{complex} + \beta_{fluid}X_{fluid} + \beta_{modality}X_{modality} + \beta_{natural}X_{natural} \\ & + \beta_{tempo}X_{tempo} + \beta_{manipulate}X_{manipulate} + \beta_{fluid*complex}X_{fluid}X_{complex} \\ & + \beta_{fluid*manipulate}X_{fluid}X_{manipulate} + \beta_{complex*manipulate}X_{complex}X_{manipulate} + \beta_{fluid*tempo}X_{fluid}X_{tempo} + \beta_{complex*tempo}X_{complex}X_{tempo} \end{aligned}$$

$$\text{Equation III: Valence} = \alpha + \beta_{complex}X_{complex} + \beta_{fluid}X_{fluid} + \beta_{modality}X_{modality} + \beta_{natural}X_{natural} + \beta_{tempo}X_{tempo} + \beta_{manipulate}X_{manipulate} + \beta_{fluid*complex}X_{fluid}X_{complex}$$

In Equations II and III,  $\alpha$  stands for the fixed intercept, while  $\beta_{fluid}$  and  $\beta_{complex}$  represent the betas of the fixed effects of fluidity and body movements and movement complexity respectively.  $\beta_{modality}$  and  $\beta_{natural}$  represent the betas of modality and perceived naturalness.  $\beta_{tempo}$  and  $\beta_{manipulate}$  represent the betas of the fixed effects of the presentation tempo and tempo manipulation.  $\beta_{fluid*complex}$  stands for the beta of the interaction between fluidity and complexity.

**Table A1**

Summary of the Results of Maximum Likelihood Ratio Tests ( $\chi^2$  Analysis)

Dependent Variable	Variable	df	AIC	BIC	Log likelihood	$\chi^2$	p	
Emotional Arousal	Baseline	3	8163.10	8180.23	-4078.55	—	—	
	+ Complexity	4	7087.83	7110.67	-3539.91	1077.27	< .001***	
	+ Fluidity	5	7033.36	7061.91	-3511.68	56.47	< .001***	
	+ Fluidity: Complexity	6	6917.87	6952.13	-3452.94	117.49	< .001***	
	+ Naturalness	7	6896.57	6936.55	-3441.29	23.30	< .001***	
	+ Modality	8	6859.19	6904.87	-3421.59	39.39	< .001***	
	+ Tempo	9	6795.47	6846.86	-3388.73	65.72	< .001***	
	+ Manipulation	10	6792.72	6849.83	-3386.36	4.74	.03*	
	+ Fluidity*Manipulation	11	6793.56	6856.38	-3385.78	1.16	.28	
	+ Complexity*Manipulation	12	6794.33	6862.86	-3385.16	1.24	.27	
	+ Complexity*Tempo	13	6790.49	6864.73	-3382.25	5.84	.02	
	+ Fluidity*Tempo	14	6783.28	6863.23	-3377.64	9.21	< .001***	
	Emotional Valence	Baseline	3	6842.48	6859.61	-3418.24	—	—
		+ Complexity	4	6785.30	6808.14	-3388.65	59.18	< 0.001***
+ Fluidity		5	6787.04	6815.59	-3388.52	0.26	.61	
+ Fluidity: Complexity		6	6778.12	6812.38	-3383.06	10.92	< .001***	
+ Modality		7	6777.25	6817.22	-3381.62	2.87	.09	
+ Tempo		8	6777.37	6823.05	-3380.68	1.88	.17	
+ Manipulation		9	6779.12	6830.51	-3380.56	0.25	.62	
+ Naturalness		10	6295.31	6352.42	-3137.66	485.81	< .001***	
+ Fluidity*Manipulation		11	6295.81	6358.63	-3136.91	1.50	.22	

Dependent Variable	Variable	df	AIC	BIC	Log likelihood	$\chi^2$	p
	+ Complexity*Manipulation	12	6297.33	6365.86	-3136.67	0.48	.49
	+ Fluidity*Tempo	13	6298.75	6372.98	-3136.37	0.59	.44
	+ Complexity*Tempo	14	6297.15	6377.10	-3134.58	3.59	.06
Perceived Naturalness	Baseline	3	7698.41	7715.54	-3846.20	—	—
	+ Complexity	4	7693.70	7716.54	-3842.85	6.71	.01*
	+ Fluidity	5	7693.98	7722.53	-3841.99	1.72	.19
	+ Tempo	6	7639.31	7673.57	-3813.65	56.67	< .001***
	+ Manipulation	7	7629.54	7669.52	-3807.77	11.76	< .001***
	+ Fluidity*Complexity	8	7630.15	7675.83	-3807.07	1.40	.24
	+ Tempo*Manipulation	9	7604.81	7656.21	-3793.40	27.34	< .001***

Note. The analyses explored the differences in perceptual changes as a function of the fluidity of body movements, the movement complexity, and the presentation tempo.

\* $p < .05$ . \*\*\* $p < .001$ .

**Table A2**

Summary of Results From the Multilinear Model Based on Equation I - Dependent Variable Perceived Naturalness

Variable	B	SE B	t	p
$\alpha$	0.92	2.34	0.39	.70
$\beta_{fluid}$	2.25	1.26	1.79	.07
$\beta_{complex}$	0.53	0.59	0.90	.37
$\beta_{tempo}$	0.28	0.29	0.97	.33
$\beta_{manipulation}$	2.62	0.65	4.03	< .001***
$\beta_{fluid*complex}$	-0.47	0.49	-0.95	.34
$\beta_{tempo*manipulation}$	-0.99	0.19	-5.24	< .001***

Note. Fluid = Fluidity of body movements; Complex = Movement complexity; Tempo = Presentation tempo; Manipulation = The extent of manipulation. For tempo manipulation, -1 to 0 represents deceleration, 0 stands for no manipulation, and 0 - 1 represents acceleration.

\*\*\* $p < .001$ .

**Table A3**

Summary of Results From The Multilinear Model Based on Equation II - Dependent Variable Emotional Arousal

Variable	B	SE B	t	p
$\alpha$	-12.52	11.82	-1.06	.29
$\beta_{fluid}$	7.65	4.90	1.56	.12
$\beta_{complex}$	2.12	9.65	0.22	.83
$\beta_{natural}$	0.10	0.02	5.86	< .001***
$\beta_{fluidity}$	0.28	0.05	6.03	< .001***
$\beta_{tempo}$	3.60	0.81	4.44	< .001***
$\beta_{manipulation}$	-2.06	1.09	-1.89	.06
$\beta_{fluid*complex}$	-1.16	5.80	-0.20	.84
$\beta_{fluid*manipulation}$	2.39	3.03	0.79	.43
$\beta_{complex*manipulation}$	-1.47	3.81	-0.38	.70



Variable	B	SE B	t	p
$\beta_{\text{beat}^* \text{tempo}}$	-1.74	0.58	-3.03	.002**
$\beta_{\text{complex}^* \text{tempo}}$	0.82	3.57	0.23	.82

Note. For modality, 1 = Visual-only, 2 = Audiovisual. Natural = Perceived naturalness, on a scale from 1 to 7. 1 stands for *low* and 7 for *high* naturalness. Values and definitions of the abbreviations are consistent with Table A2.

\*\*p < .01. \*\*\*p < .001.

**Table A4**

Summary of Results From the Multilinear Model Based on Equation III - Dependent Variable Emotional Valence

Variable	B	SE B	t	p
$\alpha$	3.61	0.88	4.09	< .001***
$\beta_{\text{beat}}$	-0.31	0.34	-0.92	.36
$\beta_{\text{complex}}$	-0.10	0.37	-0.26	.79
$\beta_{\text{natural}}$	0.37	0.02	23.27	< .001***
$\beta_{\text{modality}}$	-0.09	0.04	-2.16	.03*
$\beta_{\text{tempo}}$	0.00	0.17	-0.02	.98
$\beta_{\text{manipulation}}$	0.15	0.15	0.99	.32
$\beta_{\text{beat}^* \text{complex}}$	0.02	0.34	0.06	.95

Note. Values and definitions of the abbreviations are consistent with Table A2 and A3.

\*\*\*p < .001.

## Study 4

### **Tapping to drumbeats in an online experiment changes our perception of time and expressiveness**

Xinyue Wang, Birgitta Burger, and Clemens Wöllner

*Psychological Research. 88(1), 127 - 140*

<https://doi.org/10.1007/s00426-023-01835-7>

Reproduced with kind permission from Springer Nature



## Tapping to drumbeats in an online experiment changes our perception of time and expressiveness

Xinyue Wang<sup>1</sup> · Birgitta Burger<sup>1</sup> · Clemens Wöllner<sup>1,2</sup>

Received: 22 July 2022 / Accepted: 15 May 2023 / Published online: 10 June 2023  
© The Author(s) 2023

### Abstract

Bodily movements along with music, such as tapping, are not only very frequent, but may also have a profound impact on our perception of time and emotions. The current study adopted an online tapping paradigm to investigate participants' time experiences and expressiveness judgements when they tapped and did not tap to a series of drumming performances that varied in tempo and rhythmic complexity. Participants were asked to judge durations, passage of time (PoT), and the expressiveness of the performances in two conditions: (1) Observing only, (2) Observing and tapping regularly to the perceived beats. Results show that tapping trials passed subjectively faster and were partially (in slow- and medium-tempo conditions) perceived shorter compared to the observing-only trials. Increases in musical tempo (in tapping trials) and in complexity led to faster PoT, potentially due to distracted attentional resources for the timing task. Participants' musical training modulated the effects of complexity on the judgments of expressiveness. In addition, increases in tapping speed led to duration overestimation among the less musically trained participants. Taken together, tapping to music may have altered the internal clock speed, affecting the temporal units accumulated in the pacemaker-counter model.

### Introduction

Music offers a unique temporal context that entails varying tempi and complexities, in which we experience time differently. Jonathan Berger (Berger, 2014) has once pointed out that the composition of Franz Schubert's *String Quintet in C major, D.956* created many temporal illusions—by embedding faster, more complex rhythms in a slow, near motionless musical context, or slow and simple rhythms in a fast, temporally complicated context, Schubert successfully distorted the perceived durations of excerpts in comparison to the clock time.

Many studies have explored the effects of musical attributes on time perception, including tempo (Droit-Volet et al., 2013) and complexity (Bueno et al., 2002). The influences of proactive responses to music such as tapping (Hammer-schmidt & Wöllner, 2020; Manning & Schutz, 2013) on time perception have also been widely explored. However,

it has not been investigated how individuals perceive time while tapping to music that varies in tempo and complexity. The current study compares perceived time in tapping and no-tapping conditions with a drummer's performances. We also aim to investigate the perceived expressiveness of the performances in relation to the musical attributes and tapping, to reveal the role it serves in the timing experiences.

### Temporal attributes of music

To investigate the effects music may have on time perception, an understanding of the temporal attributes of music should be established first. The beat provides basic structures to music. It entails isochronous pulses that are subjectively perceived within the individual (Large, 2000), whereas tempo typically refers to the number of perceived beats in a certain period, defined as Beats Per Minute (BPM), yet in fact tempo perception is complex and involves further musical characteristics (London, 2011). Rhythms or rhythmic structures, on the other hand, represent the temporal patterns in which the musical notes are organized with respect to the underlying beat (Large, 2000). Grouping the musical notes by small or large cycles leads to different metrical levels (Burger et al., 2018). A lower metrical level refers to a shorter note length (e.g. adjustment to the eighth note level),

✉ Xinyue Wang  
philippapw@gmail.com

<sup>1</sup> Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany

<sup>2</sup> University of Music Freiburg, Freiburg, Germany

while a higher metrical level refers a longer note length (e.g. half note) (Hammerschmidt & Wöllner, 2020). Complexity describes the composition of rhythmic structures such that more complex rhythms temporally encompass more patterns (e.g. polyrhythms) and higher event density (Vuust & Witek, 2014). Rhythmic patterns of various tempi and complexities could lead to differences in the experience of time (Bueno et al., 2002).

### Moving to music affects time perception

Moving to music encompasses a variety of activities, for instance tapping (Polak et al., 2018), walking (Styns et al., 2007), or free whole-body movements (Burger et al., 2018). Among them, tapping has been frequently adopted in studies of sensorimotor synchronization and its effects on time perception, especially in combination with music (Drake et al., 2000; Hammerschmidt & Wöllner, 2020; Snyder & Krumhansl, 2001), as it allows participants to find and to react to beats with small amount of physical efforts.

As previous findings revealed, synchronizing with musical beats as an “organic, effortless, and...spontaneous” (Large, 2000, pp. 527) reaction, such as tapping, could affect the human timing performances to a great extent (Hammerschmidt & Wöllner, 2020; McAuley & Kidd, 1998; Wöllner & Hammerschmidt, 2021). Moreover, it was found that tapping to musical beats increased the accuracy with time keeping tasks (Manning & Schutz, 2013). Durations were perceived to be shorter when tapping to music, whereas time was perceived to pass faster when performing a working memory task (Wöllner & Hammerschmidt, 2021). It should be noted that time perception refers specifically to duration estimation (DE) and passage of time (PoT) (Grondin, 2010), while timing refers to not only passively perceiving time, but also proactively producing temporal structures in various tempi (Honing, 2001).

### Timing mechanisms

Theories and findings of human timing mechanisms, also known as the internal clock model, have shed light on how tapping moderated our timing experiences. The dynamic attending theory (DAT) (Jones & Boltz, 1989; Large & Jones, 1999) supports the central timing model based on Treisman’s (Treisman, 1963) theory, in which the key model is represented as a pacemaker-counter mechanism stands. More specifically, the model hypothesizes an internal clock that emits temporal pulses, records the accumulated pulses within the target period, and compares the recorded pulses to the ones of a reference duration before coming to judgments. Based on the pacemaker-counter model, the DAT postulates that the emission of internal temporal pulses can be synchronised with external rhythms, also known as the temporal entrainment effect (McAuley & Jones, 2003). Attending

to faster stimuli leads to increases in the internal temporal pulses, and consequently dilation of the perceived duration and reduced passage of time (PoT). The effect has been validated as robust and widely present across a variety of sensory modalities (Wang & Wöllner, 2019), thus validating tempo as a key predictor of the perceived duration and PoT. When tapping to higher metrical levels, participants attended to larger temporal units and consequently judged perceived durations to be shorter and PoT to be faster (Hammerschmidt & Wöllner, 2020). Thus, by tapping to the music, individuals explicitly synchronize the internal clock speed to the underlying temporal pulses of the external rhythms and experience changes in the perceived time.

### Effects of expressiveness and complexity

The emotional expressiveness may also mediate how individuals perceive the passing of time in relation to the internal clock speed. Fast movements were perceived to be more expressive (Allingham et al., 2021), indicating that fast stimuli positively predicted the perceived expressiveness. This finding with visual movements is in line with auditory evidence where an association between tempo of the music and expressiveness was found (Fernández-Sotos et al., 2016). Considering faster clock speed associated with higher expressiveness, perceiving expressive stimuli could also lead to overestimation of the duration and slower passage of time.

Complexity is another factor that has been found to affect temporal processing. Moderately complex stimuli containing higher event density within a fixed period of time were perceived to be longer than simple and/or highly complex ones (Aubry et al., 2008; Bueno et al., 2002; Hogan, 1975). According to the pacemaker-counter mechanism (Treisman, 1963), more temporal units accumulated in the counter device could lead to duration estimation, as the internal clock speed synchronises with frequent event changes (high number of segmentations) with the complex stimuli (Fraisse, 1978). On the other hand, Mate and colleagues (Mate et al., 2009) proposed that highly complex stimuli required more resources in the working memory, thus could be judged to be longer in comparison with the reference duration stored in the pacemaker-counter device. The effect was also found with musical stimuli, as participants overestimated the durations as they listened to 90-s excerpts of a rhythmically simple versus a complex symphony (Bueno et al., 2002).

### Effects of music training

A key influence on the perception of time is the perceivers’ musical expertise. Musicians are capable of more accurate duration estimation (Panagiotidi & Samartzi, 2012; Rammsayer & Altenmüller, 2006), sensorimotor synchronization (Drake et al., 2000; Repp, 2010), temporal phase

detection (Manning & Schutz, 2016), and higher synchronization flexibility (Scheurich et al., 2018) than non-musicians. 12- to 15-year-old students who received at least two years of musical training estimated durations of musical excerpts more accurately than those who did not (Panagiotidi & Samartzi, 2012). Moreover, modality-specific evidence supports the view that musically trained individuals exhibited more stable and more accurate sensorimotor synchronization with auditory rhythms than the untrained group, due to their extensive training in tasks such as collaborative music making that frequently involve time keeping (Repp, 2010; Repp & Doggett, 2007). Altogether, the findings suggest that musical training equipped individuals with higher accuracy and sensitivity in temporal processing.

### Online tapping paradigms

In past studies, researchers have mainly adopted in-lab setting for the consistency of environment and standardization of procedure. Experiments were run with in-lab tapping devices such as Yamaha piano keyboard (Snyder & Krumhansl, 2001), a BopPad touch pad (Hammerschmidt & Wöllner, 2020), or the space bar of the experiment computer (London et al., 2019). The recent tapping apparatus has shifted to online platforms, for instance, the Rhythm ExPeriment Platform (REPP) (Anglada-Tort et al., 2022) and web-based tapping applications (Hammerschmidt et al., 2021a). In response to the current demands, we aimed to develop an easy and direct way to implement a tapping study using an existing online survey platform SoSci Survey (Leiner, 2019).

### Aims

In the current study, we aimed to investigate whether judgments on Duration Estimation (DE), Passage of Time (PoT), and Expressiveness are affected when tapping to audio-visual stimuli of varying tempi and rhythmic complexities compared to when not tapping. In addition, we explored the effects of musical training on the perceptual ratings, as measured by the Gold-MSI (Müllensiefen et al., 2014). To further inspect possible effects of tapping, tapping speed and stability were also examined. We hypothesized that:

1. Tapping with the performance leads to faster PoT and shorter DE compared to the no-tapping conditions.
2. Due to the higher event density, fast performances are assumed to be perceived as longer, to pass more slowly, and be more expressive than slow ones. Similarly, complex rhythms are expected to be perceived as longer, pass more slowly, and be more expressive than simple rhythms.

3. Higher musical training is expected to lead to more accurate DE and PoT judgments, as past study found higher accuracy with musically trained groups (Nguyen et al., 2022).
4. Fast tapping as well as high tapping stability are predicted to be linked to duration overestimation and slower PoT, as they indicate faster and more stable internal clock speed, while slow tapping and low stability should lead to duration underestimation and faster PoT.

## Method

### Participants

A total of 109 participants were recruited for the online experiment (61.5% were females;  $M_{age} = 26$ ,  $SD_{age} = 7.04$ ). Over half of the participants (56.6%) have completed higher education (Bachelor, Master, and Ph.D. degrees). Participants were of a wide variety of nationalities, mainly Europeans (49.54%) and Africans (41.84%). Based on the summed scores of five items from the Goldsmith Music Sophistication Index (part of the Gold MSI factor 3 “Music Training”, Müllensiefen et al., 2014; see Supporting Information I), the participants’ music training scored a mean of 0.25 ( $SD_{MusicTraining} = 0.05$ ) after being normalized, ranging from 0 (no music training) to 1 (highly trained). Thirty-two participants had not received any formal music training, while 77 had received some type of formal training. Participants’ original musical training score, based on selected items, ranged from 5 (no training) to 34 (highly trained), including the years of training and hours of daily practice, among other variables. The musical training score was normalized for subsequent analyses with its range as described above.

Participants were recruited via the survey platform Prolific (<https://prolific.co/>) and university classes to participate in the online experiment on SoSci Survey (<https://www.sosci-survey.de>) (Leiner, 2019). The study was approved by the Ethics Committee at the Faculty of Humanities, University of Hamburg. All participation were consented. Each participant was either compensated by an hourly rate of €8.85/hour or course credits. On average, it took participants 20 min to complete the study.

### Stimuli

The experiment stimuli consisted of 9 different audio-visual presentations, in which a drummer (male, 27 year-old, classically trained for over 20 years) performed three rhythms in three tempi (60, 110, and 160 BPM) and three levels of complexities (simple, medium, complex, see Fig. 1). The movements were recorded via an eleven-camera motion-capture system (Qualisys Oqus 700) at a framerate of 200



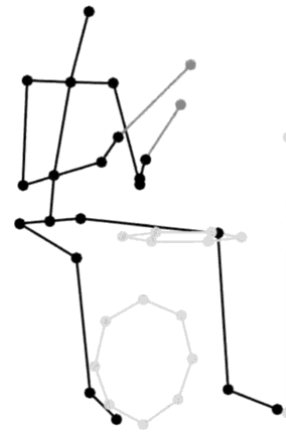
**Fig. 1** Depiction of three rhythmic complexities from simple (top), medium (middle), to complex (down) condition. The lower filled notes indicate the kick, the higher filled notes the snare, and the cross noted the hi-hat

frames per second, depicting the performer, drum sticks, and the drumming instruments (kick, snare, hi-hat) using reflective markers. Animations were created using MoCap toolbox (Burger & Toiviainen, 2013) in MATLAB (black stick figure on a white background, see Fig. 2).

Each experiment stimulus started by showing a fixation cross in the centre of a dark background lasting one second, followed by a drumming performance lasting 15 s. In addition, 4 catch trials of different lengths (2 of 8 s, 2 of 30 s) were included that did not enter any analysis, as their sole purpose was to investigate whether participants attended to the durations of the stimuli.

### Procedure

Participants were required to complete the tasks in a quiet environment on the computer using headphones. They were given the opportunity to adjust their sound level with a sample music excerpt before the start. Participants were then asked to tap the spacebar regularly for one minute upon instructions on the screen to measure their pre-test spontaneous motor tapping (SMT), and then rate how excited (emotional arousal) they felt by moving the cursor on a scale. Following this, the experiment part started containing two blocks, the first one required observing and rating the drumming animations, while the second block required observing



**Fig. 2** Depiction of the biological motion of the drummer's performance as the visual stimuli. The stimuli can be found on Zenodo (<https://doi.org/10.5281/zenodo.7987993>)

and simultaneously tapping to the drumming animations and rating them. The order of the blocks was not randomized because participants should complete the observing-only trials having no knowledge of motor involvement. The order of the stimuli within each block was randomized to avoid any effect of order. To introduce the first (i.e., no-tapping) task, a test trial was presented with participants being asked to watch the animation first and then type their estimation of the stimulus duration. They also typed their estimation of the duration in seconds, and used a slider on a scale from 1 to 101 to indicate how fast they personally felt time had passed and how expressive the performance was. The first block of 11 randomised no-tapping trials (including two catch trials) commenced after the test trial. The participants were free to take a break between the first and second block.

Before the tapping block started, participants were informed that they should tap with the performance in a regular, even, and non-rhythmic manner on the spacebar. Video examples of correct and incorrect tapping, performed by one of the authors of this study, were presented to instruct and guide the participants (material and data are available on Zenodo. For material, it can be found at: <https://doi.org/10.5281/zenodo.7987993>, for data, it can be found at: <https://doi.org/10.5281/zenodo.7988013>). Participants were again presented with a test trial, followed by the second block. During each stimulus presentation, the phrase "Please tap to the performance" was always displayed as a reminder. After the second block, participants were asked to complete a post-test SMT of the same procedure as the pre-test SMT and to rate their emotional arousal level again. They were

also required to fill in a short questionnaire of their music training and demographic information. During the process, participants were instructed not to look at any clocks or to count the time.

**Analyses**

Two streams of analyses were conducted: (1) To examine effects of tapping versus no-tapping in the context of performance and music training differences, perceptual judgements of the tapping and no-tapping trials were predicted by the stimulus characteristics and music training. (2) To investigate the impact of tapping behaviour more specifically, perceptual judgements of the tapping trials were predicted by tapping speed, tapping stability, stimulus characteristics, and music training.

After checking that the SoSci Survey tapping data allowed reliable post-processing, taps between the third and the last one were extracted for analyses purposes. This was done to reduce participant instability, as they needed a few taps to get into the tempo of the trial. Subsequently, outlier detection was conducted before the commencement of analyses. Data of participants ( $N = 12$ ) who failed to tap in less than 10 of the 11 trials in each block were completely removed from the dataset of perceptual ratings, tapping recordings, and demographics, due to failure to understand the nature of the tasks. From a pool of 29,924 taps, we eliminated outliers based on the following criteria: (1) Less than 3 taps in one trial, (2) inter-tap intervals (ITIs, temporal distance between two consecutive taps) longer than the 25% upper threshold of a whole note at the tempo, which might be due to absence of attention and indicates disruption of the task in the trial, (3) Taps of repetitive timestamps, in other words ITIs consecutively ( $N > 1$ ) equal to 0, that might be due to system failures ( $N_{Taps} = 117$ ), (4) ITIs lower than 70 ms, which suggested tapping twice by mistake in a very short period ( $N_{Taps} = 51$ ), (5) ITIs longer than 5 SDs from the mean averaged from all ITIs per participant per condition/trial except for the maximum ITI ( $N_{Taps} = 116$ ). Trials ( $N_{Taps} = 4$ ) that fit the first and second criterion were completely removed, while taps that fit the other criteria were removed from perspective trials keeping the remainder of the trial. Data of participants whose tapping trials after outlier exclusion were less than 10 out of 11 were not eliminated.

For the first stream of analysis (influence of tapping vs. no-tapping), general linear mixed models (LMMs) were adopted to answer whether (1) tapping, (2) tempo, (3) complexity, (4) musical training, also referred to training in the following, and their two-way interactions (fixed effects) affected participants’ DE, PoT, and Expressiveness judgements. Thus, independent variables include tapping (yes/no), tempo (3 levels), complexity (3 levels), and training (based on the Gold-MSI; Müllensiefen et al., 2014). To select the model of the highest goodness of fit for each dependent

variable, maximum likelihood ratio (MLR) tests were conducted (see Table 2). In the MLR tests, predictors were added one after another from the baseline model, in which only the random effects were present. Variance of participants and of conditions were considered as random effects. Models with the lowest Akaike Information Criterion (AIC), while adding significantly to the previous model were considered the final models. Therefore, for this line of models, each LMM might entail a different number of predictors. Post-hoc analyses with Tukey correction were conducted to follow up significant main and interaction effects.

$$\begin{aligned} \text{Perceptual judgments} = & \alpha + \beta_{Tap}X_{Tap} + \beta_{Tempo}X_{Tempo} \\ & + \beta_{Comp}X_{Comp} + \beta_{MT}X_{MT} \\ & + \beta_{Tap:Tempo}X_{Tap}X_{Tempo} + \beta_{Tap:Comp}X_{Tap}X_{Comp} + \beta_{Tap:MT}X_{Tap}X_{MT} \\ & + \beta_{Tempo:Comp}X_{Tempo}X_{Comp} + \beta_{Tempo:MT}X_{Tempo}X_{MT} \\ & + \beta_{Comp:MT}X_{Comp}X_{MT} + (1|Participant) + (1|Condition) \end{aligned}$$

In this equation, perceptual judgments stand for DE, PoT, or Expressiveness,  $\alpha$  stands for the fixed intercept, while  $\beta$  represents the betas of the fixed effects. Comp = Complexity, MT = Music training (normalized scores). It should be noted that components of the equation vary based on MLR test results (see Table 1).

For the second stream of analyses (influence of tapping speed and stability), LMMs were adopted to answer the research question of whether (1) tapping speed, (2)

**Table 1** Summary of the mixed linear model analysis for all trials based on the MLR results

DV	Variable	B	SE B	t	p
Duration estimation	$\beta_{Tempo:Tap}$	-0.07	0.02	-3.18	0.002**
	$\beta_{Comp}$	0.40	0.14	2.81	0.016*
Passage of time	$\beta_{Tap}$	0.33	0.15	2.18	0.049*
	$\beta_{MT}$	-0.24	0.09	-2.68	0.007**
	$\beta_{Tempo:MT}$	0.06	0.02	2.84	0.005**
	$\beta_{Tap:MT}$	0.08	0.06	2.35	0.019*
	$\beta_{Tempo}$	-0.13	0.01	-9.76	<0.001***
Expressiveness	$\beta_{MT}$	-0.25	0.11	-2.34	0.019*
	$\beta_{Tap:MT}$	0.11	0.04	2.89	0.004**
	$\beta_{Comp:MT}$	0.05	0.02	2.09	0.037*

For the full table, please refer to Table S2. Tapping was coded as 1 (no-tapping) or 2 (tapping). Tempo was coded as 3 (slow), 2 (medium), and 1 (fast). Complexity was coded as 1 (simple), 2 (medium), and 3 (complex)

Comp complexity, Tap presence of tapping, MT normalized music training scores

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

tapping stability, (3) tempo, (4) complexity, and (5) participants' music training as well as their two-way interactions affected participants' DE, PoT, and Expressiveness judgments within the tapping trials. This line of models was intended to examine the contributions of tapping speed and stability, as well as their interactions with other factors. In this regard, MLR tests were not adopted to examine the goodness of fit. Variance from participants and conditions were included as random effects. Post-hoc analyses with Tukey correction were conducted to follow up significant main and interaction effects. The analyses were performed in R (Version 3.5.3; R Core Team, 2019) using the package lme4 (Bates et al., 2015).

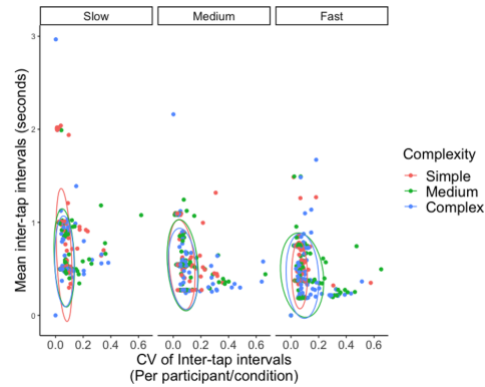
$$\begin{aligned} \text{Perceptual judgments} = & \alpha + \beta_{ITI}X_{ITI} + \beta_{ITlcv}X_{ITlcv} \\ & + \beta_{Tempo}X_{Tempo} + \beta_{Comp}X_{Comp} \\ & + \beta_{MT}X_{MT} \\ & + \beta_{ITI:Tempo}X_{ITI}X_{Tempo} + \beta_{ITI:Comp}X_{ITI}X_{Comp} + \beta_{ITI:MT}X_{ITI}X_{MT} \\ & + \beta_{ITlcv:Tempo}X_{ITlcv}X_{Tempo} + \beta_{ITlcv:Comp}X_{ITlcv}X_{Comp} + \beta_{ITlcv:MT}X_{ITlcv}X_{MT} \\ & + \beta_{ITI:ITlcv}X_{ITI}X_{ITlcv} + (1|Participant) + (1|Condition) \end{aligned}$$

In this equation, perceptual judgments stand for DE, PoT, or expressiveness,  $\alpha$  stands for the fixed intercept, while  $\beta$  represent the betas of the fixed effects. MT = Music training (normalized scores), Comp = Complexity, ITI = mean inter-tap intervals (tapping speed), ITlcv = Coefficient of variation for inter-tap intervals (tapping stability).

For this stream of models, only the ratings for the tapping trials were included as well as participants' tapping behaviour. Tapping data were transformed into two variables: (1) Tapping speed: Average ITI per participant per condition, the higher the it is, the slower the tapping tempo. (2) Tapping stability: Coefficient of variation (CV) of it is per participant per condition, which is the ratio between standard deviation and mean (standard deviation (sd)/mean). The higher the CV of ITI, the more unstable the taps. The descriptive scatter plots (see Fig. 3) show that participants' tapping behaviours were mainly isochronous, consistent, and clustered around different tempo and complexity conditions.

For both lines of analyses, dependent variables include DE, PoT, and Expressiveness.

- DE: Participants' estimation of the time passed in seconds.
- PoT: The ratings from a 1 (slowest) to 101 (fastest) scale was normalized across the full range of the scale  $(x - \min(x))/(\max(x) - \min(x))$ . Minimum was 1, while maximum was 101.



**Fig. 3** Scatter plots of the mean I (y-axis) and CVs of ITI (x-axis) clustered by rhythmic complexity, faceted by tempo

- Expressiveness: The ratings from a 1 (not at all) to 101 (very much) scale was normalized across the full range of the scale  $(x - \min(x))/(\max(x) - \min(x))$ . Minimum was 1, while maximum was 101.

## Results

LMMs were conducted to examine participants' judgements in DE, PoT, and Expressiveness in tapping versus no-tapping trials. The effects of training, tempo, and complexity as well as their interactions with tapping were included in the models.

### Influence of tapping versus no-tapping

#### Duration estimation

No significant main effects of tempo, rhythmic complexity, tapping, and training were found, while there was a significant interaction effect between tempo and tapping (see Table 1, for the full table, please refer to Table S2). Post-hoc comparison with Tukey correction suggested that, in this interaction, the effect of tapping was significant for the slow and medium tempi: No-tapping trials were perceived to last longer than the tapping trials (see Table 2, Fig. 4, for the full table, please refer to Table S3). No significant differences were found in the fast condition.

#### Passage of time

Significant main effects of complexity, tapping, and training were found, whereas no significant effect of tempo was

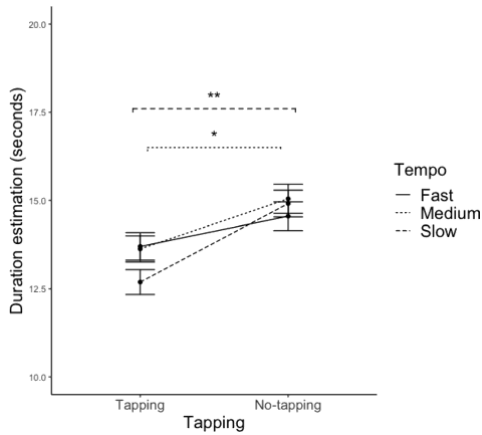


**Table 2** Summary of the post-hoc comparison with Tukey correction based on the significant interactions from Table 1

DV	IV1	IV2	Differences of estimate	SE of difference	t	p
Duration estimation	Slow	T-NT	2.22	0.30	7.29	0.011*
	Medium	T-NT	1.42	0.30	4.65	0.053*
Expressiveness	High MT	T-NT	- 0.04	0.02	- 2.84	0.043*
	High MT	Simple-complex	- 0.02	0.02	- 3.57	0.018*

For the full table, please refer to Table S3. Coding of the variables is consistent with the caption of Table 1  
*T* tapping, *NT* no-tapping, *MT* music training

\*\*\**p* < .001, \*\**p* < .01, \**p* < .05



**Fig. 4** Line plot of the interaction effect between tempo and tapping on Duration Estimation. The whiskers represent standard errors. \*\**p* < 0.01, \**p* < 0.05. Significance indicators are of the same line type as the corresponding groups (i.e. tempo)

present. More complex rhythms are related to faster PoT. However, post-hoc comparisons with Tukey corrections suggested no significant differences among the three levels of complexity (see Table 3). The model revealed significant interaction effects between training and tempo as well as between training and tapping (see Table 1, for the full table, please refer to Table S2). Post-hoc comparisons with Tukey correction suggested that the effect of tempo did not differ significantly between high and low musical training. Similarly, the effect of tapping did not differ significantly by levels of musical training (see Table 2).

**Expressiveness**

Significant main effects of tempo and training were found, whereas there was no significant effect of complexity and tapping (see Table 1, for the full table, please refer to Table S2). Faster tempo is related to higher perceived Expressiveness, as post-hoc analyses with Tukey corrections revealed significant differences between slow and fast, slow and medium, as well as medium and fast conditions. Expressiveness ratings were the highest with fast tempo, followed by medium, and lowest

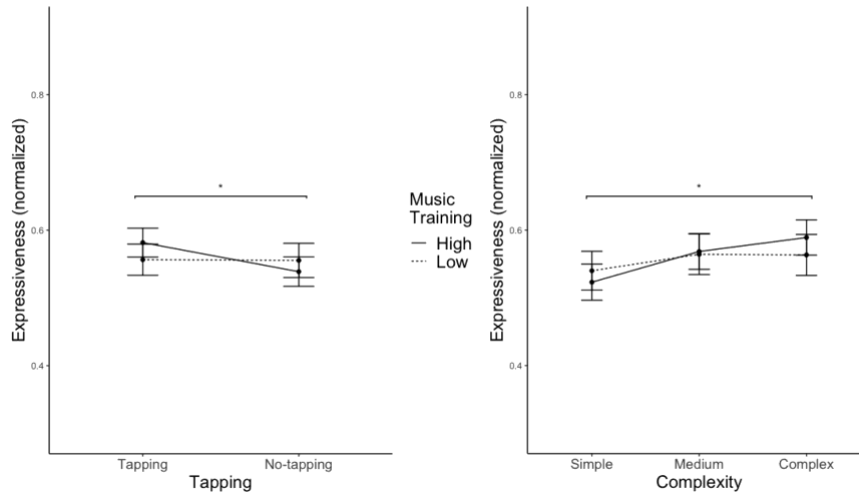
**Table 3** Summary of the post-hoc comparison with Tukey correction based on significant main effects from Table 1

DV	Condition1	Condition2	Differences of estimate	SE of difference	t	p
Passage of time	Simple	Complex	- 0.13	0.07	- 1.72	0.276
	Simple	Medium	- 0.13	0.07	- 1.72	0.268
	Medium	Complex	0.001	0.07	0.01	0.999
Expressiveness	Slow	Fast	- 0.25	0.02	16.08	< 0.001***
	Slow	Medium	- 0.18	0.02	11.55	< 0.001***
	Medium	Fast	- 0.07	0.02	4.54	0.002**

For coding of variables, see Table 1

*T* tapping, *NT* no-tapping

\*\*\**p* < 0.001, \*\**p* < 0.01, \**p* < 0.05



**Fig. 5** Line plots of the interaction effects between training and tapping (left), as well as between training and complexity (right) on Expressiveness. The whiskers represent standard errors. \*\* $p < 0.01$ ,

\* $p < 0.05$ . Significance indicators are of the same line type as the corresponding groups (i.e. musical training)

with the slow tempo (see Table 3). The model suggested significant interaction effects between training and tapping, as well as between training and complexity (see Table 1). Post-hoc comparisons with Tukey correction showed that the effect of tapping was only significant for highly musically trained participants (see Fig. 5, left panel): they perceived the performances as more expressive when tapping than when not tapping. Furthermore, the effect of complexity was only significant for highly musically trained participants (see Fig. 5, right panel): the more complex the stimuli, the more expressive they were perceived.

**Influence of tapping speed and stability**

For the tapping trials, LMMs were conducted in order to analyse specific effects of tapping in terms of speed and stability. This part of analyses focused on the effects of tapping speed, stability, and their interactions with training, tempo, and rhythmic complexity on DE, PoT, and Expressiveness.

**Duration estimation**

Significant main effects of tapping speed, stimuli tempo, and music training were found, while there was no main effect of tapping stability or stimuli complexity (see Table 4, for the full table, please refer to Table S4).

**Table 4** Summary of the mixed linear model analysis for tapping trials only

DV	Variable	B	SE	B	t	p
Duration estimation	$\beta_{ITI_{mean}}$	- 5.15	1.93	- 2.66	0.008**	
	$\beta_{Tempo}$	- 0.67	0.35	- 1.93	0.054*	
	$\beta_{MT}$	- 8.38	3.00	- 2.79	0.006*	
	$\beta_{ITI_{mean}:MT}$	4.98	2.59	1.93	0.055*	
Passage of time	$\beta_{ITI_{CV}}$	0.86	0.33	2.57	0.010*	
	$\beta_{Tempo}$	- 0.08	0.02	- 5.59	<0.001***	
	$\beta_{ITI_{mean}:CV}$	- 1.05	0.45	- 2.35	0.018*	
Expressiveness	$\beta_{Tempo}$	- 0.12	0.02	- 7.70	<0.001***	

For the full table, please refer to Table S4

$ITI_{mean}$  normalized tapping speed,  $ITI_{CV}$  tapping stability,  $Comp$  complexity,  $MT$  normalized music training scores. For coding of variables, see Table 1

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$

However, post-hoc analyses with Tukey corrections revealed no significant differences between slow and fast, slow and medium, as well as medium and fast conditions (see Table 6). A significant interaction effect between tapping speed and training on duration estimation was found. Post-hoc analyses suggest that less musically trained participants (MT score below group median) were affected by their own tapping speed: the faster they tapped, the longer they estimated the duration; whereas participants with

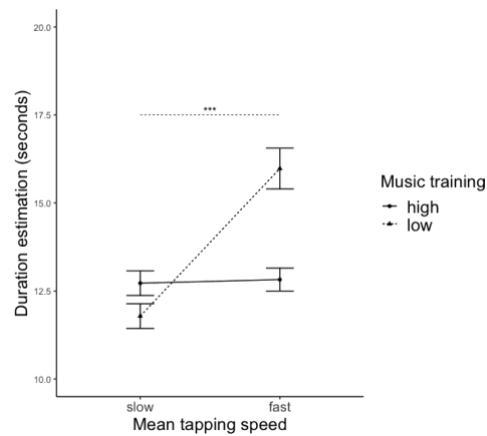
**Table 5** Summary of the post-hoc comparison with Tukey correction based on significant interactions from Table 4

DV	Condition1	Condition2	Differences of estimate	SE of difference	<i>t</i>	<i>p</i>
Duration estimation	High MT	High $ITI_{mean}$ – Low $ITI_{mean}$	0.57	0.42	1.36	0.528
	Low MT	High $ITI_{mean}$ – Low $ITI_{mean}$	1.76	0.46	3.80	<0.001***
Passage of time	High $ITI_{mean}$	High $ITI_{CV}$ – Low $ITI_{CV}$	- 0.004	0.02	- 0.22	0.996
	Low $ITI_{mean}$	High $ITI_{CV}$ – Low $ITI_{CV}$	- 0.009	0.02	- 0.50	0.959

For coding of variables, see Table 1

$ITI_{mean}$  normalized tapping speed,  $ITI_{CV}$  tapping stability, *MT* normalized music training scores

\*\*\**p* < 0.001, \*\**p* < 0.01, \**p* < 0.05



**Fig. 6** Line plots of the interaction effects between training and tapping speed on duration estimation. The whiskers represents standard errors. \*\**p* < 0.01, \**p* < 0.05. Significance indicators are of the same line type as the corresponding groups (i.e. musical training)

more musical training (MT score above group median) were not affected (see Table 5 and Fig. 6).

**Passage of time**

Significant main effects of tapping stability and tempo on PoT were found, whereas there were no main effects of complexity and training (see Table 4, for the full table, please refer to Table S4). Post-hoc analyses with Tukey corrections revealed significant differences between slow and fast, slow and medium, as well as medium and fast conditions (see Table 6). PoT was perceived the fastest with fast tempo, followed by medium tempo, and was perceived the slowest with slow tempo. Furthermore, a significant interaction between tapping stability and tapping speed was found. Post-hoc comparisons with Tukey correction suggested that the effect of mean tapping speed did not differ in relation to tapping stability (see Table 5).

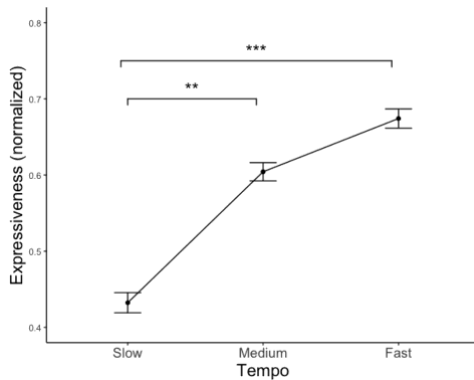
**Table 6** Summary of the post-hoc comparison with Tukey correction based on significant main effects from Table 4

DV	Condition1	Condition2	Differences of estimate	SE of difference	<i>t</i>	<i>p</i>
Duration estimation	Slow	Fast	0.44	0.31	1.41	0.387
	Slow	Medium	0.71	0.29	2.45	0.123
	Medium	Fast	- 0.27	0.28	- 0.94	0.645
Passage of time	Slow	Fast	- 0.22	0.01	15.76	<0.001***
	Slow	Medium	- 0.12	0.01	8.64	0.003**
	Medium	Fast	- 0.10	0.01	7.26	0.004**
Expressiveness	Slow	Fast	- 0.25	0.02	11.17	<0.001***
	Slow	Medium	- 0.17	0.02	7.80	0.003**
	Medium	Fast	- 0.08	0.02	3.42	0.067

For coding of variables, see Table 1

*T* tapping, *NT* no-tapping

\*\*\**p* < 0.001, \*\**p* < 0.01, \**p* < 0.05



**Fig. 7** Line plot of normalized Expressiveness grouped by performance tempo. The whiskers represent the standard errors. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

### Expressiveness

A significant main effect of stimuli tempo was found: the faster the tempo, the more expressive the performance was perceived (see Fig. 7, Table 4). Post-hoc analyses with Tukey corrections revealed significant differences between slow and fast, as well as slow and medium tempo conditions. Expressiveness was perceived to be higher with fast than slow tempo, and higher with medium than with slow tempo. However, no main effects of tapping speed, stability, complexity, training, as well as of the two-way interactions among the variables were found.

### Discussion

In this study, we performed an online experiment aiming at comparing participants' perception of time and expressiveness when tapping and not tapping to performances of a professional drummer. Participants were required to judge Duration (DE), Passage of Time (PoT), and Expressiveness in both conditions. The results suggested that time judgments and perceived expressiveness are related to motor involvement, i.e. tapping versus not tapping. Musical training also mediated the effects of motor involvement on duration estimation as well as perceived expressiveness. In addition, tempo and complexity as musical attributes have contributed to the temporal judgments.

#### Tapping versus no tapping

Regarding the effect of tapping, we found that the tapping trials were perceived shorter in durations than no-tapping

trials at slow and medium tempi. The effect of tapping is thus partially in line with our hypothesis (H1) and previous findings, such that tapping may have reduced the attentional resources allocated to the passing of time (Hammerschmidt & Wöllner, 2020; Wöllner & Hammerschmidt, 2021). As participants in the current study focused on the tapping task, they attended less to the timing task than when they were not tapping. This may have led to fewer temporal units registered in the internal clock system (Block et al., 2010), and thus shorter durations were perceived. However, the effect was only present with slow and medium tempi. One possibility is that, when tapping to fast-paced stimuli, it could be more difficult to maintain an isochronous beat which requires higher tapping stability than with slow and medium tempi, in line with audio-visual thresholds for stable tapping (Repp, 2003). In this case, participants were possibly more engaged with the timing tasks with fast tempo, which explains the absence of the tapping effect.

As an additional finding of our study, tapping had an effect on expressiveness under specific conditions: for the musically trained group, tapping trials were perceived to be more expressive than no-tapping trials. As musicians are more familiar with sensorimotor synchronisation due to training in music performance compared to non-musicians (Nguyen et al., 2022), tapping could more likely have induced higher perceived expressiveness than the no-tapping condition for them.

#### Tempo and complexity

Our study found no significant main effect of tempo on DE and PoT when both tapping- and no-tapping trials were included in the analysis. However, with only the tapping trials, tempo has a main effect on PoT: the faster the tempo, the faster the PoT. The finding partially supports our hypothesis H2. The presence of a tempo effect on PoT but not on DE has been seen in previous research. In a study where participants were asked to judge seconds-range and minutes-range durations in daily life, changes in the perceived PoT but not DE were found only for the seconds-range durations (Droit-Volet et al., 2017). This might explain our findings as we adopted 15-s stimuli. Droit-Volet and colleagues (2017) argued that duration estimation required conscious attention to time, which was less emergent in seconds-range timeframes than in minutes-range. With regard to tempo and duration estimation, the current findings are in line with a duration estimation task using tempo-shifted disco music (Hammerschmidt et al., 2021a, 2021b). Tempo differences between 105 and 125 BPM did not elicit changes in participants' perceived durations, suggesting a low sensitivity towards internal clock speed changes in the duration task.

The finding of a tempo effect on PoT only with the tapping trials highlights the role of motor involvement in time perception. This effect aligns with our hypothesis (H1) that tapping trials should pass faster than no-tapping trials, and is supported by past studies, where tapping as an additional task to the timing judgments has increased cognitive load, therefore, diverted the attention resources allocated to the timing tasks (Wöllner & Hammerschmidt, 2021). Similarly, in other prospective timing paradigms, additional tasks that entail higher cognitive load were linked to duration underestimation, as unattended temporal pulses could not register on the accumulator-counter device (Block et al., 2010).

Our finding that high complexity led to faster PoT does is in contrast to hypothesis H2 that complex music should lead to slower PoT. Previous findings revealed that participants judged complex audiovisual stimuli to last longer, indicating increases in the internal clock speed (Bueno et al., 2002; Schiffman & Bobko, 1974). One explanation could be that they adopted grouping strategies with more complex stimuli, as the musical structure exceeded one's ability to follow note by note. According to the grouping principle proposed by Lerdahl and Jackendoff (1983), listeners could segment a musical excerpt based on its hierarchical structure of the notes. It has also been pointed out that the variations in listeners' grouping strategies might be due to shifts in attention (Clarke & Krumhansl, 1990). Considering this possibility, participants in the current study may have attended to musical accents of higher metrical levels (i.e. half- or whole-note level instead of eighth- or quarter-note level) with more complex stimuli as a grouping strategy, resulting in fewer temporal units, as the internal clock was entrained to a slower pulse and faster passage of time. This is in line with findings that attention shifts to higher metrical structures led to duration underestimation, indicating a slower clock speed compared to lower metrical structures (Hammerschmidt & Wöllner, 2020). By potentially entraining the speed of the internal clock to a higher metrical level, participants might not necessarily be affected by the increased event density in complex stimuli.

For both tapping and no-tapping trials, faster tempo was related to higher perceived expressiveness. The association between performance tempo and emotional expressiveness is in line with our hypothesis (H2) that faster stimuli are perceived to be more expressive. Tempo has been regarded as one of the most important factors that facilitates the expression of emotions with music (Juslin & Madison, 1999; Juslin et al., 2001). Faster tempo has been linked to higher felt emotional arousal (Droit-Volet et al., 2013), while music perceived high in arousal level has been associated with high expressiveness (Fernández-Sotos et al., 2016). The current study is further in line with Allingham et al.'s (2021) research, as they have found an association

between increases in movement speed and a rise in perceived expressiveness.

### Musical training

For musically trained participants, the more complex the rhythms, the higher they perceived the expressiveness. The effect was absent for less trained participants. Our observation is partially in line with past research, in which non-musicians, non-drummer musicians, and drummers rated the expressiveness of drumming performances differently by allocating different weights on musical tempo, presentation modalities, and genres (Di Mauro et al., 2018). In this study, musically trained and less trained participants were both sensitive to the musical emotions expressed, whereas the trained group focused more on the technical aspects of the performances such as complexity when they judged emotional expressiveness.

Given the overall effects of motor involvement in time perception, we found that tapping speed affected duration estimation of the less musically trained group, but not the highly trained group. This partially supports our hypothesis (H3), that musical training should be associated with higher accuracy in DE and PoT. The finding highlights the role of music training in the timing and temporal judgments. The better performance among the trained group could be due to increased sensitivity towards the underlying rhythmic structure. In this way, the musically trained group (1) may have perceived the beats to be more salient than the less trained group, (2) considering that synchronizing with a beat is common practice in music training, they tapped more accurately to the drum beats, which facilitated their timing, and (3) even if there was variation in their tapping behaviour, they were less affected by it. In turn, they could better register the temporal units in the pacemaker-counter device, and estimated the stimuli duration consistently regardless of their own tapping speed. This is in line with previous studies, in which musically trained individuals outperformed the less trained group in a number of timing tasks by showing higher accuracy in duration estimation and synchronization with beats (Panagiotidi & Samartzi, 2012; Rammsayer & Altenmüller, 2006; Repp, 2010).

### Tapping speed

In addition, the effect of tapping speed on duration estimation for the less trained group also partially confirmed hypothesis H4 that faster tapping speed should lead to duration overestimation. The temporal entrainment effect (Barnes et al., 2000), describing the variation of temporal pulse emission following the rhythms of external events, offers a possible

explanation in this context. As participants changed their tapping speed, the process also elicited a changing internal clock speed by adjusting the temporal pulses emitted by the internal clock accordingly. The faster they tap, the more temporal units were accumulated, the longer a duration might be perceived. Furthermore, tapping could reinforce the temporal entrainment effect by increasing the salience of beats and drawing more attentional resources. In this way, each tap serves as a clear reference between short intervals of time that facilitate the accumulation of clock “ticks” on the counter device. Furthermore, evidence suggests that tapping can be effectively associated with the metrical levels that are registered by the variation of force (Benedetto & Baud-Bovy, 2021). The finding supports the possibility that tapping can be used to annotate how people perceive the rhythms, and consequently the perception of time passed (Hammerschmidt & Wöllner, 2020; Wöllner & Hammerschmidt, 2021).

### Limitations

A potential limitation of the current study is the sequence of tasks, especially for duration estimation and passage of time. The current study presented the PoT judgments after the duration estimation task for all participants, with both questions appearing on the webpage immediately after the stimulus. Estimation of duration may thus have affected their subsequent PoT judgments. According to the internal clock theory (pacemaker-counter mechanism), the judgment of the current time is based on the comparison of the temporal units registered to a reference duration, i.e. how fast a given amount of time should normally pass and how long it should feel, which is highly subjective (Wearden, 2015). In the current study, it is unclear what the reference duration was for each participant. Should a participant judge the stimulus to be, for example, 60 s compared to 20 s (in clock time), their PoT judgment may be under the influence of prior duration judgment. Consequently, the immediate effect of our variables such as tapping may be moderated by the duration judgments that lies between the stimulus and the PoT task. There is evidence that for an association (Droit-Volet et al., 2015) as well as a disassociation (Droit-Volet & Wearden, 2016; Droit-Volet et al., 2017) between duration estimation and PoT (Droit-Volet & Martinelli, 2023 in press). Future studies should nevertheless attempt to capture the nuances in both PoT and DE.

Another limitation of the study is that, by providing three types of duration (15 s for the main trials, 8 s and 30 s for the catch trials), there is the possibility that participants might have become consciously aware of the test durations and responded accordingly in a categorical way. Although variations both within each participant’s duration estimation and across the group have been observed, future studies should control for this potential issue by asking participant whether they had been

aware of durations and other potential control variables during their responses and had been affected accordingly.

### Conclusion

The current study investigated perceived time and expressiveness when tapping and not tapping to drumming performances that varied in tempo and complexity. Our main finding that time passed faster and felt shorter in duration (at slow and medium tempo) when tapping to drumbeats compared to when not tapping, could shed light on our time experiences in everyday scenarios such as when people are moving to music rather than passively listening to it. While motor involvement and focus of attention have clearly influenced the findings in relation to embodiment and internal clock models, more specific effects of synchronisation stability and speed call for further research.

**Acknowledgements** We are grateful to Dominik Leiner for helpful suggestions regarding the implementation of the online tapping experiment on SoSci Survey.

**Author contributions** XW, BB, and CW: contributed to the work equally.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work was supported by a Consolidator Grant to the third author from the European Research Council [grant number 725319] for the project Slow Motion: Transformations of Musical Time in Perception and Performance (SloMo).

**Availability of data and materials** Stimuli and data are accessible to the editors and reviewers.

### Declarations

**Conflict of interest** There are no competing interests.

**Ethics approval** The study was approved by the Ethics Committee at the Faculty of Humanities, University of Hamburg.

**Consent to participate** All participants have been informed and have consented to participate in the study.

**Consent for publication** The authors have agreed to the publication.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allingham, E., Hammerschmidt, D., & Wöllner, C. (2021). Time perception in human movement: Effects of speed and agency on duration estimation. *Quarterly Journal of Experimental Psychology*, 74(3), 559–572. <https://doi.org/10.1177/1747021820979518>
- Anglada-Tort, M., Harrison, P. M. C., & Jacoby, N. (2022). REPP: a robust cross-platform solution for online sensorimotor synchronization experiments. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01722-2>
- Aubry, F., Guillaume, N., Mogenicato, G., Bergeret, L., & Celsis, P. (2008). Stimulus complexity and prospective timing: Clues for a parallel process model of time perception. *Acta Psychologica*, 128(1), 63–74. <https://doi.org/10.1016/j.actpsy.2007.09.011>
- Barnes, R., Jones, M. R., Holleran, S., Hoffman, J., Large, E., Mackenzie, N., Mcauley, D., Meyer, R., & Pfor, P. (2000). Expectancy, attention, and time. *Cognitive Psychology*, 41, 254–311. <https://doi.org/10.1006/cogp.2000.0738>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v067.i01>
- Benedetto, A., & Baud-Bovy, G. (2021). Tapping force encodes metrical aspects of rhythm. *Frontiers in Human Neuroscience*, 15, 196. <https://doi.org/10.3389/fnhum.2021.633956/BIBTEX>
- Berger, J. (2014). *How music hijacks our perception of time*. Nautilus. <https://nautilus.us/how-music-hijacks-our-perception-of-time-1558/>
- Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: a meta-analytic review. *Acta Psychologica*, 3, 330–343. <https://doi.org/10.1016/j.actpsy.2010.03.006>
- Bueno, J. L. O., Firmino, É. A., & Engelman, A. (2002). Influence of generalized complexity of a musical event on subjective time estimation. *Perceptual and Motor Skills*, 94(2), 541–547. <https://doi.org/10.2466/pms.2002.94.2.541>
- Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research Psychologische Forschung*, 82(6), 1195–1211. <https://doi.org/10.1007/S00426-017-0894-2/FIGURES/7>
- Burger, B., & Toiviainen, P. (2013). MoCap Toolbox-A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the Sound and Music Computing Conference 2013* (pp. 172–178). Logos Verlag Berlin. <http://urn.fi/URN:NBN:fi:jyu-201401211091>
- Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception*, 7(3), 213–252. <https://doi.org/10.2307/40285462>
- Di Mauro, M., Toffalini, E., Grassi, M., & Petrini, K. (2018). Effect of long-term music training on emotion perception from drumming improvisation. *Frontiers in Psychology*, 9, 2168. <https://doi.org/10.3389/fpsyg.2018.02168/BIBTEX>
- Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception*, 18(1), 1–23. <https://doi.org/10.2307/40285899>
- Droit-Volet, S., & Martinelli, N. (2023). The psychological underpinnings of feelings of the passage of time. In C. Wöllner & J. London (Eds.), *Performing Time: Synchrony and Temporal Flow in Music and Dance*. Oxford University Press (In press)
- Droit-Volet, S., Ramos, D., Bueno, J. L. O., & Bigand, E. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal? *Frontiers in Psychology*, 4, 417. <https://doi.org/10.3389/fpsyg.2013.00417>
- Droit-Volet, S., Trahambas, P., & Maniadas, M. (2017). Passage of time judgments in everyday life are not related to duration judgments except for long durations of several minutes. *Acta Psychologica*, 173, 116–121. <https://doi.org/10.1016/j.actpsy.2016.12.010>
- Droit-Volet, S., & Wearden, J. (2016). Passage of time judgments are not duration judgments: Evidence from a study using experience sampling methodology. *Frontiers in Psychology*, 7, 176. <https://doi.org/10.3389/fpsyg.2016.00176/BIBTEX>
- Droit-Volet, S., Wearden, J. H., & Zélandi, P. S. (2015). Cognitive abilities required in time judgment depending on the temporal tasks used: a comparison of children and adults. *Quarterly Journal of Experimental Psychology*, 68(11), 2216–2242. <https://doi.org/10.1080/17470218.2015.1012087>
- Fernández-Sotos, A., Fernández-Caballero, A., & Latorre, J. M. (2016). Influence of tempo and rhythmic unit in musical emotion regulation. *Frontiers in Computational Neuroscience*, 10, 80. <https://doi.org/10.3389/fncom.2016.00080/BIBTEX>
- Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Perceptual coding* (pp. 203–254). Academic Press.
- Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, and Psychophysics*, 72(3), 561–582. <https://doi.org/10.3758/APP.72.3.561>
- Hammerschmidt, D., Frieler, K., & Wöllner, C. (2021a). Spontaneous motor tempo: Investigating psychological, chronobiological, and demographic actors in a large-scale online tapping experiment. *Frontiers in Psychology*, 12, 2338. <https://doi.org/10.3389/fpsyg.2021.677201/PDF>
- Hammerschmidt, D., & Wöllner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time. *Music Perception*, 37(4), 263–277. <https://doi.org/10.1525/mp.2020.37.4.263>
- Hammerschmidt, D., Wöllner, C., London, J., & Burger, B. (2021b). Disco time: the relationship between perceived duration and tempo in music. *Music & Science*, 4, 2059204320986384. <https://doi.org/10.1177/2059204320986384>
- Hogan, H. (1975). Time perception and stimulus preference as a function of stimulus complexity. *Journal of Personality and Social Psychology*, 31(1), 32. <https://doi.org/10.1037/h0076162>
- Honing, H. (2001). From time to time: the representation of timing and tempo. *Computer Music Journal*, 25(3), 50–61.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459–491. <https://doi.org/10.1037/0033-295X.96.3.459>
- Juslin, P. N., Friberg, A., & Bresin, R. (2001). Toward a computational model of expression in music performance: the GERM model. *Musicae Scientiae*, 5(1\_suppl), 63–122. <https://doi.org/10.1177/10298649020050S104>
- Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, 17(2), 197–221. <https://doi.org/10.2307/40285891>
- Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, 19(4), 527–566. [https://doi.org/10.1016/S0167-9457\(00\)00026-9](https://doi.org/10.1016/S0167-9457(00)00026-9)
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, 106(1), 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>
- Leiner, D. J. (2019). *SoSci Survey* (3.1.06). <https://www.socisurvey.de/>
- Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, 229–252. <https://www.jstor.org/stable/40285257>
- London, J. (2011). Tactus ≠ tempo: Some dissociations between attentional focus, motor behavior, and tempo judgment. *Empirical Musicology Review*, 6, 43–55.
- London, J., Thompson, M., Burger, B., Hildreth, M., & Toiviainen, P. (2019). Tapping doesn't help: Synchronized self-motion

- and judgments of musical tempo. *Attention, Perception, and Psychophysics*, 81(7), 2461–2472. <https://doi.org/10.3758/S13414-019-01722-7>
- Manning, F. C., & Schutz, M. (2013). “Moving to the beat” improves timing perception. *Psychonomic Bulletin and Review*, 20(6), 1133–1139. <https://doi.org/10.3758/S13423-013-0439-7/FIGURES/2>
- Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research Psychologische Forschung*, 80(4), 532–542. <https://doi.org/10.1007/S00426-015-0678-5/FIGURES/4>
- Mate, J., Pires, A., Campoy, G., & Estaún, S. (2009). Estimating the duration of visual stimuli in motion environments. *Psicológica*, 30(2), 287–300.
- McAuley, J. D., & Jones, M. R. (2003). Modelling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-Interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102–1125. <https://doi.org/10.1037/0096-1523.29.6.1102>
- McAuley, J. D., & Kidd, G. R. (1998). Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology*, 24(6), 1786–1800. <https://doi.org/10.1037/0096-1523.24.6.1786>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS One*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nguyen, T., Sidhu, R. K., Everling, J. C., Wickett, M. C., Gibbins, A., & Grahn, J. A. (2022). Beat perception and production in musicians and dancers. *Music Perception*, 39(3), 229–248. <https://doi.org/10.1525/MP.2022.39.3.229>
- Panagiotidi, M., & Samartzi, S. (2012). Time estimation: musical training and emotional content of stimuli. *Psychology of Music*, 41(5), 620–629. <https://doi.org/10.1177/0305735612441737>
- Polak, R., Fischinger, T., & Holzapfel, A. (2018). Rhythmic prototype across cultures: a comparative study of tapping synchronization. *Music Perception*. <https://doi.org/10.1525/MP.2018.36.1.1>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Rammisayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24(1), 37–48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Repp, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior*, 35(4), 355–370. <https://doi.org/10.1080/00222890309603156>
- Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Human Movement Science*, 29(2), 200–213. <https://doi.org/10.1016/j.humov.2009.08.002>
- Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Perception*, 24(4), 367–376. <https://doi.org/10.1525/mp.2007.24.4.367>
- Scheurich, R., Zamm, A., & Palmer, C. (2018). Tapping into rate flexibility: musical training facilitates synchronization around spontaneous production rates. *Frontiers in Psychology*, 9, 458. <https://doi.org/10.3389/FPSYG.2018.00458/BIBTEX>
- Schiffman, H. R., & Bobko, D. J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, 103(1), 156. <https://doi.org/10.1037/h0036794>
- Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: cues to pulse finding. *Music Perception*, 18(4), 455–489. <https://doi.org/10.1525/mp.2001.18.4.455>
- Styns, F., van Noorden, L., Moelants, D., & Leman, M. (2007). Walking on music. *Human Movement Science*, 26(5), 769–785. <https://doi.org/10.1016/J.HUMOV.2007.07.007>
- Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the “internal clock.” *Psychological Monographs*, 77(13), 1–31. <https://doi.org/10.1037/h0093864>
- Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5, 1111. <https://doi.org/10.3389/FPSYG.2014.01111/BIBTEX>
- Wang, X., & Wöllner, C. (2019). Time as the ink that music is written with: a review of the internal clock models and their explanatory power in audiovisual perception. *DGM Jahrbuch*, 29(2019), 1–22. <https://doi.org/10.5964/jbdgm.2019v29.67>
- Wearden, J. H. (2015). Passage of time judgements. *Consciousness and Cognition*, 38, 165–171. <https://doi.org/10.1016/J.CONCOG.2015.06.005>
- Wöllner, C., & Hammerschmidt, D. (2021). Tapping to hip-hop: Effects of cognitive load, arousal, and musical meter on time experiences. *Attention, Perception, & Psychophysics*, 83(4), 1552–1561. <https://doi.org/10.3758/S13414-020-02227-4>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.