Interplay and Malleability of Multisensory Integration and Recalibration across multiple Timescales

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

an der Universität Hamburg

Fakultät für Psychologie und Bewegungswissenschaft,

Institut für Psychologie

vorgelegt von

Alexander Kramer

Hamburg, 2023

Promotionsprüfungsausschuss:

- Prof. Dr. phil. Frank Rösler (Vorsitzender)
- Prof. Dr. rer. nat. Brigitte Röder (1. Dissertationsgutachterin)
- Prof. Dr. rer. nat. Sebastian Gluth (2. Dissertationsgutachter)
- Prof. Dr. ing. Timo Gerkmann (1. Disputationsgutachter)
- PD. Dr. rer. nat. Patrick Bruns (2. Disputationsgutachter)

Tag der Disputation: 13.09.2023

Abstract

Continuous interaction between vision and audition allows for a coherent multisensory representation of the external world, which helps us to orient our gaze towards unexpected noises or guide our auditory attention towards specific sounds. Combining information of different senses into multisensory representations is referred to as multisensory integration. To optimally use the information available in vision and audition, the perceptual system must infer which information belongs together. Spatio-temporal features are particularly important in this process. However, sensory signals from a single source might be misaligned in space and time, due to noise or inaccuracies in either vision or audition. Noise determines sensory reliabilities, i.e., the consistency of the output of the sensory system over multiple observations of an identical stimulus, whereas sensory accuracy, determines the degree to which the sensory output reflects the characteristics of the stimulus in an unbiased manner. Moreover, the perceptual system does not even know a-priori how many sources are in the world but must infer their number from sensory evidence and prior beliefs, a process called Causal Inference (CI). This challenging inferential problem is always present when distinct sensory cues provide information about a particular common feature.

Audio-visual spatial discrepancies can be induced in experimental setups. When participants localize the auditory component, the reported auditory position is shifted towards the visual position, which is commonly referred to as the Ventriloquism effect (VE). The VE is considered an example for multisensory integration and more specifically CI. Continuous exposure to audio-visual spatial discrepancies leads to shifts in subsequent unimodal auditory localization, which is often referred to as cumulative ventriloquism aftereffects (CVAE). Aftereffects can be induced by a single exposure to an audio-visual spatial discrepancy too, but this instantaneous ventriloquism aftereffect (IVAE) has been hypothesized to be mechanistically distinct from the CVAE. These aftereffects are examples of multisensory recalibration, whereby information across senses serves to maintain unisensory representation accurate. Understanding the computational principles of the Ventriloquism effect and its aftereffects is essential in order to understand how multisensory integration and recalibration interact to provide a coherent multisensory representation of the world.

In Study 1 **(Chapter III),** two sounds were paired with visual stimuli and presented with opposite directions of audio-visual spatial discrepancies. Either the auditory or the visual component had to be localized. The reliability of the visual stimulation was high in one session and low in another. Both auditory and visual unimodal stimuli were intermixed to measure auditory and potential visual aftereffects. Whereas no visual aftereffects were found reliable auditory aftereffects were found across all conditions. The auditory CVAE as well as the auditory IVAE were reduced in the low visual reliability condition. In addition, we found a visual VE when the visual reliability was low.

The paradigm of Study 2 **(Chapter IV)** followed Study 1 with some alterations. Across sessions, the absolute audio-visual spatial discrepancy was varied, changing the sensory evidence for a common cause. Furthermore, an association paradigm was applied before aftereffects were induced, where one audio-visual pair was presented spatio-temporally aligned, presumably increasing the system's prior belief of a common cause, and another was presented spatio-temporally randomly misaligned, presumably decreasing the system's prior belief of a common cause. VE, IVAE as well as CVAE increased with increasing audio-visual spatial disparity. Spatio-temporal alignment during association blocks led to an increased VE and CVAE in initial test blocks compared to misalignment during association blocks. In subsequent test blocks this pattern reversed. This modulation of CVAE and VE was limited to the large audio-visual disparity.

Model based analysis of Study 1 revealed that learning mechanisms for the CVAE and IVAE are sensitive to sensory reliabilities. Study 1 and Study 2 both suggested that the CVAE is based on a rather distinct process from the VE, that depends however on the output of multisensory integration. By contrast, the IVAE seems to be an additional outcome of the same process that underlies the VE. While in Study 2 the CVAE did depend on the posterior of a common cause, it did not in Study 1indicating that the sensory context might influence which information the perceptual system considers for recalibration.

Study 3 **(Chapter V)** investigated whether the VE and CVAE integrate explicit reward feedback to identify which of the sensory cues, vision, or audition, is inaccurate. When feedback indicated accurate audition, the VE decreased over time and no CVAE was observed. These results suggest that crossmodal recalibration and multisensory integration incorporate top-down driven feedback resulting in more accurate audio-visual spatial perception.

In summary, our results are in line with a common computational process for multisensory integration and instantaneous recalibration. Both effects result from an inference process that dissociates whether audio-visual disparities are likely due to noise, distinct causes, or inaccuracies that vary dynamically over time. The CVAE on the other hand is rather a distinct process which relies on the output of multisensory integration. Importantly, the relation between multisensory recalibration and integration is neither linear nor monotonous with respect to size of the audio-visual disparity and sensory reliabilities. Furthermore, the CVAE is fine tuned to less volatile sources of inaccuracies compared to the IVAE. Thus, it seems that the perceptual system does learn the temporal dynamics of typical sources of inaccuracies. Moreover, it accounts for these distinct sources by evolving multiple recalibration mechanisms which are adjusted to the specific dynamics of these sources. External feedback might provide an important tool when it comes to learning about these sources of sensory inaccuracies and might therefore shape multisensory recalibration.

Contents

Abbreviations

aCVAE	Auditory cumulative ventriloquism aftereffect
aIVAE	Auditory instantaneous ventriloquism aftereffect
ANOVA	Analysis of variance
aVE	Auditory ventriloquism effect
BOLD	Blood oxygenation level-dependent
CI	Causal Inference
CVAE	Cumulative ventriloquism aftereffect
EEG	Electroencephalography
eF	Estimated frequency
EMM	Estimated marginal mean
fMRI	Functional magnetic resonance imaging
GABA	Gamma-Aminobutyric acid
ILD	Interaural level difference
ISI	Inter stimulus interval
ITD	Interaural time difference
ITI	Inter trial interval
IVAE	Visual instantaneous ventriloquism aftereffect
LMM	Linear mixed effects model
MANOVA	Multivariate analysis of variance
MVN	Multivariate normal distribution
NPN	Non-perceptual noise
PEP	Protected exceedance probability
RMSE	Root mean squared error
SOA	Stimulus onset asynchrony
vCVAE	Visual cumulative ventriloquism aftereffect
vIVAE	Visual instantaneous ventriloquism aftereffect
VE	Ventriloquism effect
vVE	Visual ventriloquism effect

Chapter I - General Introduction

Spatial Perception is probably one of the most important servants for successful interaction with our environment. While engaging with our environment visual and auditory perception are continuously interacting to provide coherent representation of the external world, allowing us to orient our view to unexpected noises (Arnott & Alain, 2011; J. X. Maier & Groh, 2009) or guide our auditory attention in the direction of a specific person, to whom we would like to listen. The benefits of the interaction between vision and audition are multifaceted, reaching over better recognition (Kriegstein & Giraud, 2006), faster reaction times (Diederich & Colonius, 2004), improved speech intelligibility (Sumby & Pollack, 1954) to improved localization precision (Alais & Burr, 2004). To make the most of the information available in vision and audition, the perceptual system must infer which information belong together. Whether a face and a voice constitute the same speaker in a crowded environment is often a-priori unknown to the perceptual system. Importantly, vision and audition do not only convey complementary information. They also share redundant information that can be used to integrate auditory and visual objects in audio-visual objects, the process of multisensory integration. The voice and face of a speaker can be localized at the same position, there are temporal correlations between the movement of the mouth and the envelope and spectrum of the auditory input. Time points and positions are so called supra-modal features that are not bound to only one specific sense and the importance of especially spatio-temporal features has been highlighted early (Welch, 1999).

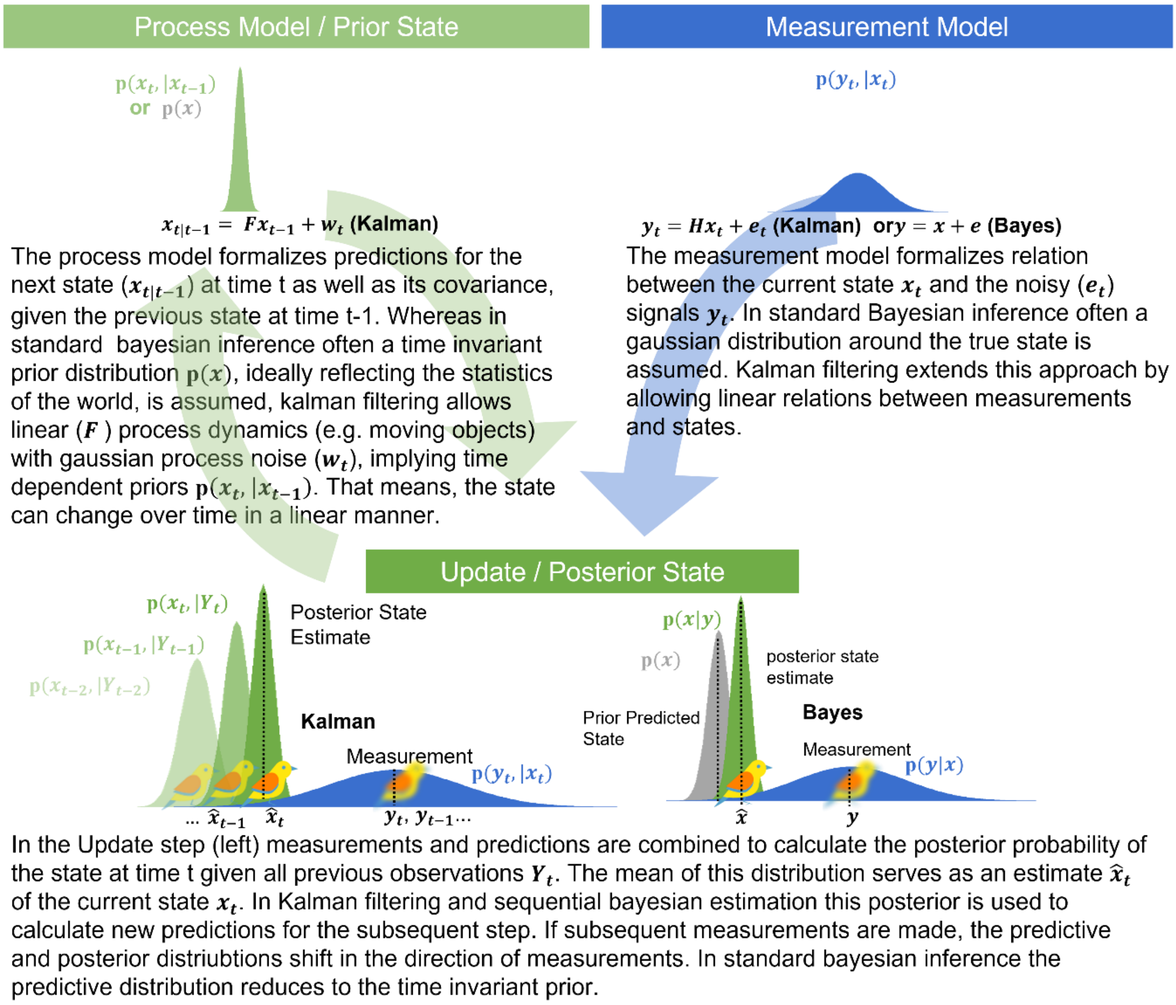
It is evident that, in order to be useful in terms of which sound belongs to which object, auditory and visual spatial representations have to be aligned. However, the information available to the perceptual system is always noisy. The spatial resolution of vision is orders of magnitude higher than the auditory resolution, the latter being in the range of several degrees (Middlebrooks & Green, 1991). If vision and audition would be processed entirely independent, then simply due to the relatively low spatial resolution of the auditory system, auditorily and visually perceived object positions would diverge by several degrees rather as a rule than exception. Yet it might just as well be, that two different sources led to the auditory and visual inputs. Imagine two birds in a tree from, one of whom is hidden but singing; an observer will likely misperceive the visible one as singing. Moreover, auditory spatial perception is highly malleable by environmental statistics (Keating & King, 2015) and the underlying physical cues change throughout development. (King, 2009). These two exemplary sources of inaccuracies differ with respect to their temporal volatility and environmental context. Moreover, they are an additional and continuous source of discrepancies in audio-visual perception. More severely, the perceptual system cannot even directly infer whether these inaccuracies should be attributed to the visual or the auditory system.

In summary, disagreement between visual and auditory spatial estimates lead to a highly non-trivial *credit assignment problem*. This credit assignment problem is not specific to audio-visual spatial perception but generalizes to multisensory and even sensorimotor processing per se, i.e., whenever noise and inaccuracies might occur.

Audio-visual spatial perception is especially suited to investigate this credit assignment problem, since audio-visual discrepancies can be induced in puristic experimental setups by presenting visual stimuli (e.g., light flashes) simultaneously with spatially offset tones. Participants must localize either auditory, visual or both stimulus components. A common finding is that the reported auditory position is shifted towards the visual position, which has been termed the Ventriloquism effect (VE, e.g., Bertelson & Aschersleben, 1998; Howard & Templeton, 1966; Lewald et al., 2001). Furthermore, exposure to audio-visual discrepancies leads to shifts in subsequent unimodal auditory localization. Different types of these so-called ventriloquism aftereffects (VAE) have been observed that differ particularly with respect to the amount of exposure needed to build them up and their temporal decay rate (Bosen et al., 2018; Bruns & Röder, 2019; Watson et al., 2019; Wozny & Shams, 2011a). These difference in build-up and decay rate potentially reflect the dynamics of different natural sources of inaccuracies. The ventriloquism effect has been interpreted as a mechanism to reduce noise (Alais & Burr, 2004; Battaglia et al., 2003) and infer the underlying causal structure (Körding et al., 2007) in an audio-visual scene, whereas its aftereffects might serve for recalibrating audition (King, 2009). All these effects are highly robust and thus provide an excellent tool to investigate the general mechanism of multisensory processing (Bruns, 2019), especially to address the credit assignment problem that arises due to cue discrepancies. Moreover, understanding the computational principles of the Ventriloquism effects and its aftereffect can clarify how integration and recalibration interact and thereby guide the search for potential neural implementations in the brain.

Perception as Bayesian Inference

The inspiration for the Bayesian approach to perception traces back to Helmholtz (1896), who framed visual perception as an inverse inference problem in which perception tries to infer which visual scene produced the retinal input. It becomes apparent, that the perceptual system must therefore itself generate models that define how the visual scene generates the input (Yuille & Kersten, 2006). The probability distribution of the visual inputs given a visual scene is called the likelihood function. Given a prior distribution that defines the a-priori probability of the visual scene, the most likely visual scene can be recovered via Bayes Rule. Assuming gaussian priors and likelihoods the posterior distribution that characterizes the probability of a certain visual scene after observations have been made, is calculated as the product of likelihood and prior. A schematic depiction of *Bayesian Inference* is given in Figure 1.



The process model formalizes predictions for the next state ($x_{t|t-1}$) at time t as well as its covariance, given the previous state at time t-1. Whereas in standard bayesian inference often a time invariant prior distribution $p(x)$, ideally reflecting the statistics of the world, is assumed, kalman filtering allows linear (F) process dynamics (e.g. moving objects) with gaussian process noise (w_t), implying time dependent priors $p(x_t, |x_{t-1})$. That means, the state can change over time in a linear manner.

The measurement model formalizes relation between the current state x_t and the noisy (e_t) signals y_t. In standard Bayesian inference often a gaussian distribution around the true state is assumed. Kalman filtering extends this approach by allowing linear relations between measurements and states.

In the Update step (left) measurements and predictions are combined to calculate the posterior probability of the state at time t given all previous observations Y_t. The mean of this distribution serves as an estimate $\hat{x}_t$ of the current state x_t. In Kalman filtering and sequential bayesian estimation this posterior is used to calculate new predictions for the subsequent step. If subsequent measurements are made, the predictive and posterior distriubtions shift in the direction of measurements. In standard bayesian inference the predictive distribution reduces to the time invariant prior.

Figure 1. A schematic overview and comparison of standard Bayesian Inference and Kalman Filtering. The Figure has been adapted from Okorokova et al. (2015).

The mean of the posterior distribution can be used as a point estimate for the state of the visual scene. This Bayesian estimator is optimal in the sense that given the uncertainty in the prior and the input, it minimizes the uncertainty in the posterior estimate, which is equivalent to minimizing the mean squared error. Unless explicitly specified otherwise we refer to this minimum mean squared error definition of optimality.

If gaussian priors and likelihoods are assumed, the posterior estimate can be interpreted as a weighted sum of the prior mean and likelihood mean, each weighted inversely proportional to its uncertainty. The Bayesian approach is especially fruitful due to its direct interpretability and can be related to feedforward (via the likelihood) and feedback (via the prior and generative model) driven processing in the brain (Yuille & Kersten, 2006). Moreover, it explicitly formalizes how noisy and often ambiguous information can be combined with prior knowledge to form consistent representations.

Ventriloquism as Bayesian Inference

The Bayesian approach has been successfully applied to describe the combination of prior knowledge with unisensory cues in a plethora of studies and sensory domains (Knill, 2003) (Jacobs, 1999; Knill and Saunders, 2003; Hillis et al., 2004). Importantly, it can be extended to multiple cues by weighting the prior and likelihood of each cue inversely proportional to its normalized uncertainty (see Chapter II for a formal definition). The inverse of cue uncertainty is referred to as reliability. Several multisensory studies manipulated the cue-reliability in one sensory modality and found that reliability determined the relative weight of the sensory modalities in multisensory integration (Fetsch et al., 2009; Gori et al., 2012; Helbig & Ernst, 2007). This pattern was also observed in audio-visual spatial integration (Alais, 2004; Battaglia et al., 2003). Battaglia et al. (2003) degraded the reliability of the visual stimulus by corrupting a random dot stereogram of a bump with noise, while Burr et al. (Alais, 2004) used stimuli with gaussian luminance envelope. Spatially disparate visual stimuli were concurrently presented with auditory stimuli. Participants had to indicate the position of the auditory stimulus and showed the typical Ventriloquism effect. Importantly, the size of the ventriloquism effect was reduced for less reliable visual stimuli compared to reliable stimuli. Hence, the weighting of audition and vision does indeed seem to depend on the relative reliabilities. Whether the weighting is actually optimal is currently put up for debate (Rahnev & Denison, 2018; Rosas & Wichmann, 2011). Some studies report suboptimal weights (Burr et al., 2009; Meijer et al., 2019), while other studies fail to demonstrate optimal reduction of uncertainty (Battaglia et al., 2003). Moreover, the size of the ventriloquism effect is not strictly determined by the relative reliabilities. Other rather top-down driven influences as associated reward (Bruns et al., 2014; Study 3, Chapter V) or task-relevancy (Rohe & Noppeney, 2018) modulate the magnitude.

Sensory Uncertainty and Bayesian Likelihoods

The unisensory representations of auditory and visual objects are usually interpreted as the phenomenal equivalent of the likelihood functions in the Bayesian framework. On the other hand, consistent distortions of space are often interpreted as priors.

For the auditory system, the width of the likelihood function (or inversely its precision) depends on multiple physical characteristics. Psychoacoustic studies show that sound localization accuracy and reliability are best in frequency ranges below 1.5 kHz and above 3.0 kHz where in former case ILDs and in the latter ITD provide most information. In the range between (1.5 –3.0 kHz range, John C Middlebrooks, 1991) where neither ILD and ITD provide reliable information, sound localization is worst. Moreover, auditory precision decreases from midline towards the auditory periphery (Middlebrooks & Green, 1991). Further auditory precision increases with wider bandwidth (Blauert, 1996). That might be because compared to pure sinusoidal sounds, broadband sounds provide additional spectral cues (Butler, 1986; Carlile et al., 1999, 2005) and facilitate robust neural coding through the engagement of more neurons (Recanzone & Sutter, 2008).

The auditory resolution is limited to several degrees (Middlebrooks & Green, 1991), but relative changes in auditory positions can be detected for as low as 1° (Grantham et al., 2003; Mills, 1958). In contrast, the spatial resolution of vision is orders of magnitude higher. Reference visual acuity in optometric tests is 1 arc min which corresponds to 1/60 degree and relative visual localization thresholds can be as low as several arc sec (1/3600 degree).

Often consistent biases in localization response patterns have been interpreted as indicators of priors, i.e. they might reflect the statistical spatial distribution of important auditory and visual objects (Odegaard et al., 2015a; Parise et al., 2014). Biases towards the eccentricity or to the center are common in auditory spatial perception and seem to reflect interindividual differences (Odegaard et al., 2015a). Similarly, biases towards the center are observed in visual spatial perception (Odegaard et al., 2015a) as well as eccentricity biases (Fortenbaugh et al., 2012; Temme et al., 1985; Werner & Diedrichsen, 2002). Recent studies have challenged the view that biases reflect spatial priors, but rather argue that distortions already occur at the sensory level, which is often associated with the computational level of likelihoods (Hong et al., 2021; Odegaard et al., 2015a). Additionally to eccentric and central biases, auditory perception is often shifted constantly in one direction (Garcia et al., 2017; Hong et al., 2021) and computational studies show that similar constant localization shifts induced through audio-visual training are rather in line with shifts in the likelihood functions (Wozny & Shams, 2011a).

Causal Inference

Standard Bayesian Inference allows the perceptual system to increase precision and fuse cues that likely have a common cause and only deviate due to noise. Yet, this implies that the system has perfect knowledge of the causal structure in the world, i.e., it knows a-priori which cues belong to the same source. As pointed out earlier, in the ventriloquism situation it might just as well be that there are distinct sources. This means that the perceptual system must take multiple generative models into account: One in which the two signals are caused by one source and another in which two sources caused the signals. Inferring the causal structure of a scenery is referred to as *Causal Inference* (Körding et al., 2007). This framework generalizes Bayesian Inference to discrete latent states. This extension allows to apply this framework to a wide range of perceptual problems, for instance whether an object is rather a tilted circle or an upright ellipse (Knill, 2007a). According to the Causal Inference framework, the perceptual system infers the posterior probability of different possible scenarios, given the available sensory cues and prior beliefs about the probability of a common cause. Studies have shown that the multisensory percept is formed as a weighted average of estimates derived from two optimal models, one for the common cause scenario, the other assuming distinct causes (Körding et al., 2007; Shams & Beierholm, 2010). This approach has been found to fit empirical data in several audio-visual localization studies (Odegaard & Shams, 2016; Rohe & Noppeney, 2015; Wozny et al., 2010).

Learning in Bayesian Inference

The benefits from Bayesian Inference emerge from the combination of sensory evidence with prior knowledge. Yet, there are many unknowns, that must be assumed a-priori to perform these computations. The perceptual system must learn which possible models of the world are likely candidates. In the ventriloquism scenario, this comes down to the question whether there are one or two objects. But if situations with an increasing number of stimuli are considered, the structural inference problem becomes far more complicated. Furthermore, to each of these scenarios an a-priori probability must be assigned, that might in reality change dynamically with changing contexts. Moreover, Bayesian likelihoods assume a correct mapping between sensory cues and states of the world. As already pointed out, in spatial audition these cues are highly dynamic throughout development and in the mature auditory system. Consequently, the parameter and structure of Bayesian computations must be dynamic and subject to learning as well, to keep perception accurate over time.

Learning Causal Priors

It is yet an open question how priors of a common cause are acquired and how these priors change over time to fit to the actual statistic of a scene. Causal priors should be highly flexible due to the volatility of causal structures per se in natural scenes. In fact, in the presence of multiple objects, the system must even be able to store multiple causal priors simultaneously. Several studies showed that priors are indeed malleable by experience. Adams, Graf, & Ernst (2004) showed that "the light from above prior" can be changed via tactile feedback. In a depth perception paradigm, Knill (2007b) demonstrated that the prior probability of perceiving slanted circles in contrast to upright ellipses could be altered by giving tactile feedback to participants consistent with one the two possible interpretations. Van Wanrooij et al. (2010) paired LED flashes with auditory noise bursts by presenting them either consistently spatially aligned or randomly misaligned. Participants had a faster reaction time when orienting towards the audio-visual pair that was previously presented spatially aligned. This difference in multisensory facilitation indicates different tendencies to integrate, i.e., different prior probabilities of a common cause, for the auditory and visual stimuli depending on the stimulus history.

Similarly, when distinctively colored visual stimuli are uniquely paired with sine tones of different frequencies by either presenting them spatially and temporally aligned or randomly misaligned with respect to space and time, subsequently measured VEs are larger for pairs with spatially and temporally consistent stimulus history (Tong et al., 2020). Again, indicating that the distinct stimulus histories led to differences in the prior probability of a common cause. Similar results have also been found in the audio-visual temporal domain (Habets et al., 2017). However, in a similar paradigm compared to Tong et al. (2020), Odegaard et al. (2017) observed increased Ventriloquism effects when sound and visual stimuli were presented temporally aligned but spatially unaligned in advance. The authors hypothesized that the spatial disparity indicated a distinct cause, whereas the temporal synchronicity implied a common cause. This mismatch might have served as an error signal to update the prior probability of a common cause. However, synchronous but spatially disparate stimuli might as well indicate that albeit the stimuli are caused by one source, they must not be perfectly aligned. A proper strategy for the perceptual system for this scenario would be to widen the prior under the common cause assumption, i.e., the shape of the prior is changed so that larger audio-visual discrepancies become more likely even under a common cause. Such a change of the shape of a prior has been demonstrated in sensorimotor recalibration (Burge et al., 2008). Hence, whereas the paradigm of Tong et al. (2020) seemed to have changed the integral of the common cause prior, the paradigm of Odegaard, Wozny, & Shams (2017) might have changed the shape of the common cause prior. Therefore, the common cause prior might be malleable in multifaceted ways.

Knill (2007b) provide a computational model for a potential update procedure that can also be applied to the common cause prior. Under the assumption of that the prior changes over time in accordance with a gaussian random walk this model learns the new priors sequentially over time. Importantly, for audio-visual spatial perception, it has never been tested whether the changes observed by aligned or misaligned pairing are in line with learning models of the prior probability of a common cause.

Learning Likelihood Functions

Auditory spatial cues depend on head size, body shape and the shape of the pinnae, thus plasticity in the auditory system is necessary throughout development to compensate for altered spatial cues (Keating & King, 2015; Mendonça, 2014). Monaural lesions or simply a flue are just two examples of how auditory spatial cues can change in adulthood and lead to inaccurate spatial perception. Spatial fine-tuning is however paramount to maintain consistent multisensory representation over time, which form the base for many of the benefits of multisensory perception and are implicitly assumed in Causal Inference. Indeed, it has been shown that auditory horizontal perception remains plastic in adulthood, beyond ILD and ITD adaptation. Several studies show that adults can accommodate to altered binaural cues (Bauer et al., 1966; Mendonça et al., 2013) as well as altered monaural cues (Carlile et al., 2014; Carlile & Blackman, 2014; Mendonça et al., 2013). Importantly, when normal cues were restored in these studies, auditory perception also returned with its initial properties, devoid of any aftereffects. This indicates that rather new mappings between cues and auditory space were learned and kept in parallel (Keating & King, 2015). Yet genuine remapping has been observed in adult barn owls when they were allowed to hunt (Bergan et al., 2005). The prism goggles first induced a constant shift of the visual input, relative to the auditory input in the direction of prism refraction. Throughout exposure to the prism glass, midbrain representations of auditory space got remapped to realign visual representations (DeBello et al., 2001). After removal of the prism glasses visual input was not shifted anymore, hence auditory space was misaligned again, now in the opposite direction. This shift of the auditory map can be measured as a behavioral aftereffect that manifests in a changed head-orienting responses (Knudsen & Knudsen, 1990).

The Aftereffects of Ventriloquism

Similar aftereffects can be induced in the Ventriloquism paradigm (Bertelson et al., 2006; Radeau & Bertelson, 1974; Recanzone, 1998). When the audio-visual disparity between visual and auditory stimulus remains constant over time, rapid aftereffects emerge, that become apparent in subsequent unimodal auditory localization shift in direction of the audio-visual disparity. Usually the ventriloquism aftereffect is only a fraction of the audio-visual disparity (10% - 50%) (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009) and when ventriloquism effect and aftereffect are reported within one study the ventriloquism effect is significantly larger (Bosen et al., 2017, 2018; Rohlf et al., 2020). Like the ventriloquism effect (Lewald et al., 2001; Slutsky & Recanzone, 2001), the ventriloquism aftereffect is reduced when visual and auditory stimuli are not presented in synchrony (Radeau & Bertelson, 1977).

Taken these similarities and given that both effects are a direct cause of exposure to audio-visual discrepancies, both effects have been previously regarded as two sides of the same coin and therefore the aftereffect has been considered as a measure for the ventriloquism effect (Radeau, 1994; Vroomen & Stekelenburg, 2014). Several studies investigating the neural underpinning of both effects demonstrated at least partially overlapping networks in the planum temporal (Bonath et al., 2007, 2014; Callan et al., 2015; Zierul et al., 2017). However, with respect to temporal marker of the processing stages of ventriloquism effect and its aftereffect, differences emerge: EEG studies show that early processing stages around 100ms after stimulus onset are altered by the aftereffect (Bruns, Liebnau, et al., 2011), while the ventriloquism effect and effects of Causal Inference seem to occur after 200ms (Aller & Noppeney, 2019; Bonath et al., 2007). Recanzone et al. (1998) argued that the tuning properties in primary auditory cortices as well as the observed frequency dependency (Frissen et al., 2003, 2005 for frequency transfer) are well in line with the behavioral aftereffect. In sum the results led to a relative agreement that the aftereffect resembles remapping of auditory representations early along the cortical pathway. Yet, this does not exclude the possibility that the ventriloquism aftereffect might be the direct consequence of the ventriloquism effect and thereby not a dissociate process. A recent developmental study (Rohlf et al., 2020) showed that children already at the age of 5 years show a ventriloquism effect that seems to follow the rules of Causal Inference, whereas the ventriloquism aftereffect appeared not before the age of 8 years. This implies that proper audio-visual integration might be a prerequisite for recalibration. However, this interpretation still does not rule out that recalibration might be induced by the ventriloquism effect. In Chapter II we provide a taxonomy for computational models of recalibration, which allows to dissociate integration and recalibration on a computational level.

Multiple Timescales of Learning

The existence of several learning mechanisms over distinct time courses seems to be a general pattern in unisensory (Bao & Engel, 2012; Dhruv et al., 2011), multisensory (D. M. Simon et al., 2018) as well as sensorimotor plasticity (Inoue et al., 2015; Smith et al., 2006). Fast and slow adaptation mechanisms have for instance been found in contrast adaptation (Bao & Engel, 2012), face and motion adaptation (Mesik et al., 2013) and in saccadic adaptation (Robinson et al., 2006). A common finding is that only the slow mechanism leads to persistent remapping whereas the effect of the fast mechanism dissipates quickly (Bao & Engel, 2012; Bosen et al., 2018; Smith et al., 2006; Watson et al., 2019) . This pattern has been used to demonstrate the relative independence of the learning mechanisms in several studies. The general idea is to induce long-term remapping along a stimulus dimension in an initial training period, afterwards fewer trials that train in the opposite direction along the stimulus dimension are presented in an erasure period. Throughout the erasure period the training effects are either erased or even reversed. Importantly, after a short time responses shift back in the training direction of the initial period. This so-called rebound (Bao & Engel, 2012; Smith et al., 2006; Watson et al., 2019) occurs because fast learning does not erase the effect of long-term learning but simply overrides it temporarily. When fast learning dissipates, the long-term effects dominate again. More recent results from audio-visual spatial perception (Bruns & Röder, 2019) and sensorimotor control (Inoue et al., 2015) even argue for a third recalibration mechanism that accumulates over days rather than minutes or hours. The underlying observation is that the effects of recalibration peak within one session and after training slowly decay. If training is repeated the next day or within 24h, savings from the previous training can be observed that manifest in faster learning rates or higher initial levels.

The ventriloquism aftereffect has initially been induced by audio-visual exposure over several minutes. Although longer training leads to larger aftereffects especially for large audio-visual disparities, robust aftereffects can be obtained already after 24 bimodal trials (Frissen et al., 2012). Indeed, several studies show now, that even a single exposure to discrepant audio-visual stimuli can induce an aftereffect (Bruns & Röder, 2015; Wozny & Shams, 2011a).

Mendonça, Escher, van de Par, & Colonius (2015) demonstrated that the audio-visual discrepancy in the directly preceding bimodal trial had the strongest influence on a subsequent unimodal trial. Wozny et al. (2011a) found auditory perceptual shifts induced by a single trial of remarkable 5% of the audio-visual discrepancy. If the effects of subsequent audio-visual trials accumulated at a similar rate one would have expected aftereffects of the full size of the audio-visual disparity in previous studies that used hundreds of trials (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009) of audio-visual training. Therefore aftereffects in these studies might not be the cumulative outcome of single trial recalibration (Wozny & Shams, 2011a). To test if rather two distinct forms of recalibration occur in parallel similar to the previous findings for adaptation, Bruns et al. (2015) paired a high frequency tone with a visual stimulus displaced to one side, and a low frequency tone with a visual stimulus displaced to the other side, thereby inducing aftereffects in opposite directions for the two tones. The paradigm made use of assumption that the aftereffect does not transfer across frequencies (Lewald, 2002; Recanzone, 1998). In line with this assumption average unimodal responses differed for the two tones consistent with the different disparities. However, responses were also modulated in the direction of the disparity of the directly preceding trial. In consequence two processes were postulated, a ventriloquism aftereffect that emerges slowly by accumulating evidence over trials (cumulative Ventriloquism Aftereffect, CVAE) and one that emerges instantaneously (instantaneous Ventriloquism Aftereffect, IVAE). In line with two simultaneous learning mechanisms that concurrently determine auditory localization a rebound effect has also been observed in the ventriloquism paradigm (Watson et al., 2019), providing further evidence for distinct mechanisms. Since the IVAE occurs immediately and dissipates quickly (Bosen et al., 2018) it has been argued that the IVAE unlikely induces remapping of early cortical representations. Rather it might be that the IVAE directly recruits information and the neural circuitry (Park & Kayser, 2019) of the ventriloquism effect (VE). Further supporting the hypothesis of distinct processes, it has been found that the IVAE emerges earlier across the lifespan than the CVAE (Rohlf et al., 2020).

The Processing Stage of Fast Learning

Since the first description of fast and slow learning mechanisms it has been a point of discussion whether fast learning resembles early processing stages, i.e. remapping (in sensory recalibration) and implicit learning (for sensory motor perturbations) or rather late cognitive stages, i.e. perceptual decision making or explicit strategy learning (McDougle et al., 2015). In the ventriloquism situation the possibility of explicit cognitive or decisional contributions is especially apparent since the auditory and visual precepts are usually not fully fused but a residual reportable discrepancy remains, opening the possibility of conscious compensatory behavior (Radeau, 1994; Vroomen & Stekelenburg, 2014). Bertelson & Aschersleben (1998) presented sounds left or right from the center and reduced the eccentricity stepwise until participants could not distinguish anymore on which side stimuli had been presented. Importantly when sounds were accompanied by central and synchronous visual stimulus participants failed to distinguish between left and right for larger eccentricities. This implied that albeit participants were not consciously able to report on which side of the visual stimulus the sound had been presented the apparent position of the sound was still shifted in the direction of the visual stimulus. This suggests that the VE at least contains genuine perceptual component. Moreover, it has been shown that unconsciously presented visual stimuli lead to VEs, albeit profoundly reduced (Delong et al., 2018).

Whereas these studies indicate a genuine perceptual component for the VE similar support for a perceptual component of the IVAE is yet lacking. In sensorimotor learning several studies show that fast learning is often associated with explicit learning and malleable by top-down factors as explicit feedback and task instructions (Bond & Taylor, 2015; McDougle et al., 2015; Redding & Wallace, 2002; Schween et al., 2020; Taylor et al., 2014). Only recently it has been shown that the IVAE in older adults is to a larger degree driven by the previous response history than for younger adults, demonstrating the general possibility of non-sensory drivers of the IVAE (Park et al., 2021). In a visuo-vestibular recalibration study investigating heading perception, sensory signals were complemented by explicit feedback about the true heading direction and explicit feedback seemed to induce another recalibration on top of the standard multisensory recalibration. Recalibration that was attributed explicitly to the feedback seemed to involve neural plasticity in the ventral parietal area which is rather associated with choice behavior than sensory processing (Zaidel et al., 2021). In summary, recent studies provide indications that response shifts like those observed from perceptual recalibration can evolve at the level of decision making and it is not clear whether the IVAE belongs to one or the other category of learning mechanisms.

From the Bayesian perspective Körding et al. (2007) argued that the purpose of multiple learning mechanisms might be to account for multiple sources of inaccuracies in our environment and the perceptual system itself. The processes which induce inaccuracies differ in magnitude and volatility. Previously in this Chapter we gave several examples of how inaccuracies can emerge on a developmental timescale, but also within days or hours in case of illness or injuries and even minutes or seconds when for instance room acoustics change. To better account for these distinct causes of inaccuracies the perceptual system might fine tune different learning mechanisms to the most prominent sources. By estimating the volatility of inaccuracies over time, the system can than solve the credit assignment problem which type of process is the source of inaccuracy (Kording et al., 2007). It follows, that the perceptual system should cover presumably fast transient changing sources by learning with fast learning rates and fast forgetting, and presumably lasting changes by sustainable learning.

Recalibration as Kalman Filtering

Bayesian learning formalizes the updating of prior beliefs by incoming information. Whereas in the CI-model and Bayesian Inference this procedure was only used to read out posterior estimates of visual and auditory position, the framework can be generalized to account for learning and more complex dynamics (see Figure 1 for a comparison of Bayesian Inference and Kalman Filtering). The natural extension of these models is the *Kalman Filter* (Kalman, 1960). It is again assumed that the actual state of the world, e.g., the position of a singing bird, is a latent state, that has to be inferred from sensory observations, whereby the Kalman Filter assumes a linear mapping between observations and latent states. The inference can then be separated in two steps, first given our current belief about the state of the world, e.g. the position, movement direction and speed of the bird, the future state is predicted (Figure 1, panel Process Model) accounting for the assumed dynamics (e.g., the bird rests at a branch). Second, once a new observation is made, predictions and observations are compared. Based on the prediction error (e.g., the bird flew away) the estimates of the latent states and prior beliefs are updated (Figure 1, panel Update). The Kalman Filter is closely related to the predictive coding framework and can be interpreted as predictive coding with linear dynamics and mappings between latent states and observations (Bastos et al., 2012).The ability to incorporate the own actions and the consequences withing the Kalman Filter approach made it a widely applied model for sensorimotor control (Burge et al., 2008; Franklin & Wolpert, 2011; Izawa & Shadmehr, 2011; Körding & Wolpert, 2006) but it has also been proposed as a mechanism for multisensory recalibration (Burge et al., 2010; Di Luca et al., 2009; Ernst & Luca, 2011) and sensory adaptation (Barraza & Grzywacz, 2008; Grzywacz & De Juan, 2003).

The Kalman Filter generalizes the idea to combine prior knowledge and observations weighted by uncertainty via adjusting the learning rate (or Kalman gain) accordingly. High sensory reliability and high prior uncertainty lead to fast learning because the system has little prior information and incoming observations are highly informative. On the other hand, low sensory reliability, and low prior uncertainty, lead to slow learning since incoming observation are only weakly informative and more reliable prior information is available. Overall, the Kalman Filter can well capture the proposed requirements for complementary learning mechanisms on different timescales, for instance highly volatile process rapidly induce uncertainty about their state and thereby lead to fast learning, mimicking the findings of the IVAE (Figure 1). Whereas several studies in sensorimotor control are well in line with Kalman-filtering, visuo-vestibular heading recalibration does not seem to depend on uncertainty in the sensory cues (Zaidel et al., 2011). Thus, it is not clear whether Kalman Filtering is also a suitable model for multisensory recalibration, investigating whether IVAE and CVAE follow the principles of Kalman Filtering. Up to now it is however not clear whether CVAE or IVAE approximate this optimal dynamic. Theoretical work on neural implementations of the Kalman Filter (Denève et al., 2007; Wilson & Finkel, 2009) demonstrate, that the hardware of the perceptual system is in principle capable of implementing Kalman Filters, albeit the involved computations are complex and resource demanding. Assuming that the primary objective of recalibration might be to provide coherent audio-visual representation, this objective can be achieved by simpler but suboptimal heuristics. Hence, simple heuristics as for instance exponential learning must be considered as alternative mechanisms.

Above we described how Bayesian Inference can be extended to Causal Inference by combining two cue integration models for the common cause and the distinct cause scenario and average over these models. Analogously, if the scenery is compatible with multiple causal or structural models, e.g., is there one hidden bird singing in a tree and moving from branch to branch or are their multiple resting birds singing alternately, each scenario can be represented by a dedicated Kalman Filter (a detailed description is given in Chapter II). Again, for each scenario the posterior probability is calculated, and a weighted average can provide an estimate reflecting the structural uncertainty. Early modelling studies of the CVAE suggested that recalibration should only occur, when the posterior probability of common cause is high (Sato et al., 2007), however it is not clear whether recalibration similarly to integration follows the principles of Causal Inference. Moreover, within the Causal Inference framework fully fused, fully segregated as well as partially fused audio-visual spatial estimates need to be calculated and theoretically all of them could be used to calculate errorterms for recalibration.

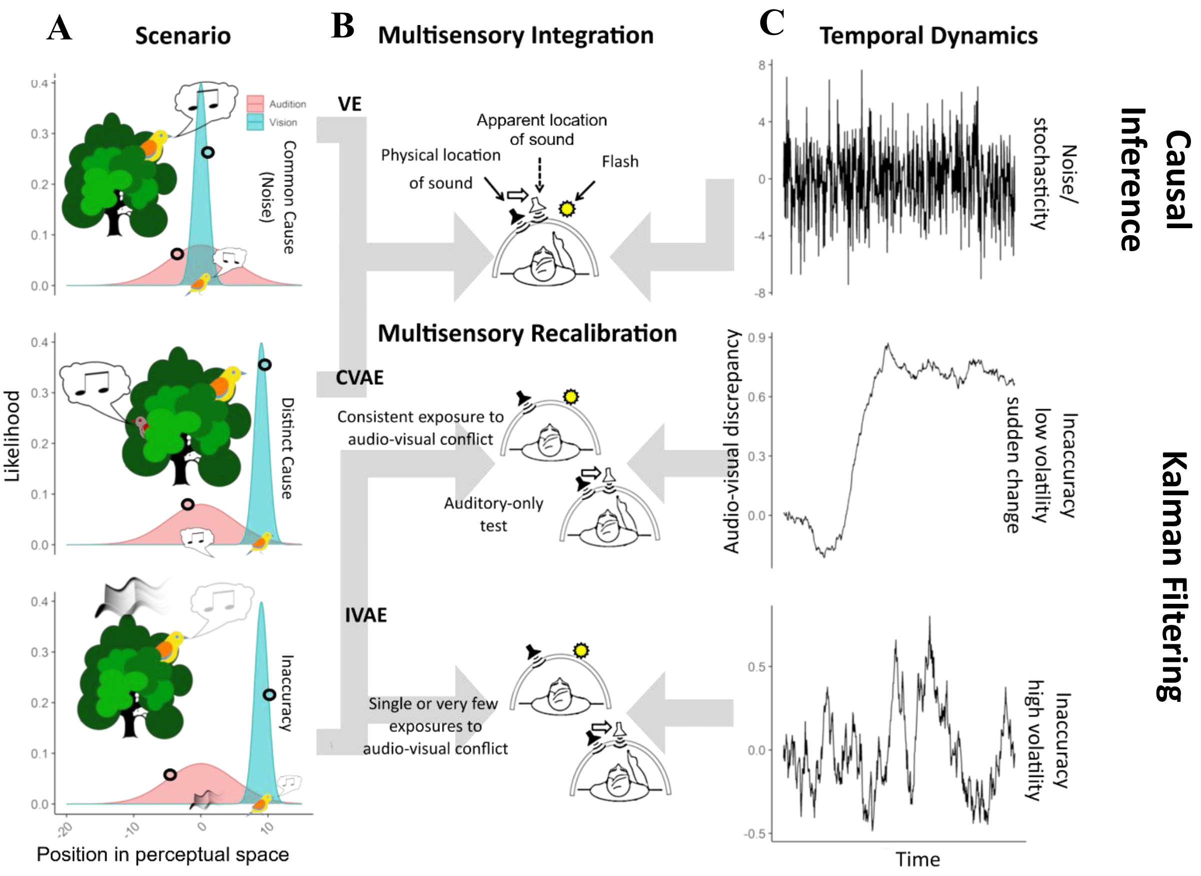


Figure 2. The relation of distinct sources of uncertainty, behavioral phenomena, and temporal dynamics of the distinct sources of uncertainty. **A:** Sceneries with distinct external structure or internal characteristics of the perceptual system that lead to similar sensory input, i.e., an audio-visual discrepancy between the perceived position of a bird singing and the bird. In the upper row there is one singing bird but due to external and internal sensory noise a discrepancy is perceived. In the central two birds are present in the tree but the more apparent one is not singing; sound source and visual source are distinct. In the bottom row there is again one singing bird, but the senses are not aligned. **B:** Behavioral phenomena that presumably help attribute the perceived audio-visual disparity to the right cause. The VE is the consequence of the perceptual systems attempt to take the common and distinct cause scenario optimally into account. CVAE and IVAE serve to realign senses. Figures are adapted from Chen & Vroomen (2013). **C:** Audio-visual discrepancies vary differently across time. Noise is centered around zero, varies quickly and has no autocorrelation (top row) whereas inaccuracies vary across time depending on the degree of volatility of the source of inaccuracy (central and bottom row) and audio-visual disparities are auto-correlated. Slowly varying sources of inaccuracies (central row) might be captured by the CVAE, whereas highly volatile sources of inaccuracies require faster recalibration, which is associated with the IVAE.

A systematic investigation, which errorterm is used for the CVAE and IVAE could further clarify the relation of integration and recalibration. If no multisensory estimates are involved in the errorterms this argues for independence of integration of recalibration and vice versa. Figure 2 summarizes the hypothetical relation between VE, CVAE and IVAE and on the one hand which part of the credit assignment problem they might solve and on the other hand which computational principle they might follow.

Learning Accuracy

As mentioned above, the concept behind recalibration as filtering (Kording et al., 2007) is that the perceptual system tries to identify the sources of its own inaccuracy and estimate their impact. CVAE and IVAE are thus interpreted as the attempt to compensate for the estimated inaccuracies. Given that the priors do reflect the accuracy of vision and audition well, recalibration can occur in an unsupervised manner to maintain internal consistency of the audio-visual representation of space (Zaidel et al., 2013). The system does not know a-priori which sensory modality is more likely to be accurate than the others and several authors have proposed that a heuristic could be to attribute inaccuracies to individual cues based on their reliabilities similar to Bayesian cue integration (Burge et al., 2010; Ghahramani et al., 1997; Makin et al., 2013; van Beers et al., 2002). When biases are small Bayesian cue combination completely ignoring biases (Scarfe & Hibbard, 2011) can under some circumstances provide better estimates than unisensory estimates. Another possibility is to assess whether the sensory estimates by one modality does lead to more rewarding interactions with the environment (Di Luca et al., 2009; Ma & Jazayeri, 2014; Zaidel et al., 2013) than the other. In this case external reward could serve as a teaching signal indicating which of the sensory modalities is more accurate. A mismatch between prior beliefs of the perceptual system and evidence provided by the reward signal could then trigger learning of more accurate priors. If priors accurately reflect the accuracy of the system, recalibration can occur selectively for only the inaccurate senses (Figure 3). Zaidel et al. (2013) found that when participants were rewarded for correct heading responses based on either the visual or the vestibular cues in a visuo-vestibular recalibration paradigm, both cues were recalibrated in the rewarded direction. In contrast, when feedback was absent, both cues were recalibrated towards each other. Zaidel et al. (2013) interpreted these results in terms of two superimposed recalibration mechanisms, where recalibration with reward can provide accuracy and recalibration without reward internal consistency. So far, it is not clear whether reward feedback can be incorporated in audio-visual spatial recalibration to maintain not only coherent but also accurate auditory and visual representations. If so, learning might be realized by updating priors or superimposing another recalibration mechanism.

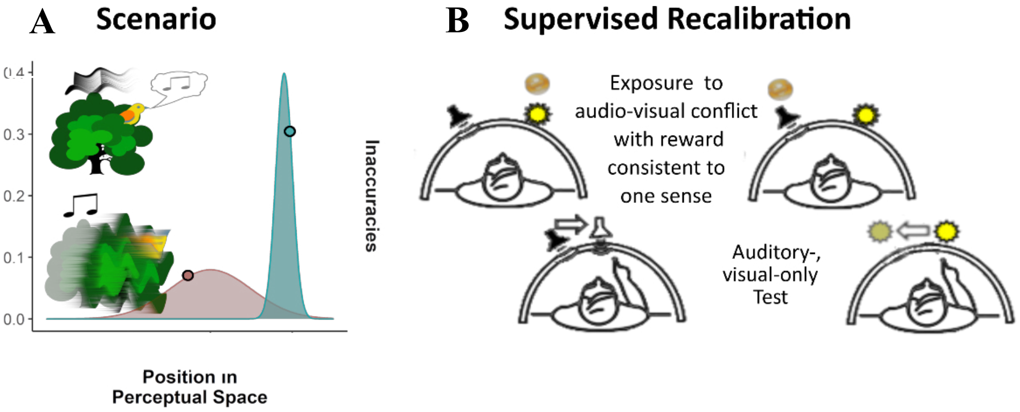


Figure 3. Supervised Recalibration to infer sensory accuracy. **A:** In this scene there is one singing bird, but the senses are not aligned. It is however not possible to determine which sense is misaligned (maybe even both). **B:** Supervised learning, whereby reward consistent with one sense serves as teaching signal, presumably helps to attribute the perceived audio-visual disparity to the right sensory modality. Only the inaccurate sense should ideally be recalibrated. Alternatively (not depicted) reward could superimpose another recalibration process dissociable from the standard CVAE and IVAE (Zaidel et al., 2013). Figures are adapted from Chen & Vroomen (2013)

Distinct Neurophysiological Prerequisites

The Bayesian computational framework for multisensory integration and recalibration largely abstracts away fundamental differences between the auditory and visual neural processing streams for spatial perception. Importantly, there have been increasing numbers of studies providing evidence against optimal perception in the Bayesian sense (Battaglia et al., 2003; Fetsch et al., 2012; M. Maiworm & Röder, 2011; Meijer et al., 2019; Rosas et al., 2005) questioning the usefulness of the ideal observer approach in general (Rahnev & Denison, 2018). With respect to audio-visual spatial perception overweighting of the visual cues is often observed (Battaglia et al., 2003; Meijer et al., 2019). Yet, from the perspective of *bounded rationality* suboptimality can be explained by bounded resources or hard constraints provided by the underlying neural circuitry (Gigerenzer & Goldstein, 1996). Bounded rationality implies that suboptimal decisions or the outcome of perceptual processes often provide "satisficing" solutions (H. A. Simon, 1956) or solutions that are only optimal if limited time, information, and for instance computational resources are considered (Gigerenzer & Goldstein, 1996).

Such considerations require to link the distinct neurophysiological principles of spatial audition and vision to computational models and subsequently behavior. Electrophysiological (Cao et al., 2019; Rohe et al., 2019) as well as an fMRI studies (Rohe & Noppeney, 2015) have linked neural activity to specific computational stages of the CI-model, yet they lack forward models that explain how computations at the neural level on the one hand approximate the behavioral findings (i.e. the VE) and on the other hand the electrophysiological or BOLD signal on the other hand. Several computational neural models have been proposed to fill this explanatory gap (Cuppini et al., 2017; Fang et al., 2019; Magosso et al., 2013; Ursino et al., 2019), however similar to the Bayesian approach visual and auditory architectures in these models are often homogeneous and only differ for instance by assuming a larger receptive field size for auditory compared to visual unimodal neurons. However, auditory and visual spatial processing streams differ remarkably.

While in vision the spatial configuration of the light pattern transduced by the retina is preserved in a topological way. That means, spatially proximal receptors on retina encode spatially proximal inputs and similarly in layers of e.g., the primary visual cortex spatially proximal neurons encode spatially proximal inputs. In contrast, the topological organization of the cochlear and auditory nerve is predominantly based on frequency. Horizontal localization is dominated by binaural cues, which are based on small differences in the input of the left and right ear, either with respect to the time of arrival or the sound level. In the low frequency spectrum (below ~1.5 kHz, Middlebrooks & Green, 1991) sound localization is dominated by interaural time differences (ITDs) whereas interaural level differences are dominant for higher frequencies (above ~ 3 kHz, Middlebrooks & Green, 1991). In the frequency range of 1.5 – 3.0 kHz, localization performance is worst since both ILDs and ITDs do provide unreliable cues (Middlebrooks & Green, 1991).

It seems that sound position is encoded by two broadly tuned channels. One channel represents the left hemispheres of auditory space and the other the right. The relative activation of these channels is used to decode azimuthal sound position (Młynarski, 2015; Stecker et al., 2005; Werner-Reiss & Groh, 2008). Several studies (Brugge et al., 1996; Lee & Middlebrooks, 2011; Rajan et al., 1990; Stecker et al., 2005) can be interpreted in terms of a third central channel and (Dingle et al., 2012) have found small proportions of neurons (~10%), that preferentially fire for central positions. Going even further Carlile et al. (2016) argue that at the stage where binaural and monaural cues are integrated, multiple spatial channels emerge with approximately 6° of separation based on psychophysical results. No matter the number of channels, these studies commonly argue for spatially broadly tuned neurons (Groh et al., 2003; Recanzone, 2000; Werner-Reiss & Groh, 2008; Woods et al., 2006). The tuning of these neurons does not allow for a so-called population code (Grothe et al., 2010), whereby single cells have narrow perceptual field tuned to specific locations (or ILDs and ITDs).

However, we are only aware of one neuro-computational CI-model (Tong et al., 2018) that models auditory spatial neurons with tuning properties in accordance with channel-based spatial encoding, several other models assume neurons with narrow perceptual fields analogous to visual neurons (Cuppini et al., 2017; Fang et al., 2019; Magosso et al., 2013; Ursino et al., 2019). Moreover, while there are fMRI studies providing results which are in line with channel-based coding of auditory space and a modulation of this coding by the aVE (Bonath et al., 2007; Callan et al., 2015) and CVAE (Zierul et al., 2017) these studies do not link the behaviorally observed aVE and aCVAE to its computational principles. Hence, a proper connection of neurophysiological findings and computational principles is yet missing. However, the stark contrast between broadly tuned auditory channels and the topographical organization in the visual cortex, where the spatial tuning is in the range of arc mins and thereby close to physical limit that is defined by the size of the receptors (Geisler, 1984) might define physiological constraints that lead to different computational properties of integration and recalibration in vision and audition.

Potential Implication for Multisensory Integration

First, several studies show a reliability dependence of the VE (Alais & Burr, 2004; Battaglia et al., 2003; Meijer et al., 2019), albeit none of these studies explicitly assessed the visual VE. Participants in these studies were instructed to localize the audio-visual stimulus as if it was caused by a single source. But localizing spatially disparate stimuli as if they originate from the same source, does not necessarily imply, that these stimuli were perceived at the same location. The former implies an explicit judgement about the causal structure which follows distinctive principles then implicit perceptual Causal Inference (Acerbi et al., 2018). Hence, visual VEs are only implicitly assumed but not explicitly measured by for instance directly instructing participants to localize the visual stimulus (as for instance in Badde et al., 2020). Hence, it is not clear whether the topographical organization of the visual cortex does allow for visual perceptual shifts via multisensory integration. A lack of visual VEs would require a reformulation of computational and neurocomputational models. If a visual VE exists, it is further unclear, whether it also follows the predictions made by the CI-model.

Potential Implication for Multisensory Recalibration

The differences in the neural organization of spatial vison and audition might have implications for the plasticity of both systems. The various aCVAE studies (Bertelson et al., 2006; Radeau & Bertelson, 1974; Recanzone, 1998) already demonstrate that auditory spatial perception is malleable by visual input. Additionally, even synchronous tactile stimulation can induce auditory aftereffects (Bruns, Spence, et al., 2011).

The plasticity of auditory spatial perception goes even beyond recalibration of spatial representations but comprises more general forms of adaptation. We use adaptation as an umbrella term for dynamic, context- and history-dependent changes of the relation between sensory stimuli and neural activity (Weber et al., 2019). Psychophysical studies with human subjects have revealed that when a sound is continuously presented at a specific adaptation location, the perceived location of subsequently presented sounds is shifted away from the adaptation position (Carlile et al., 2001; Kashino & Nishida, 1998; Stange et al., 2013; Vigneault-MacLean et al., 2007). Moreover, the auditory system quickly adapts to the ILD distributions in the environment by shifting the preferred range of neurons towards the mean of the ILD distribution and adjusting their spatial sensitivity to variance in the ILD distribution (Dahmen et al., 2010). Importantly. the changed tuning properties are also reflected in human localization behavior, indicating that the changes of early representations of auditory space propagate further to cortical representations of space that underly localization behavior. The adaptation of tuning properties to the stimulus statistic has been referred to as adaptive coding and is generally assumed to provide a mechanism to map a large range of stimulus values (which would otherwise exceed the ranges of values that can be encoded via e.g. firing rates) on neural firing patterns (Willmore & King, 2023). It is an example for the efficient coding principle, that states that neurons maximize the amount of information that they can encode about sensory inputs (Weber et al., 2019), given substantially limited resources (Willmore & King, 2023). Similar adaptive coding has been found on a neural level for ITDs in the inferior colliculus of gerbils, again psychophysical studies in humans showed behavioral effects the fit well to changes in ITD processing on the neural level of gerbils (J. K. Maier et al., 2012).

Dean et al. (2005) show that around the adapted location accuracy in neural encoding increases. Indeed, it seems that adaptation facilitates sound source separation around the adapted position (Getzmann, 2004). Sharpened spatial tuning in the presence of competing sounds has also been found in the auditory cortex (Maddox et al., 2012; Middlebrooks & Bremen, 2013). Interestingly, neural patterns of two spatially separated sounds that lead to the percept of an integrated auditory object rather correspond to the perceived position of the integrated object than to the physical positions of its two subcomponents. These results suggest that auditory spatial processing might be inherently dynamic and rather relative (with respect to previous sounds) than absolute (Lingner et al., 2018), leading to adaptive compression and dilation of auditory space as function of stimulus history. Adaptive coding of space might reflect a compromise between beneficial behavioral effects and efficient coding.

The visual system demonstrates adherence to the principles of efficient coding too. This is impressively highlighted by its sensitivity over wide ranges of stimulus strength. It provides relatively invariant perceptual representations under drastically changing illumination with respect to luminance and chroma (Webster, 2015). Moreover, it dynamically adjusts to the contrast statistics of the input thereby likely boosting the saliency of novel stimuli (McDermott et al., 2010; Wissig et al., 2013). This is again achieved by adjusting neural response functions to capture relevant scene statistics (Gardner et al., 2005). Importantly, chroma, luminance and contrast are all inherently volatile features with wide intensity ranges that must be mapped on retinotopically organized neural populations. Adaptation and particularly adaptive coding (Webster, 2015) is highly beneficial such a situation.

However, the spatial layout, at least in the horizontal and vertical spatial dimension, is directly accounted for by the neural organization on a population level. There is no need for neurons to adapt to different statistics in the vertical or horizontal dimension at least not on a magnitude comparable to non-spatial stimulus dimensions. Even under the most drastic distortions of visual space in the horizontal plane over weeks induced by prism adaptation no compensatory visual aftereffects are reported (Welch, 1978). Rather the sensorimotor system adapts to the altered visual input by recalibrating the mapping between eye and head (Redding & Wallace, 1997). In line with the hypothesis, that the visual system rather shows plasticity with respect to stimulus dimensions that are not directly inferable from retinotopic maps, the visual system is alterable by multisensory recalibration in the domains of time (Di Luca et al., 2009) and movement (C. C. Berger & Ehrsson, 2016). The former is obviously not encoded in a topographic manner and the visual system is known to have relatively low temporal resolution compared to for instance audition (Roach et al., 2006). Similarly, motion is itself inferred from spatiotemporal patterns (Lu & Sperling, 2001) and receptive fields of motion sensitive neurons become increasingly large across the processing hierarchy for motion, even up to 10° at 10° eccentricity (Andersen, 1997).

Taken together it seems that due to the organization of the auditory system, auditory spatial perception is inherently plastic. In contrast, it is an open question of whether multisensory recalibration of visual spatial perception is possible at all. It has been argued early that the visual system is rather a reference for recalibrating the other sense with respect to stimulus position (King, 2009). A study investigating the mechanism behind the aCVAE highlighted the special role of vision as a teaching signal for audition (Pages & Groh, 2013). Usually synchrony of audio-visual stimulation is a major driver for integration (Slutsky & Recanzone, 2001) and recalibration (Radeau & Bertelson, 1978), albeit in the study of Pages

& Groh (2013)., slightly delayed visual stimulation led to larger aftereffects then synchronous visual stimulation. The author assumed that the delayed visual stimulation provided feedback about the true stimulus position and thereby triggered recalibration. Prism experiments in barn owls demonstrated that on the neural level the teaching signal underlying recalibration of auditory representations in the external nucleus of the inferior colliculus were explicitly provided by the visual system (Knudsen, 2002).

However, usually highly reliable visual stimuli are used in audio-visual recalibration studies and the high reliability of the visual system might suppress visual recalibration. Recent studies manipulating visual reliability in a CVAE paradigm (Hong et al., 2021; Mahani et al., 2017) or introducing reward as an additional potential teaching signal in a VE paradigm (Bruns et al., 2014) did not test for visual aftereffects. Hence, it is not clear whether the lack of reported visual aftereffects is due to incapability of the visual system to recalibrate or whether additional teaching signals as for instance reward, or artificially lowered visual resolution can cause visual recalibration.

Aim of this Study

The underlying assumption of this introduction can be summarized from a normative perspective, namely that VE, IVAE and CVAE might serve different purposes and thereby might reflect distinct but likely interdependent processes specifically tuned to their tasks. In line with this assumption behavioral studies showed that the VE emerges from Causal Inference (Körding et al., 2007) and can therefore promote audio-visual spatial segregation and fusion depending on the most likely causal structure of the environment. Moreover, the VE can reduce noise in spatial perception (Alais, 2004; Battaglia et al., 2003; Meijer et al., 2019). On the other hand, IVAE and CVAE can at least partially recalibrate auditory spatial perception leading to more consistent representation of auditory and visual space (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009). Since the IVAE builds up instantly and decays rapidly, it seems fine-tuned to capture volatile sources of inaccuracies (Körding et al., 2007) whereas the slowly emerging CVAE might promote more persistent recalibration to temporally more stable sources of inaccuracies. This functional segregation is supported by at least partially distinct underlying neural circuitries and distinct developmental trajectories.

The objective of the present thesis is to further elucidate the interdependences of these three processes based on a computational modelling approach. In Chapter II we provide a model taxonomy that systematically arranges possible extensions of the Causal Inference for recalibration. Initially the model taxonomy allows to thoroughly formalize six research questions. First it formalizes whether the recalibration mechanisms for IVAE and CVAE consider sensory uncertainties and priors about sensory accuracies or rather uses simpler heuristics. Second the question whether VE, IVAE and CVAE are dissociable with respect to the information they recruit can be addressed by testing which errorterm underlies recalibration. The third question is whether recalibration follows the principles of Causal Inference, i.e., whether the posterior probability of a common cause modulates the course of recalibration. The fourth question refers to the processing level of the IVAE, whereby we dissociate perceptual and response level stages. The fifth question addresses the time point when the IVAE emerges. If a bimodal trial alters the IVAE does this change already affect the response to this very same bimodal trial or does this change become effective in subsequent trials? In a subsequent step the model taxonomy was extended to address the seventh question, whether the prior of a common cause can be adapted to the statistics in the experimental context.

Since the ventriloquism effect is highly sensitive to the sensory uncertainties Study 1 (Chapter III) investigated whether the CVAE is affected by sensory uncertainties at all or rather unaffected (Zaidel et al., 2011). Therefore, the visual reliability was varied across two sessions in a paradigm, that allowed to estimate VE, IVAE and CVAE within one session (Bruns & Röder, 2015).

We used a modelling approach to analyze to what extent the effects of altered sensory uncertainties in the IVAE and CVAE are mediated by distinct learning rates or changes in the process of multisensory integration. Sensory uncertainties alter multisensory integration, more specifically the size of the VE which would affect several potential errorterms. Moreover, the learning rate in Kalman Filtering depends on sensory uncertainties, this is not expected for exponential learning as simple heuristic. Therefore, Study 1 was particularly suited to address the first and second question.

In Study 2 (Chapter IV), an association paradigm adapted from Tong et al. (2020) was used to selectively increase and decrease the prior probability of a common cause for one audio-visual pair each. Afterwards CVAE and IVAE were induced, to test whether distinct prior probabilities of a common cause induce IVAEs and CVAEs of different sizes. This was done in two sessions, whereby one session included a large audio-visual disparity and the other session a small audio-visual disparity. This procedure allowed to test the third question, that recalibration should ideally only occur when a common cause for the two discrepant cues is likely. The posterior probability of common cause depends not only on the prior probability of a common cause, but also sensory evidence. Particularly the posterior probability of a common cause should vary as function of audio-visual disparity. Explicit model-based predictions were made via model simulations based on the results of Study 1 and are reported in Chapter IV. Causal Inference based (CI-based) recalibration would on the one hand suggest a further computational interdependence between recalibration and integration. On the other hand, such a finding would imply that recalibration is also fine-tuned to the causal structures in the scenery. Thereby sensory inaccuracies, which would emerge from recalibration based on cues from different sources, can be avoided. Study 1 and 2 were used in common to exploratorily investigate question 4 and 5.

To further test whether the perceptual system can learn new priors in the presence of a changing environment and more specifically incorporate explicit feedback about localization accuracy, Study 3 (Chapter V) used an audio-visual recalibration design adapted from Zaidel et al. (2013). Across four session the visual reliability was either set high or low and reward feedback was given contingent with either the veridical auditory or visual position. Several outcomes seemed likely for this study. Similarly, to Zaidel et al. (2013), a shift of auditory localization in the direction from the multisensory auditory percept to the unisensory auditory percept (i.e. in the opposite direction of the audio-visual discrepancy) would suggest a distinct reward based recalibration process. A simple suppression of the auditory CVAE, i.e., no unisensory localization shifts after recalibration would be in accordance with the interpretation that the perceptual system learns about the sensory inaccuracies. Feedback could sharpen the prior about inaccuracies in auditory spatial perception which means the perceptual system becomes more certain that it is currently not inaccurate and therefore does not recalibrate. This eighth research question was inspired by computational modelling ideas, yet we only tested behavioral predictions and did not perform model comparisons.

Finally, to assess whether visual spatial perception is malleable by audio-visual stimulation we assessed vVEs, vIVAEs and vCVAEs in Study 2, assuming that a lower visual reliability might enhance the malleability of the visual system. Moreover, we tested for vCVAEs in Study 3 to investigate whether consistent reward feedback implying accurate audition and inaccurate vision can induce shifts of visual spatial perception. A lack of recalibration of vision, even if vision is presumably inaccurate would be suboptimal from a normative perspective and in terms of bounded rationality, could imply hardwired constraints of visual spatial plasticity.

Chapter II

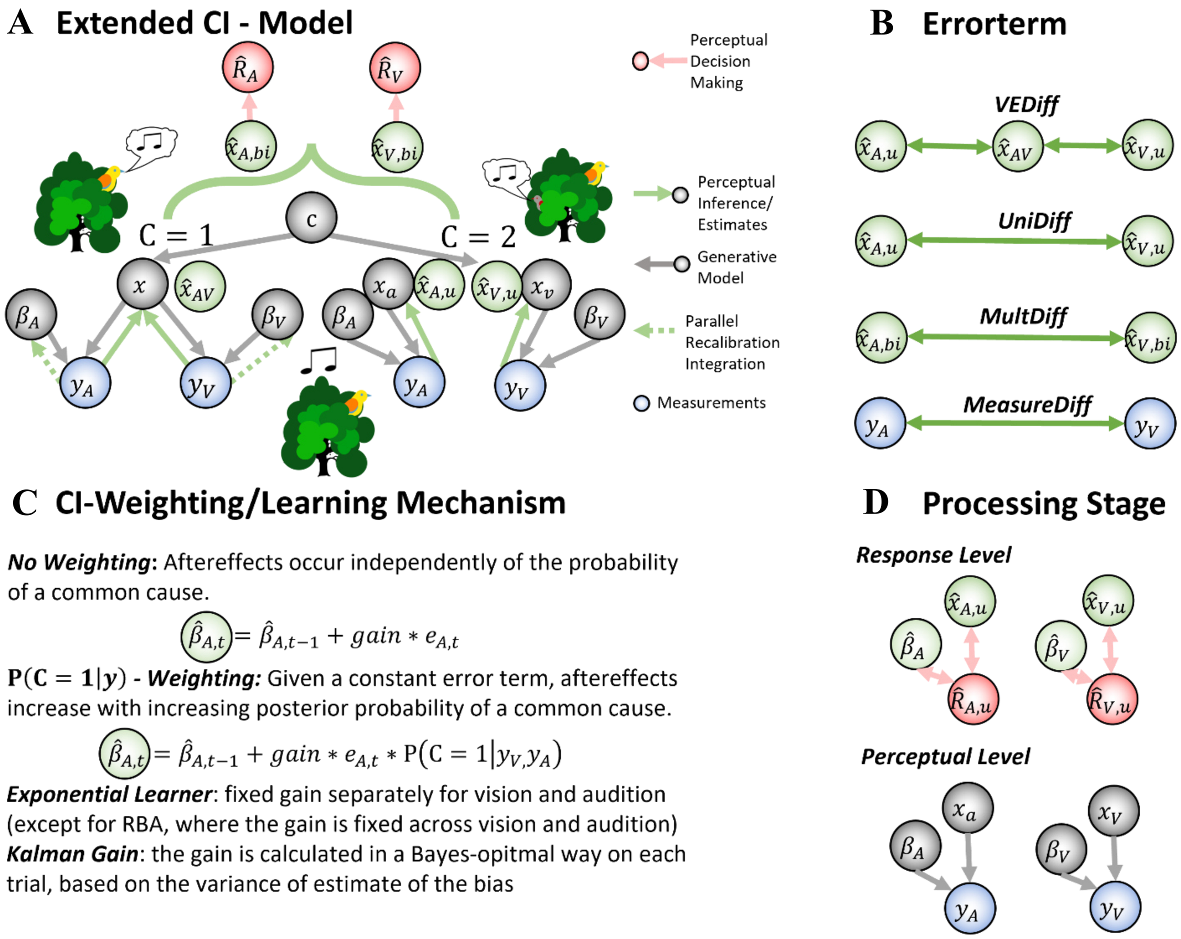
A Factorial Model Taxonomy for Multisensory Recalibration

Introduction

The standard CI-model can well explain, how the perceptual system infers the causal structure of a scenery in the presence of unsystematic noise. However, auditory and visual spatial representations can be systematically distorted. For instance, the perceptual space can be biased towards the center, periphery (Bruns et al., 2020; Odegaard et al., 2015b; Zwiers et al., 2003) or constantly in one direction (e.g. Knudsen & Knudsen, 1989; Radeau & Bertelson, 1974). If either vision or audition is distorted this leads to systematic audio-visual discrepancies. These discrepancies vary dynamically as a function of space for central or periphery biases and are well described by linear mappings between physical and perceptual spaces (Hong et al., 2021; Odegaard et al., 2015b; Shinn-Cunningham, 2000) where slopes larger or smaller than one indicate periphery or centrality biases and the intercept indicates the constant bias. In the section "Modelling of Integration", the standard CI-model is extended to account for constant and spatially varying systematic distortions in vision and audition, whereas section "Modelling of Recalibration" further extends the standard CI-model with potential mechanisms to compensate for these systematic distortions, i.e. mechanisms of multisensory recalibration.

A schematic overview of several model factors used to extend the CI-model by recalibration is given in Figure 4. First, models are differentiated based on the errorterm used (section Errorterms). The extended CI-model (Figure 4, A) provides information on the level of measurements, unisensory estimates as well as multisensory estimates to calculate erroterms (Figure 4, B) for subsequent learning. The difference between unisensory percepts (UniDiff), the difference between partially fused auditory and visual percepts (MultDiff) as well as the difference between measurements (MeasureDiff), i.e., the noisy raw cues available to the system, provide an estimate of the sum of visual and auditory biases. The difference between fully fused percept and auditory (or visual) unisensory percept provides an estimate for the bias in the auditory (or visual) system. Of particular interest is, whether errorterms are based on multisensory information (VEDiff, MultDiff in Figure 4, B) or not (MeasureDiff, UniDiff in Figure 4, B), because the former implies a direct computational link between multisensory integration and recalibration.

Further, we dissociated whether recalibration is directly affected by Causal Inference (section Posterior Weighting) or not (Figure 4,C). A direct influence of Causal Inference was modelled by weighting the recalibration step with $p(C = 1|y)$ (Badde et al., 2020; Sato et al., 2007).



No Weighting: Aftereffects occur independently of the probability of a common cause.

$$\hat{\beta}_{A,t} = \hat{\beta}_{A,t-1} + gain * e_{A,t}$$

$\mathbf{P(C = 1|y)}$ **- *Weighting:*** Given a constant error term, aftereffects increase with increasing posterior probability of a common cause.

$$\hat{\beta}_{A,t} = \hat{\beta}_{A,t-1} + gain * e_{A,t} * \mathrm{P}(\mathrm{C} = 1|y_{V,}y_A)$$

Exponential Learner: fixed gain separately for vision and audition (except for RBA, where the gain is fixed across vision and audition)
Kalman Gain: the gain is calculated in a Bayes-opitmal way on each trial, based on the variance of estimate of the bias

Figure 4. Model Taxonomy of audio-visual integration and recalibration. **A**: The extended CI–model accounts for biases (β_A, β_V) in sensory measurements (y_A, y_V) and provides several perceptual estimates (green circles) that could be used for recalibration. True states of the scenery are represented by grey filled circles, measurements are represented by blue circles and observable (by the experimenter) response behavior is represented by red circles. The figures is adapted from Körding et al. (2007). **B**: Possible errorterms for recalibration based on the extended CI–model. In the CI-model theoretically measurements, unimodal estimates as well as multisensory estimates can be used to calculate errorterms (a detailed description is given in *Errorterms*). **C**: Recalibration mechanisms. The sensory system tries to estimate its own biases ($\hat{\beta}_{A,t}$) based on an errorterm ($e_{A,t}$). The gain defines the speed of recalibration and can either be fixed (Exponential learner) or vary as a function of reliability (Kalman gain). Moreover, the gain can be scaled by the posterior probability of a common cause ($p\left(C = 1\middle|y_{V,}y_A\right)$). **D**: Recalibration models assume, that the perceptual system tries to estimate its own biases and then corrects for them. This correction can for instance occur at the perceptual level, whereby several studies suggest, that the likelihood functions of the measurements are corrected or at the level of perceptual decision making, whereby the perceptual outcome is not altered, but the response is shifted relative to the percept.

Moreover, different recalibration mechanism (Mechanisms of Recalibration) were formalized. We already introduced the optimal Kalman learning mechanism in the introduction. Since several previous modelling approaches relied on simple exponential learning (see Table

1) and as there is an ongoing debate on whether the perceptual system rather approximates optimal algorithms by simpler heuristics (Gardner, 2019), we included exponential learning as potential mechanism (Figure 1, C).

Table 1
Several well reviewed modelling approaches of the CVAE, that are covered by our proposed model categorization.

Model Name/ Description	Ref.	CI-Weighting	Errorterm	Learning Mechanism
CI-Model of Recalibration	(Hong et al., 2021; Sato et al., 2007),	$p(C = 1\|y)$ – Weighting	VEDiff	Exponential (distinct or same gains for A and V)
Remapping based on partially fused estimates	(Ernst & Luca, 2011)	No Weighting	MultDiff	Exponential (distinct gains for A and V)
Fixed Ratio Adaptation	(Zaidel et al., 2011),	No Weighting	MeasureDiff	Exponential (distinct gains for A and V)
Double Exponential Model	(Bosen et al., 2018), (Watson et al., 2019)	No Weighting	UniDiff	Exponential (distinct gains for A and V)
Reliability Based Adaptation	(Burge et al., 2010; Ghahramani et al., 1997)	No Weighting	VEDiff	Exponential (same gain for A and V)
Cue combination and cue calibration	(Burge et al., 2008)	No Weighting	UniDiff	Kalman

Finally, different processing stages (Figure 4,D) were considered for the IVAE (section Processing Stage of immediate Recalibration). In a recent study Zaidel, Laurens, DeAngelis, & Angelaki (2021) found that supervised recalibration (in which explicit spatial feedback was given) is associated with rather decision related cortical areas in the medial superior temporal area, whereas no activity changes were found in low-level perceptual areas in the ventral intraparietal area.

Likewise, several authors (Aller et al., 2022; Park et al., 2021) have argued that audio-visual spatial recalibration might not only affect perceptual stages but also post-perceptual stages of decision making. Whereas for the cumulative aftereffect neurophysiological and computational studies indicate a remapping between cues and spatial representations in early sensory processing stages (Bruns, Liebnau, et al., 2011; Recanzone, 1998; Wozny & Shams, 2011a; Zierul et al., 2017) it is not clear whether the same accounts for the IVAE. Alternatively, the IVAE might occur due to remapping between percepts and decisional choices (Aller et al., 2022; Park et al., 2021). This hypothesized dichotomy between long lasting effects of perceptual remapping and fast short-term learning on decisional or cognitive levels has been observed across multiple domains (see Chapter 1 section *Multiple Timescales of Learning*). Therefore, we explicitly modelled the IVAE as perceptual or response level effect (Figure 1, D).

Modelling of Integration

The Causal Inference Model of Integration

Several studies (extensive literature reviews are given in (Y.-C. Chen & Spence, 2017; Noppeney, 2021; Shams & Beierholm, 2010, 2022)) have shown that multisensory integration and more specifically audio-visual spatial integration is well described by a process of Causal Inference (CI). The underlying assumption of the CI-model is that two stimuli, in the ventriloquist situation an auditory stimulus s_A and a visual stimulus s_V, might either have a common cause (C = 1) or two distinct causes (C=2). Due to external and internal noise the perceptual system only has access to noisy measurements of y_A (y_V) of the actual locations x_A (x_V). Hence, the perceptual system must estimate the auditory ($\hat{x}_A$) and visual position ($\hat{x}_V$) given uncertainty about the causal structure C of the event. Generally, we use the ^-Operator for terms that the perceptual system estimates (or models describing the perceptual system). If a parameter of a model describing the perceptual system is approximated by the experimenter, we use the ~-Operator. In a common version of the CI-Model it is assumed that the perceptual system calculates separate estimates for the common cause ($\hat{x}_{A,C=1}, \hat{x}_{V,C=1}$) and distinct cause ($\hat{x}_{A,C=2}, \hat{x}_{V,C=2}$) scenario. These estimates are than combined by a weighted average, whereby the weights are the posterior probabilities of a causal scenario derived via bayes rule $p(C = k \mid y_A, y_V) = \mathrm{p}(y_A, y_V \mid C = k) * p(C = k)\ for\ k\ in\ \{1,2\}$. Hereby $\mathrm{p}(y_A, y_V \mid C = k)$ is the likelihood of the measurements given a certain causal scenario and $p(C = k)$ is the prior probability of certain scenario. Hence, $p(C = 1)$ quantifies the tendency of the perceptual system to assume a common cause before any stimuli have been observed. Several versions of the Causal Inference have been proposed.

In the standard version of the CI-Model it is assumed that the measurements are unbiased estimators of the veridical positions.

$$y_m = x_m + e_m \tag{1}$$

$, m \in \{A, V\}\ with\ e_m \sim N(0, \sigma_m)$ and $p(y_m|x_m) \sim N(x_m, \sigma_m)$. However, biases towards the eccentricity as well as the center are common in auditory spatial perception (Lewald, 2002; Lewald et al., 2000; Odegaard et al., 2015b), similarly biases towards the center are observed in visual spatial perception (Lewald & Ehrenstein, 2000; Odegaard et al., 2015b) as well as eccentricity biases (Fortenbaugh et al., 2012; Lewald & Ehrenstein, 2000; Temme et al., 1985; Werner & Diedrichsen, 2002). Hereby the difference between perceived and veridical position varies as a function of eccentricity. Moreover, constant directional biases to the left or to the right occurred in our study at baseline as in previous studies (Badde et al., 2020; Hong et al., 2021), hereby the difference between perceived and veridical position is constant across the azimuth. To account for these biases we reformulate the CI-model as a switching Kalman Filter (Murphy, 1998) similar to previous approaches (Knill, 2007a; Shams & Beierholm, 2011).

$$y = \boldsymbol{H}x + \boldsymbol{H_\beta}\beta + e \quad with\ e \sim MVN(0, \boldsymbol{R}) \tag{2}$$

Now $y = \begin{pmatrix} y_A \\ y_V \end{pmatrix}$ and $x = \begin{pmatrix} x_A \\ x_V \end{pmatrix}$ are vectors of the measurements and veridical positions, H is a 2x2 Matrix that defines a linear relationship between veridical positions and measurements. Similarly, $\beta = \begin{pmatrix} \beta_A \\ \beta_V \end{pmatrix}$ is a vector of constant biases and $\boldsymbol{H_\beta}$ is a 2x2 Matrix that defines a linear relationship between constant biases and measurements. $\boldsymbol{R}$ is the 2x2 covariance matrix of the measurement noise $e = \begin{pmatrix} e_A \\ e_V \end{pmatrix}$. We assume that $\boldsymbol{R}$ is a diagonal matrix $\begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_V^2 \end{pmatrix}$, hence the visual and auditory noise is independent. If we set $\beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\boldsymbol{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ we recover the sensation model used in the standard CI-model (Körding et al., 2007). For this study we assume, that $\boldsymbol{H}$ is a diagonal matrix indicating that auditory measurements only depend on the auditory position and analogous for vision. If the diagonal elements of $\boldsymbol{H}$ are below 1, the perceptual system assumes that the measurements are biased towards the center. Accurately recovering s from y would mean to compensate for this bias and lead to more eccentric estimates $\hat{s}$. Similarly, diagonal elements above 1 lead to more central estimates $\hat{s}$ compared to the standard model (1). Moreover, we assume that $\boldsymbol{H_\beta}$ is the identity matrix.

The basic idea of the Kalman Filter is, that the system makes a prediction for the next measurement, compares the incoming observation with these predictions and then updates the state estimates and its estimated covariance. In a first step the predictions for the visual and auditory states $\hat{x}_{k|k-1}$ and biases $\hat{\beta}_{k|k-1}$ are made

$$\hat{x}_{k|k-1} = \boldsymbol{F}\hat{x}_{k-1|k-1} \tag{3}$$

$$\hat{\beta}_{k|k-1} = \boldsymbol{F}_{\beta}\hat{\beta}_{k-1|k-1} \tag{4}$$

$$\widehat{\boldsymbol{P}}_{k|k-1} = \boldsymbol{F}\widehat{\boldsymbol{P}}_{k-1|k-1}\boldsymbol{F}^{t} + \boldsymbol{Q}_{k-1} \tag{5}$$

$$\widehat{\boldsymbol{P}}_{k|k-1} = \boldsymbol{Q}_{k-1} \tag{6}$$

Hereby $\boldsymbol{F}$ and $\boldsymbol{F}_{\beta}$are matrices that define internal forward models about how the states evolve over time from trial $k-1$ to k. For simplicity we assume that previous auditory and visual stimuli do not influence predictions for following stimuli, hence $\boldsymbol{F}$ is the $\boldsymbol{0}$ matrix. We nevertheless include $\boldsymbol{F}$ here because a choice different from $\boldsymbol{0}$ would for instance allow to model simple sequential effects as for instances biases towards previous stimulus positions or the perception of moving stimuli. $\widehat{\boldsymbol{P}}_{k|k-1}$ and $\widehat{\boldsymbol{P}}_{\beta,k|k-1}$ are the estimated covariance matrices of the predicted states and biases. They describe the systems belief about the uncertainty of the predictions. $\boldsymbol{Q}_{k-1}$ is a matrix that describes the uncertainty in the transition process, i.e., how noisy the transition from time $k-1$ to k is. Note that $\widehat{\boldsymbol{P}}_{k|k-1}$ only depends on $\boldsymbol{Q}_{k-1}$ if $\boldsymbol{F}$ is $\boldsymbol{0}$. $\widehat{\boldsymbol{P}}_{k|k-1}$ can be interpreted as the covariance of a gaussian spatial prior for auditory and visual stimulus positions.

Importantly, we define $\boldsymbol{Q}_{k-1}$ depending on the causal structure. If there are two causes the trial-to-trial noise should be independent and $\boldsymbol{Q}_{k-1,C=2} = \begin{pmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_q^2 \end{pmatrix}$ is a diagonal matrix with equal diagonal elements σ_q^2 assuming equal variance for auditory and visual spatial priors. If there is one cause, we derive $\boldsymbol{Q}_{k-1,C=1}$ by setting an intermediate matrix $\begin{pmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_m^2 \to 0 \end{pmatrix}$ rotating it by 45° (Shams & Beierholm, 2011). If $\sigma_m \to 0$ the model becomes equivalent to the standard CI, however small values of σ_m indicate that stimuli are a priori not expected to be perfectly aligned under a common cause. The resulting $\boldsymbol{P}_{k-1,C=1}$ can be interpreted as a narrow coupling prior (Ernst & Luca, 2011; Shams & Beierholm, 2011).

If $\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ then the interpretation of σ_q is identical to the standard deviation of the spatial prior in the standard CI. However, in the standard CI-model σ_q is primarily used to model centrality biases, which $\mathbf{H}$ accounts for in our model. This means that σ_q is not necessarily identifiable in our study, therefore we set $\sigma_q = 60$. This ensured that σ_q did not induce any significant additional centrality biases, but $\boldsymbol{Q}_{k-1}$ can nevertheless be properly defined. Note that if we allow for simple sequential effects or moving objects σ_q has a second interpretation, namely if σ_q is small previous stimuli have a strong effect on localization in an ongoing observation whereas this effect decreases with increasing σ_q. This interpretation justifies the formulation of the model and allows it to be applied two a wider range of phenomena.

The diverging definitions of $\boldsymbol{Q}_{k-1,C=1}$ and $\boldsymbol{Q}_{k-1,C=2}$ make it obvious that computations differ from this point onwards for the assumption of a common cause and a distinct cause. In fact, analogous to the standard CI, two distinct estimates $\hat{x}_{A,\ k|k,C=1}$ and $\hat{x}_{A,\ k|k,C=2}$ for auditory (and visual) positions are calculated reflecting the two possible causal structures. However, a strength of the Kalman Filter framework is that except for the choice of Q the computational steps are equivalent.

Given the a-priori predicted state estimates, an error is calculated when a new observation is made:

$$\hat{e}_k = y_k - (\boldsymbol{H}\hat{x}_{k|k-1} + \boldsymbol{H}_\beta\hat{\beta}_{k|k-1}) \tag{7}$$

Two auxiliary matrices are calculated:

$$\boldsymbol{S}_{\boldsymbol{k},\boldsymbol{C}=\boldsymbol{c}} = \boldsymbol{H}\widehat{\boldsymbol{P}}_{\boldsymbol{k}|\boldsymbol{k}-\boldsymbol{1},\boldsymbol{C}=\boldsymbol{c}}\boldsymbol{H}^{\mathbf{t}} + \boldsymbol{R} \tag{8}$$

$$\boldsymbol{K}_{k,C=c} = \widehat{\boldsymbol{P}}_{k|k-1,C=c}\boldsymbol{H}^{\mathbf{t}}\boldsymbol{S}_{k,C=c}^{-1} \tag{9}$$

Here $S_{k,C=c}$ is the covariance matrix of the errorterm $\hat{e}_k$ for the possible causal structures $c \in \{1,2\}$ and the $\boldsymbol{K}_{k,\boldsymbol{C}=\boldsymbol{c}}$ is the Kalman gain (Kalman, 1960). Based on the predicted states and its covariance, $\boldsymbol{S}_{k,\boldsymbol{C}=\boldsymbol{c}}$ and $\boldsymbol{K}_{k,\boldsymbol{C}=\boldsymbol{c}}$ are used to calculate the final estimates for a trial k and each causal structure c,

$$\hat{x}_{k\,,\boldsymbol{C}=\boldsymbol{c}} = \hat{x}_{k|k-1} - \boldsymbol{K}_{k,\boldsymbol{C}=\boldsymbol{c}}\hat{e}_{k|k-1} \tag{10}$$

$$\widehat{\boldsymbol{P}}_{k\,,C=c} = \widehat{\boldsymbol{P}}_{k|k-1,C=c} - \boldsymbol{K}_{k,\boldsymbol{C}=\boldsymbol{c}}\boldsymbol{S}_{k,C=c}\boldsymbol{K}_{k,C=c}{}^{-1} \tag{11}$$

The Kalman gain specifies how much the estimates depend on the observation and can be interpreted as a ratio between prediction uncertainty and observation uncertainty, which becomes more obvious in the following formulation:

$$\boldsymbol{K}_{\boldsymbol{k},\boldsymbol{C}=\boldsymbol{c}} = \widehat{\boldsymbol{P}}_{\boldsymbol{k},\boldsymbol{C}=\boldsymbol{c}}\boldsymbol{H}^{\mathbf{t}}\boldsymbol{R}_{\boldsymbol{k}}^{-1} \tag{12}$$

Analogous to the standard CI-model we have now a spatial estimate $\hat{x}_{k,\boldsymbol{C=1}}$ assuming a common cause and a spatial estimate $\hat{x}_{k,\boldsymbol{C=2}}$ assuming two causes. In a final step these estimates have to be merged into a single spatial estimate:

$$\hat{x}_k = p(C = 1|y_k)\hat{x}_{k,\boldsymbol{C}=1} + p(C = 2|y_k)\hat{x}_{k,\boldsymbol{C}=2} \tag{13}$$

whereby

$$p(C = c|y_k) = p(y_k|C = c)p(C = c)$$
$$= p\big(\hat{e}_{k|k-1}\big|C = c\big)p(C = c)$$

and

$$p\big(\hat{e}_{k|k-1}\big|C = c\big) = \varphi(\hat{e}_{k|k-1}, \boldsymbol{S}_{k,\boldsymbol{C=c}}) \tag{14}$$

with $p\big(\hat{e}_{k|k-1}\big|C = c\big) = \varphi(\hat{e}_{k|k-1}, \boldsymbol{S}_{k,\boldsymbol{C=c}})$. Hereby $\varphi(x, \Sigma)$ is the density function of a multivariate normal with covariance Σ at x . Similarly, an approximation of $\boldsymbol{\hat{P}}_k$ can be derived by the method of moments (see Murphy, 1998 for details). The components of $\hat{x}_k = \begin{pmatrix} \hat{x}_{Ak} \\ \hat{x}_{Vk} \end{pmatrix}$ are the final auditory and visual spatial percepts. Note again that if we choose $\boldsymbol{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, then with $\sigma_m \rightarrow 0$ our formulation of the CI-model converges to the standard CI-model. However, there are a couple of advantages in our formulation. Our model naturally accounts for linearly distorted spatial percepts in audition and vision with only two parameters and produces optimal estimates in presence of linear distortions (Kalman, 1960). In fact, there is a rich literature about senor fusion with biased sensors in the Kalman Filter framework (Drécourt et al., 2006), that can nurture modelling attempts of multisensory integration and recalibration. Finally, the use of a forward transition model F and the general multivariate formulation of our version allows trivial generalization to dynamic stimuli, multiple stimuli and more than one stimulus dimension.

Response Model

In study I (Chapter III) and II (Chapter IV) we used a pointing task to evaluate perceived auditory or visual localization. The pointing process is affected by multiple sources of noise independent from the perceptual process (Tassinari et al., 2006; Trommershäuser et al., 2003). We model all these sources of noise in one component as non-perceptual noise $e_{np} \sim N\big(0, \sigma_{np}\big)$. In bimodal trials participants only responded to one of the senory components $m \in M := \{A, V\}$. The resulting model for a response r_k any trial k is as follows:

$$r_k = \hat{x}_{k,M=m} + e_{np} \tag{15}$$

Approximation of R and H

In order to reduce the number of free parameters, we introduce an analytical approximation of $\boldsymbol{R}$ and make use of the following relations. From (2) it follows that:

$$\boldsymbol{E}(\boldsymbol{Cov}(\boldsymbol{y_k}, \boldsymbol{y_k})) = \boldsymbol{R} \tag{16}$$

Moreover, due to $F = \mathbf{0}$, (7) simplifies to:

$$\hat{e}_k = y_k - \hat{\beta}_{k|k-1} \tag{17}$$

Were $\boldsymbol{H}_\beta \hat{\beta}_{k|k-1}$ reduces to $\hat{\beta}_{k|k-1}$ as $\boldsymbol{H}_\beta$ is the identity matrix in our models. Then the response in a unimodal trial simplifies to:

$$r_k = \boldsymbol{K}_{k,C=2}(y_k - \hat{\beta}_{k|k-1}) + e_{np} \tag{18}$$

As this is a simple linear system, we can approximate $\boldsymbol{K}_{k,C=2}$ and $\hat{\beta}_{k|k-1}$ by linear regression with the responses as dependent variable and the true stimulus positions as regressors.
By using this approximation in:

$$\boldsymbol{R} = E\big(Cov(y_k, y_k)\big) = \tag{19}$$

$$E\left(\boldsymbol{K}_{k,C=2}\left(Cov(r_k, r_k) - Cov\big(e_{np}, e_{np}\big)\right)\boldsymbol{K}_{k,C=2}{}^{T}\right)$$

We can approximate the true R as follows:

$$\widetilde{\boldsymbol{R}} = \widetilde{\boldsymbol{K}}_{k,C=2}\left(Cov(r_k, r_k) - Cov\big(e_{np}, e_{np}\big)\right)\widetilde{\boldsymbol{K}}_{k,C=2}{}^{T} \tag{20}$$

The term $Cov\big(e_{np}, e_{np}\big)$ can be approximated based on the residual covariance of non-perceptual-noise-trials (NPN-trials), i.e. trials in which by design the perceptual noise is neglectable small, whereas $Cov(r_k, r_k)$ can be approximated by the residual covariance of the responses in unimodal trials. Given the approximations $\widetilde{\boldsymbol{K}}_{k,C=2}$ and $\widetilde{\boldsymbol{R}}$ we can solve (12) for an approximation of $\widetilde{\boldsymbol{H}}$.

If we abandon the assumption of perfect mappings between physical and perceptual space, this has important methodological implications. If we assume that eccentricity or centrality biases occur, then not only the responses are shifted to the eccentricity (or the center) but also the variance in the responses is inflated (or shrunk). That means we must estimate the mapping between physical and perceptual space to correctly approximate sensory reliabilities from participant responses. Because simple left/right paradigms or two-AFC task with only one standard do not allow to estimate this mapping, they cannot recover the correct sensory reliabilities as well.

Bias Correction

Equation (18) allows to estimate the constant biases on grouped data. If we ignore the noise in the sensory input y_k(approximated by the physical stimulus position) and the NPN e_{np} we can approximate $\hat{\beta}_{k|k-1}$ for each trial:

$$\widetilde{\boldsymbol{K}}_{k,C=2}{}^{-1} r_k - y_k = \tilde{\beta}_{k|k-1} \tag{21}$$

This noisy estimate of the bias term can be used to approximate the CVAE and moreover to correct responses for the CVAE. Thereby it is possible to fit models for integration and immediate recalibration without explicitly modelling cumulative recalibration. Otherwise, one must model all combinations of IVAE, CVAE and VE models which can quickly lead to several hundred versions of combined models.

Suboptimal Weighting in Integration

Recent studies have suggested that although integration is reliability dependent, vision is relatively overweighted (Battaglia et al., 2003; Meijer et al., 2019). We investigated two possible scenarios that could lead to a visual overweighting.

On the one hand (suboptimal estimation), the visual system might not have access to the actual visual reliabilities. Beierholm (2020) shows that in an environment of varying visual reliabilities the perceptual system does not instantly infer changes in the visual reliability (as for instance proposed by Ma & Jazayeri, 2014) but rather learns over time compatible with a Bayesian learner. Hence, a strong prior for high visual reliabilities might for instance lead to a temporary overestimation of the visual reliability (Battaglia et al., 2003). We implement this hypothesis by introducing a suboptimal estimation parameter $\omega_{SE,V}$ and define $\boldsymbol{R}_{\omega_V} = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \omega_{SE,V} * \sigma_V^2 \end{pmatrix}$ and replace $\boldsymbol{R}$ with $\boldsymbol{R}_{\omega_V}$ in (**8**) and (14). Note that not only the relative weighting of vision and audition changes as a function of $\omega_{SE,V}$ but also $p(C = c|y_k)$. For any fixed audio-visual discrepancy, the posterior probability of a common cause decreases with increasing $\omega_{SE,V}$.

On the other hand (suboptimal weighting), the visual system might have access to the actual visual reliabilities but simply overweight the visual sensory input. In this case we analogously define a suboptimal weighting parameter $\omega_{SW,V}$ so that $\boldsymbol{R}_{\omega_V} = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \omega_{SW,V} * \sigma_V^2 \end{pmatrix}$. Yet, we would calculate $\boldsymbol{S}_{k,C=c}$ according to (8) and $\boldsymbol{S}_{\omega_V,k,C=c}$ as:

$$\boldsymbol{S}_{\boldsymbol{\omega_V},\boldsymbol{k},\boldsymbol{C}=\boldsymbol{c}} = \boldsymbol{H}\widehat{\boldsymbol{P}}_{\boldsymbol{k}|\boldsymbol{k}-\boldsymbol{1},\boldsymbol{C}=\boldsymbol{c}}\boldsymbol{H}^{\mathrm{t}} + \boldsymbol{R}_{\boldsymbol{\omega_V}} \tag{22}$$

And use $\boldsymbol{S}_{k,C=c}$ in (8) and $\boldsymbol{S}_{\omega_V,k,C=c}$ in (14). Importantly, $p(C = c|y_k)$ would not vary as a function of $\omega_{SW,V}$, but would rather be correctly computed.

Modelling of Recalibration

In the previous section we described a version of CI-model that considers linearly biased sensors. The model allows for eccentricity and centrality biases by allowing that H is not the identity matrix. Constant biases in one direction are implemented in the bias vector β. The general conception is, that the perceptual system tries to estimate a mapping between the sensory inputs and the true states in the world by estimating the best values for H and β. The perceptual system has usually access to multiple spatial cues even within one sensory modality. Here we will focus on the combination of cues from different modalities, more specifically vision and audition. The calibration of H in the multisensory context is often referred to as multisensory enhancement (Bolognini et al., 2007; Frassinetti et al., 2002; Stein et al., 1988) and is usually observed when veridical auditory and visual positions are aligned during stimulation. In contrast, the estimation of β (i.e., multisensory recalibration) emerges as a relative shift of auditory and at least theoretically visual localization. These shifts are mainly observed when visual and auditory stimuli are consistently spatially disparate and usually limited to the auditory domain. Although it is possible to adaptively estimate H in the Kalman framework (Rao, 1999) we will focus on the calibration of β which is a correction of the likelihood function, that is usually identified with the ventriloquism aftereffect (Sato et al., 2007; Wozny & Shams, 2011a). This is moreover justified as there is increasing evidence that multisensory enhancement and the ventriloquism aftereffect are distinct processes with respect to their neural implementation (Bruns, 2020; Passamonti, Frissen, Ladavas, & Làdavas, 2009).

Recalibration Models

The behavioral results already demonstrated an effect of reliability on the CVAE and the IVAE. Although the VE is also modulated by the relative reliability, this does not necessarily mean that CVAE, IVAE and VE depend on the same mechanism. To our knowledge most previous computational models of the CVAE (for instance all models listed in Table 1), can be interpreted as special cases of a simplistic learning model:

$$\hat{x}_t = \hat{x}_{t-1} + a e_t \tag{23}$$

Where $\hat{x}_{t-1}$ is the estimate of some property to be learned before a new observation arrives, e_t is an errorterm calculated based on the new observation and a is a learning rate, that determines how quickly the system learns based on new observations. Neural network models of the CVAE resemble an exception from this general formulation, whereby the CVAE is for instance based on Hebbian learning (Magosso et al., 2013). However, if a is fixed the model becomes a simple exponential learner, however a could also be a function of reliability itself. For linear systems with gaussian observation noise the Kalman Filter (Kalman, 1960) provides an optimal learning rate based on the reliability of observations and the certainty about the current estimate, which allows e_t to vary as a function of reliability.

Some previous studies directly used the VE as errorterm for the CVAE (Badde et al., 2020; Hong et al., 2021) and as the VE varies as a function of reliability one would then expect the CVAE to also vary as a function of reliability. In this framework the question whether VE, CVAE and IVAE are dissociable processes reduces to the question whether the VE serves as errorterm for the CVAE and the IVAE. It becomes obvious that we have to disentangle potential effects of the errorterm and the recalibration mechanism. Hence, in the following paragraphs we formally describe potential errorterms and learning mechanism in the ventriloquism paradigm and compare the resulting models for the VE and CVAE in a Bayesian framework.

Mechanisms of Recalibration

Kalman Filtering

We already introduced the concept of Kalman Filtering in the previous section and can now conveniently reuse it to model recalibration. Therefore, we define a second filter that we apply after the integration process. We reuse $\boldsymbol{F}_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ as transition matrix for $\hat{\beta}_k$ and define $\widehat{\boldsymbol{P}}_\beta = \begin{pmatrix} \sigma_{\beta A}^2 & 0 \\ 0 & \sigma_{\beta V}^2 \end{pmatrix}$. Here $\sigma_{\beta A}^2$ and $\sigma_{\beta V}^2$ define the width of spatial priors for auditory and visual constant biases i.e., with increasing $\sigma_{\beta A}^2$ and $\sigma_{\beta V}^2$ larger biases become a priori more and more likely, and more weight is given to incoming observations relative to the prior. Thus, overall recalibration becomes faster. Moreover, for simplicity we assume a steady state filter

(Anderson & Moore, 1979), therefore we set. $\boldsymbol{Q} = \boldsymbol{0}$ and (11) is not applied to update $\widehat{\boldsymbol{P}}_\beta$, instead $\widehat{\boldsymbol{P}}_\beta$ remains fixed across trials. Additionally, we define new observations $y_{\beta k}$ and a corresponding measurement matrix $\boldsymbol{H}_\beta$ that defines an observation model for $y_{\beta k}$ and $\hat{\beta}_k$ as well as a measurement noise matrix $\boldsymbol{R}_\beta$. Thus, the Kalman Filter equations for updating the bias estimate in each trial k become:

$$\hat{\beta}_{k|k-1} = \hat{\beta}_{k-1|k-1} \tag{24}$$

$$\widehat{\boldsymbol{P}}_{\beta,k|k-1} = \widehat{\boldsymbol{P}}_\beta \tag{25}$$

$$\hat{e}_{\beta,k} = y_{\beta k} - \boldsymbol{H}_\beta \hat{\beta}_{k|k-1} \tag{26}$$

$$\boldsymbol{S}_{\beta,k} = \boldsymbol{H}_\beta \widehat{\boldsymbol{P}}_\beta \boldsymbol{H}_\beta{}^{\mathbf{t}} + \boldsymbol{R}_\beta \tag{27}$$

$$\boldsymbol{K}_{\beta,k} = \widehat{\boldsymbol{P}}_\beta \boldsymbol{H}_\beta{}^{\mathbf{t}} \boldsymbol{S}_{\beta,k}{}^{-1} \tag{28}$$

$$\hat{\beta}_k = \hat{\beta}_{k|k-1} - \boldsymbol{K}_{\beta,k} \hat{e}_{\beta,k} \tag{29}$$

Exponential Learning

The exponential learner in (23) can be formulated as a Kalman Filter were the measurement noise $\boldsymbol{R}_\beta$ is restricted to be equal across conditions and modalities (Harvey, 1986). In this case, $\boldsymbol{K}_{\beta,k}$ does not anymore depend on the actual reliabilities of the auditory and visual stimuli, rather we can use an arbitrary constant $r_\beta = 1$ across all stimuli as entrance for the diagonal of $\boldsymbol{R}_\beta$. The size of the aftereffect does then only depend on the ratio between $\sigma^2_{\beta A}$ and $\sigma^2_{\beta V}$. By assuming a constant errorterm across conditions, one obtains fixed ratio adaptation (Zaidel et al., 2011).

Errorterms

If we assume the most simplistic learning model described in (23) our initial question, whether VE, IVAE and CVAE share computational principles, reduces to the question which errorterm is used. If we consider the in- and outputs of the CI-model four distinct pieces of information are theoretically available for the recalibration process, first the measurements $y_k = \begin{pmatrix} y_{A,k} \\ y_{V,k} \end{pmatrix}$, second the two estimates under different causal models $\hat{x}_{k,C=1} = \begin{pmatrix} \hat{x}_{A,k,C=1} \\ \hat{x}_{V,k,C=1} \end{pmatrix}$ and $\hat{x}_{k,\boldsymbol{C=2}} = \begin{pmatrix} \hat{x}_{A,k,C=2} \\ \hat{x}_{V,k,C=2} \end{pmatrix}$ and finally the merged estimate $\hat{x}_k = \begin{pmatrix} \hat{x}_{Ak} \\ \hat{x}_{Vk} \end{pmatrix}$.

Errorterms independent of Multisensory Integration

Two errorterms that are independent of the magnitude of multisensory integration are based on y_k and $\hat{x}_{k,C=2}$:

$$\hat{e}_{MeasureDiff,k} := y_{A,k} - y_{V,k} \tag{30}$$

$$\hat{e}_{UniDiff,k} := \hat{x}_{A,k,C=2} - \hat{x}_{V,k,C=2} \tag{31}$$

While $\hat{e}_{UPk}$ is based on $\hat{x}_{k,\boldsymbol{C=2}}$ which assumes 2 causes and thus coincides with the unimodal percepts without any integration, $\hat{e}_{Mk}$ is based on the measurements which do obviously not depend on later processing steps. The major difference between $\hat{e}_{UniDiff,k}$ and$\hat{e}_{MeasureDiff,k}$ is, that $\hat{e}_{MeasureDiff,k}$ does not consider any possible distortions of the perceptual space.

Errorterms depending on Multisensory Integration

Moreover, we consider two errorterms that directly depend on multisensory integration:

$$\hat{e}_{MultDiff,k} := \hat{x}_{A,k} - \hat{x}_{V,k} \tag{32}$$

$$\boldsymbol{\hat{e}_{VEDiff,k}} := \boldsymbol{\hat{x}_{k,C=1}} - \boldsymbol{\hat{x}_{k,C=2}} \tag{33}$$

The error signal is hereby the residual difference between auditory and visual percepts after integration. Note that $\hat{e}_{VEDiff,k}$ is the difference between the common cause percept and the separate cause percept. If we use this errorterm in an exponential learner, the resulting recalibration mechanism is equivalent to the reliability based adaptation (Burge et al., 2010; Ghahramani et al., 1997). In both cases recalibration would need access to final or intermediate outputs of multisensory integration, speaking in favor of a computational entanglement of integration and recalibration. For an overview of errorterm used in the literature we refer to Table 1. We can approximate covariance matrices of the errorterms via the covariance matrices of the contributing estimates ($\boldsymbol{\hat{P}}_k, \boldsymbol{\hat{P}}_{k,C=1}, \boldsymbol{\hat{P}}_{k,C=2}$) and the sensory noise covariance (***R***), as these are all linear functions of the measurements and the estimates. We robustify the filter further by diagonalizing these matrices, thereby implying uncorrelated noise. For the sake of readability, we will only use the annotation terms (MeasureDiff, UniDiff, MultDiff and VEDiff) when referring to models incorporating one of these errorterms.

Pooled Recalibration and Integration

Another approach to model integration and recalibration in common is an approach that is commonly used in engineering (Friedland, 1969). We can simply augment the state vector, that usually incorporates visual and auditory positions, by the auditory and visual biases and try to estimate them in one step. We will refer to this as pooled recalibration and integration (PRI). This can be implemented by a simple adjustment of the CI-model. We summarize the observation matrices $\boldsymbol{H}$ and $\boldsymbol{H}_\beta$ in one block matrix $\boldsymbol{H}_{PRI} = \begin{pmatrix} \boldsymbol{H} & 0 \\ 0 & \boldsymbol{H}_\beta \end{pmatrix}$ and do the same for $\widehat{\boldsymbol{P}}_{PRI} = \begin{pmatrix} \widehat{\boldsymbol{P}} & 0 \\ 0 & \widehat{\boldsymbol{P}}_\beta \end{pmatrix}$ and $\boldsymbol{Q}_{PRI} = \begin{pmatrix} \boldsymbol{Q} & 0 \\ 0 & \boldsymbol{Q}_\beta \end{pmatrix}$ whereby $\boldsymbol{Q}_\beta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{P}_{\beta,c=2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, implying no update of the bias estimate, when two distinct causes are assumed. Again, we assume a steady state filter for the bias, hence P_β is not updated. This model balances integration and recalibration based on the prior uncertainty with respect to the true external precision and the sensory bias. If the system has a broad prior, where the next stimulus should occur and a narrow prior with respect to the sensory bias, it will resolve multisensory conflicts rather by integration. With increasing width of the bias prior, the contribution of recalibration to conflict resolution will increase. The pooled recalibration and integration model is the most stringent implementation of the idea of a single process for recalibration and integration, since integration and recalibration are not split in a stepwise procedure.

Posterior Weighting

All the errorterms introduced so far monotonically increase as a function of the veridical audio-visual discrepancy, which in turn would also lead to a monotonical increase of the aftereffect. However, the same constraint for integration -namely, only to integrate, when a common cause is likely, should apply to a beneficial recalibration system.

We chose a heuristic approach to make the recalibration process sensitive to the possible causal structure. In the posterior weighted version ($p(C = 1|y_k) -$ Weighting), we replace equation (29) with:

$$\hat{\beta}_k = \hat{\beta}_{k|k-1} - \boldsymbol{K}_{\beta,k} * p(C = 1|y_k) * \hat{e}_{\beta k} \tag{34}$$

Note that $p(C = 1|y_k) * \hat{e}_{VEk} = p(C = 1|y_k) * (\hat{x}_{k,C=1} - \hat{x}_{k,C=2})$ is essentially the Ventriloquism effect in a trial and equivalent to the errorterm $\hat{x}_k - \hat{x}_{k,C=2}$. Combined with an exponential learner we recover the causal-inference-based model of recalibration in Hong et al. (2021).

Transfer of Recalibration

Although on a group level primarily leftward adaptation seemed to be suppressed, on an individual level we observe, that several participants show either a dominant aftereffect to the right or to the left that transferred to the tone, that was adapted in the opposite direction. To capture this response pattern, we introduced a transfer factor for audition τ_A^{VAE} and vision τ_V^{VAE} in the range between 0 and 1, whereby zero would indicate no transfer and one full transfer. If recalibration occurred for audio-visual pair AV1 according to (29) the bias term for AV2 will also be updated as follows (and vice versa):

$$\hat{\beta}_{AV2,k} = \hat{\beta}_{AV2,k|k-1} - \boldsymbol{T}\boldsymbol{K}_{AV1,\beta,k}\hat{e}_{\beta,k} \tag{35}$$

With $\boldsymbol{T} = \begin{pmatrix} \tau_A^{VAE} & 0 \\ 0 & \tau_V^{VAE} \end{pmatrix}$.

Processing Stage of immediate Recalibration

Perceptual level

Immediate recalibration differs from cumulative recalibration in three ways. First, immediate recalibration occurs faster, sometimes even after a single trial (Bruns & Röder, 2015; Wozny & Shams, 2011a). Second, cumulative recalibration leads to temporally stable aftereffects (Frissen et al., 2012; Machulla et al., 2012) whereas immediate recalibration dissipates quickly over time, even in absence of counter evidence (Bosen et al., 2018). Thirdly, immediate recalibration transfers across frequencies in a more consistent way (Bruns & Röder, 2015) than cumulative recalibration (Bruns & Röder, 2019). If a tone is consistently presented with an audio-visual discrepancy e.g., to the right, only a few subsequent audio-visual trials with a different tone (differing up to several octaves in frequency) and the opposite audio-visual discrepancy are necessary to shift the perception in opposite direction. The previously presented framework for modelling recalibration can well account for all three phenomena. The speed of recalibration depends on the width of spatial priors for the bias term, hence, choosing $\widehat{\boldsymbol{P}}_{\beta}$ with larger diagonal entries leads to faster recalibration (see Kording, Tenenbaum, & Shadmehr, 2007 for a similar approach to model fast and slow adaptation). Bosen et al.(2018) used a temporal exponential decay factor that depended on the interstimulus interval. For simplicity we assumed an equal interstimulus interval across all trials, thereby the decay factor reduces to a constant d_{iVAE} between 0 and 1. The natural way to incorporate d_{iVAE} in the Kalman Filter is to redefine $\boldsymbol{F}_{\beta iVAE} = \begin{pmatrix} d_{iVAE} & 0 \\ 0 & d_{iVAE} \end{pmatrix}$ in the prediction step. Note that each model consisted of two chained Kalman Filter, first the integration mixture filter, second the recalibration filter. Importantly, only in the first filter we used $\boldsymbol{F}_{\beta iVAE}$ whereas in the second filter we used the standard transition matrix $\boldsymbol{F}_{\beta}$. Otherwise we would have decreased the immediate aftereffect twice per trial. We modelled transfer across frequencies analogously to cumulative recalibration according to (35) by introducing τ_A^{iVAE} and τ_V^{iVAE} as additional parameter.

Response Level

Previous modelling attempts implicitly assumed that the IVAE operates at the same level as the CVAE by shifting the likelihood functions. However, recent discussions highlight, that multisensory influences can be observed at multiple stages across the multisensory processing hierarchy (Rohe & Noppeney, 2015) and especially also at the level of perceptual decision making (Aller et al., 2022; Park et al., 2021). Centrality biases in multisensory time estimation are for instance likely due to post-perceptual integration of prior information (Murai & Yotsumoto, 2018). If the IVAE would be better described by a model at the level of decision making this would also speak for the initial hypothesis of two distinct recalibration processes. We assume that the IVAE $\hat{\beta}_{k,iVAE}$ does not influence multisensory integration, therefore (2) and (7) become:

$$y = \boldsymbol{H}x + e \quad with\, e \sim MVN(\boldsymbol{0}, \boldsymbol{R}) \tag{36}$$

$$\hat{e}_k = y_k - (\boldsymbol{H}\hat{x}_{k|k-1}) \tag{37}$$

In contrast the response model is now affected by $\hat{\beta}_{k,iVAE}$:

$$r_k = \hat{x}_{k,M=m} - \hat{\beta}_{k-1,iVAE} + e_{np} \tag{38}$$

$$r_k = \hat{x}_{k,M=m} - \hat{\beta}_{k,iVAE} + e_{np} \tag{39}$$

In perceptual models the response is based on the perceptual estimates which are in turn affected by the IVAE. So, the IVAE affects the perceptual estimates, before it is updated. Similarly, we can model the effect on the response level, based on the not updated IVAE (38). Importantly, the model can produce fully fused responses ($r_A = r_V$) without fully fused percepts ($\hat{x}_A = \hat{x}_V$). Thereby this model can predict suboptimal multisensory reliabilities $\sigma_{AV}^2 > \frac{\sigma_A^2\sigma_V^2}{\sigma_A^2+\sigma_V^2}$. Moreover, it is also possible that we use the updated IVAE (39) after a new observation came in to model the response. In this case of instantaneous updates, the effect of the IVAE on bimodal trials is smaller than for perceptual models. Delayed updates lead to similar effects of the IVAE for response level models compared to perceptual-level models.

Importantly, we use $\hat{\beta}_{k,iVAE}$ not $\hat{\beta}_{k-1,iVAE}$, that means we use the updated IVAE after a new observation came in. This affects especially bimodal trials. During bimodal trials, this model predicts a reduced influence of previous trials, as the IVAE is immediately updated, hence the model predicts a smaller IVAE in bimodal trials compared to unimodal trials. Note that the argument also holds for PRI.

Note that this model of the IVAE is only reasonable when the errorterm is based on the multisensory merged percepts (30). For any other errorterm consider the scenario were the auditory and visual percepts are fused. Then the errorterm based on the merged percept is zero, all other potential errorterms are larger than zero and therefore the auditory stimulus would consistently be perceived on the opposite site of the visual stimulus. To our knowledge this pattern has never been observed so far.

Model Comparison and Averaging with Approximate Bayesian Inference

Model and parameter inference propose a strikingly harder problem for recalibration models than for integration models due the inherently nonstationary nature of the underlying process. As all our models are based on the CI-model for which no closed form solution for the log-likelihood exists, there is no closed form solution for the log-likelihood of the recalibration models as well. Due to the non-stationarity, standard Monte-Carlo approaches are not applicable to recalibration models. Previous studies either used simpler integration models (Bosen et al., 2018) so that closed form solutions for the log-likelihood exist, or simply did not fit data during the recalibration phase where the log-likelihood is non-stationary (Badde et al., 2020; Hong et al., 2021; Sato et al., 2007). In the former case, it is obviously not possible to make use of the most widely accepted model of audio-visual integration: the CI-model. In the latter case most information from the recalibration phase is ignored. This information is especially important to investigate the interaction between recalibration and integration. If there is interdependence between recalibration and integration, all parameter estimates of integration must be backed up by data. More specifically, the errorterms as well as the posterior probability of a common cause must be estimated on a trial-wise base and accordingly also be backed up by actual empirical data.

To overcome these limitations, we used a Bayesian Inference paradigm that does not depend on an explicit formulation of likelihood, which is referred to as approximate Bayesian computation (ABC) (Beaumont, 2019). The fundamental idea of ABC methods is to simulate data for a given model and parameter set and calculate a distance between simulated data and real data as a replacement for the likelihood. If the distance falls under a predefined threshold (rejection sampling), the parameter set is accepted as a posterior sample for the model. Thereby samples can be iteratively created to approximate the posterior probability (Beaumont et al., 2002). A drawback of this procedure is that sampling can become inefficient for high dimensional parameter spaces or when the posterior is narrow compared to the prior (Beaumont et al., 2009; Sisson et al., 2007). ABC methods based on sequential Monte Carlo sampling (ABC-SMC) can overcome this limitation by gradually shifting sampling from the prior to posterior via importance sampling (Beaumont et al., 2002, 2008).

Moreover, ABC-SMC sampler on the joint model space (Beaumont, 2019; Marin et al., 2012; Toni et al., 2012; Toni & Stumpf, 2010) can approximate the relative model evidence, allowing standard Bayesian model comparison analysis techniques. Estimating the model evidence for each of participant as well as each model allows to use random effects Bayesian model selection for group studies (Stephan et al., 2009). Hereby, it is assumed that the true model is not fixed across participant but rather a random factor. This has several advantages over the global bayes factor, which is simply the product of participant-wise bayes factors. Most importantly, since we use ABC-SMC participant-wise estimates of model evidence can become zero quite frequently. If the model evidence is zero for a single participant, the global model evidence also becomes zero, albeit this model might explain best the data of all other participants. Random effects Bayesian model selection does not suffer from this problem since it is based on protected exceedance probabilities (*PEP*) which quantify the probability that one model is more frequent in the sample than any other model, accounting for the possibility that frequency differences might be due to chance (Rigoux et al., 2014).

Importantly random effects Bayesian model selection can also account for the factorial design of our model taxonomy. In fact, *PEPs* cannot only be calculated for individual models but also for a specific level of a model factor. This is accomplished by aggregating evidence over all models that share the same level in the factor of interest. Such a collection of models is referred to as model family. Naturally, factorial model taxonomies lead to large numbers of models, of which many might not be distinguishable (Ma, 2018). Even if a best fitting model does not exist, family-wise model comparisons might still allow to make inference about levels of some model factors. Moreover, we can make inferences about distinct model factors in distinct studies. Assuming that a study design is well suited to identify the level for a factor A but not for factor B, we can still account for the fact that we do not know the true level of B by marginalizing over all possible levels of B. In a second study we can analogously try to identify the level of factor B. Thereby one avoids wrong generalizations that would have seem reasonable if only certain candidate models were tested that differed with respect to factor A in Study 1 and analogously for factor B (see van den Berg et al., 2014 for an example).

Summary

The proposed model taxonomy (see Figure 5 for an overview) allows to classify most modelling approaches in the literature in a systematic way. This thorough classification can serve for rigorous model comparison studies since alternative models are directly implied by the taxonomy and a hand selection of concurrent models, which might result in straw man alternative models (Bowers & Davis, 2012; Jones & Love, 2011; Wilson & Collins, 2019), can be avoided.

The model taxonomy spans five model factors. First, we consider which errorterms derived from the CI-model could serve for recalibration. Second, we dissociate whether the learning mechanism of recalibration is sensitive to sensory uncertainties or not. Third, we address if recalibration is modulated by the posterior probability of a common cause. Fourth, the processing stage of the IVAE is dissociated. Fifth, the time of the update of the IVAE is considered. All combinations of factor levels are realized. The taxonomy is further extended by reliability-based adaptation and pooled recalibration and integration. In total the taxonomy contains 18 CVAE models and 30 IVAE models. The combination of SMC-ABC and random effects Bayesian model selection allows to test the whole set of models across several studies and identify most likely levels for the model factors.

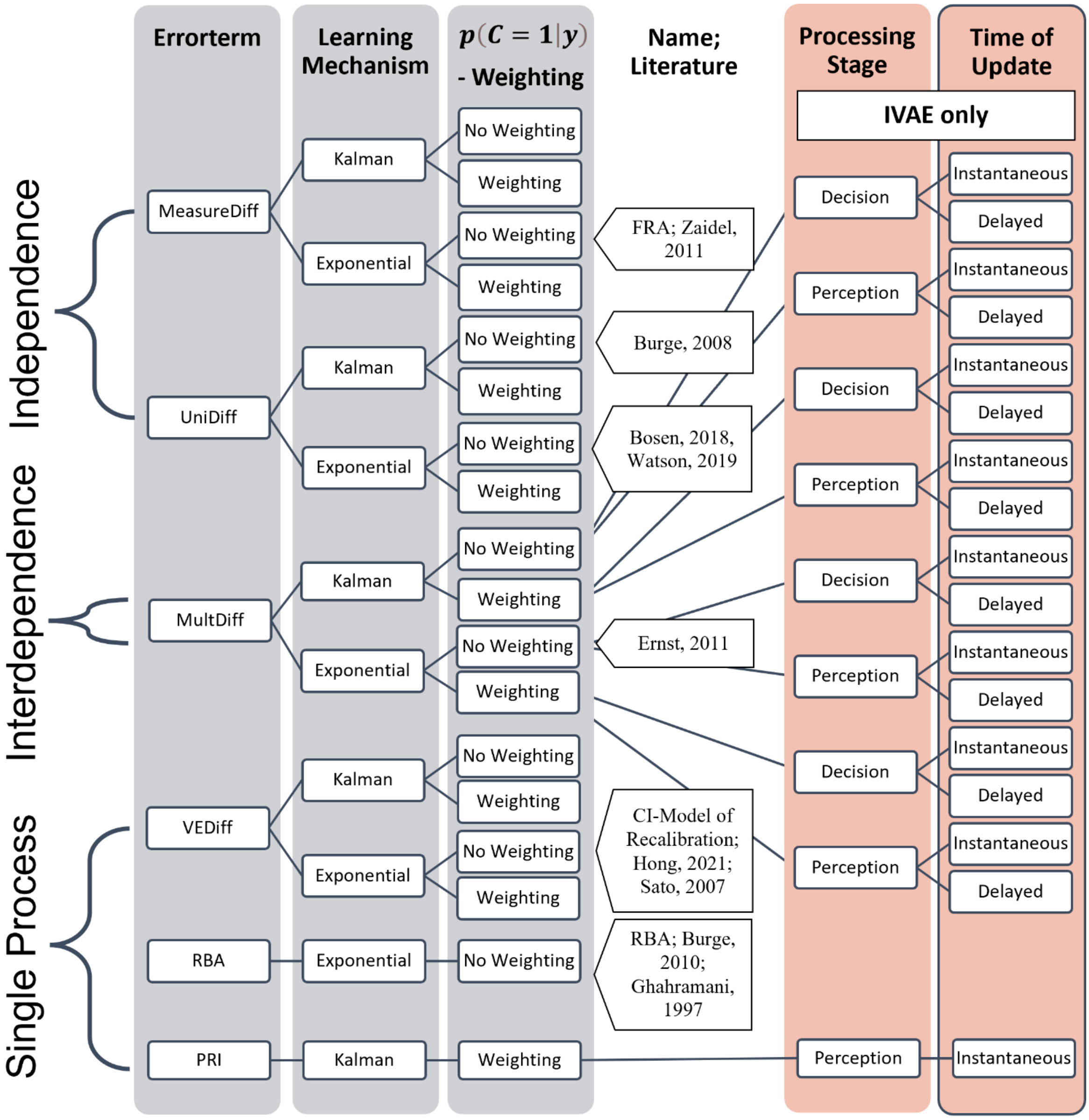


Figure 5. Overview of the model taxonomy for CVAE and IVAE models. All levels of Errorterm, Learning Mechanism and Weighting were realized for the IVAE as well as the CVAE (grey background). For the IVAE additionally the levels of Processing Stage and Time of Update were realized, but only for the MultDiff errorterm.

Chapter II
A Factorial Model Taxonomy for Multisensory Recalibration

Chapter III - Study1

Cumulative and Instantaneous Recalibration Mechanisms are Sensitive to Sensory Uncertainties and Differentially Relate to Integration

Chapter III - Study1
Cumulative and Instantaneous Recalibration Mechanisms are Sensitive to Sensory Uncertainties and Differentially Relate to Integration

Introduction

The ability to combine auditory and visual input into representations of distal objects (Welch & Warren, 1980) or to shift the gaze in the direction of a sound that is out of sight (Frens & Van Opstal, 1995) are remarkable features of our perceptual system, given the underlying computational challenges. Imagine the situation of an approaching car on a street. The perceptual system can form estimates of the car position solely based on either the sound or the visual input. If there is only one car and the auditory and visual spatial estimates align, the perceptual system could easily merge the unisensory percepts (the visual image of the car and for instance the noise of the motor) into a coherent multisensory percept (an approaching noisy car).

However, in general auditory and visual estimates are not necessarily aligned due to independent sources of uncertainty. Firstly, the perceptual system cannot know a priori whether a honk and the retinal image of the car belong to the same distal object; for example, a car outside of the visual field might have honked. Secondly, the input itself is always noisy. To minimize noise, the perceptual system calculates a weighted average of visual and auditory cues, whereby the weights are inversely proportional to the uncertainty in each cue. Moreover, to avoid integration when two causes are likely, the cues are not always fully fused. Rather the magnitude of integration is proportional to the probability that both cues belong to one source. Since visual reliability (the inverse of cue uncertainty) usually vastly exceeds auditory reliability, a shift of auditory spatial perception towards the visual input is often observed, the auditory Ventriloquism Effect (aVE).

While noise leads to unsystematic spatial discrepancies between vision and audition, auditory and visual spatial representations can also be systematically distorted. For instance, the perceptual space can be biased towards the center, periphery (Bruns et al., 2020; Odegaard et al., 2015b; Zwiers et al., 2003) or constantly in one direction (e.g. Knudsen & Knudsen, 1989; Radeau & Bertelson, 1974). If either vision or audition is distorted, this leads to systematic audio-visual discrepancy. These discrepancies vary dynamically as a function of space for central or periphery biases and are well described by linear mappings between physical and perceptual spaces (Hong et al., 2021; Odegaard et al., 2015b; Shinn-Cunningham, 2000). When audio-visual stimuli are presented with a fixed discrepancy, a constant shift of auditory perception in the direction of the discrepancy is induced, which is referred to as the auditory Ventriloquism Aftereffect (Radeau & Bertelson, 1974; Recanzone, 1998).

Several studies (Bosen et al., 2018; Bruns & Röder, 2015; Park & Kayser, 2019; Watson et al., 2019; Wozny & Shams, 2011a) propose a dissociation between slowly emerging but sustainable shifts of auditory maps, induced by cumulating evidence (the auditory cumulative Ventriloquism Aftereffect, aCVAE) and almost instant but highly transient shifts, which dissipate quickly (the auditory instantaneous Ventriloquism Aftereffect, aIVAE).

The aCVAE is associated with remapping of auditory representations early along the auditory pathway (Bruns, Liebnau, et al., 2011; Recanzone, 1998; Zierul et al., 2017). Those transient changes between the mapping of auditory and visual input are unlikely to be compensated by remapping of early sensory representations as these are assumed to be relatively stable over time. Instead, it is reasonable to assume that the perceptual system needs to accumulate evidence over time for stable perceptual shifts (Bruns & Röder, 2015). In line with this assumption, the IVAE seems to recruit partially distinct neural circuits (Park & Kayser, 2021) and dissipates more rapidly over time even in the absence of counterevidence (Bosen et al., 2018; Watson et al., 2019) in comparison to the CVAE.

While several studies have provided evidence for reliability weighting in the VE paradigm (Alais & Burr, 2004; Battaglia et al., 2003; Meijer et al., 2019) the role of reliability for recalibration remains unclear. On the one hand the perceptual system could use reliability as an indicator for accuracy (Block & Bastian, 2011) and recalibrate each sense inversely proportional to its relative reliability (Reliability-Based-Adaptation, RBA (Burge et al., 2010; Ghahramani et al., 1997)). Alternatively, senses could be recalibrated according to a fixed ratio (Fixed-Ratio-Adaptation, FRA, (Zaidel et al., 2011)). Ideally, the ratio should reflect the relative probabilities of vision and audition to be inaccurate. More recent models of Causal Inference account for dynamical and constant biases during audio-visual integration (Hong et al., 2021; Odegaard et al., 2015b) and explicitly model the ventriloquism aftereffect (Hong et al., 2021; Sato et al., 2007) as part of the Causal Inference process. Whereas Hong et al. (2021) find reliability-dependence for audio-visual recalibration, Rohlf et al. (2021) did find recalibration independent from visual reliability. Hong et al. (2021) argue, that the latter result can be explained by the non-monotonic relation (skewed bell-shaped) between the aCVAE and visual reliability that is implied by their CI-based recalibration model.

The objective of this study was two-fold. First, we wanted to confirm the reliability dependence of the aCVAE in an extensive study (32 participants, 2400 localization trials). More importantly, we systematically investigated the computational principles of aVE, aIVAE and aCVAE based on the model taxonomy proposed in Chapter II. To our knowledge all previous approaches of modelling recalibration are covered by this taxonomy (Table 1) and have never been compared in an exhaustive and systematic manner. We differentiated models primarily based on the errorterm and the recalibration mechanism (Kalman Filtering vs. exponential learning) used. More exploratively we investigated whether $p(C = 1|y_k)$ – Weighting occurs and for the IVAE, at which processing stages (perceptual or response stage) recalibration occurs.

We made use of a behavioral experimental paradigm that allowed to evaluate the aVE, aIVAE and aCVAE in common (Bruns & Röder, 2015). Additionally, we manipulated the visual reliability across two sessions. The reliability changes multisensory percepts and thereby the magnitude of the errorterms. We hypothesized that a low visual reliability leads to a smaller auditory VE and a larger visual VE compared to high visual reliability (Alais & Burr, 2004; Battaglia et al., 2003; Meijer et al., 2019). The aVE, aIVAE and aCVAE seem to recruit partially overlapping but also distinct neural circuitry (Bonath et al., 2007; Park & Kayser, 2019, 2021; Zierul et al., 2017). Yet two independent processes can share the same neural resources for distinct computations and vice versa a computationally integrated network might activate distinct subregions, depending on the actual task performed. With respect to the question whether multisensory integration, immediate and cumulative recalibration are computationally dissociable processes, the errorterms based on multisensory percepts would be an indicator of an interdependence of the aCVAE or aIVAE with the aVE. If the errorterm was based directly on the VE, this would even point towards a single process. Finally, errorterms based on unisensory cues or percepts would speak in favor of completely independent processes. Hence, this manipulation allows to rule out several errorterms of the proposed model taxonomy. Since visual perceptual shifts might occur when the visual reliability is sufficiently lowered, we also tested for visual Ventriloquism Effects (vVE), visual instantaneous aftereffects (vIVAE) and visual cumulative aftereffects (vCVAE). We will use VE, IVAE and CVAE when we refer to effects in the visual and auditory modality and VAE when we refer to both the CVAE and the IVAE.

Additionally, the manipulation of visual reliability allowed to investigate which type of learning might underly recalibration. All previous approaches to model audio-visual recalibration implicitly made use of a simple exponential learner (see Table 1 Chapter II for an overview). Hereby, the learning rate is fixed throughout the whole process. However, an optimal learner (in the Bayesian sense) would weigh the incoming information based on their reliability much like in integration, but also consider the likelihood that a sensory cue is inaccurate. For a linear system the Kalman Filter (Kalman, 1960) provides the optimal learning rate to this problem (See Section "Mechanisms of Recalibration" for a detailed description of the learning algorithms). The Kalman Filter approach extends the idea that recalibration should primarily consider accuracy by the aspect that precise information should lead to faster recalibration than imprecise information (Burge et al., 2008; Di Luca et al., 2009; Körding & Wolpert, 2004). To our knowledge Kalman Filtering has not been considered for audio-visual recalibration so far, especially when the errorterms are derived from Causal Inference. It is important to consider Kalman Filtering as learning mechanism when reliability dependence of recalibration is investigated, since an effect of reliability can either be due to altered multisensory percepts (Hong et al., 2021) or altered learning rates throughout the recalibration process (Burge et al., 2008; Di Luca et al., 2009) or even both. If IVAE and CVAE are Kalman-like processes, they should be reduced by a decrease of visual reliability.

In a more explorative way, we investigated whether the empirical data is better described by $p(C = 1|y_k)$ - Weighting and whether the IVAE occurs at a post-perceptual processing stage or rather at the perceptual level.

Methods

Participants

In order to counterbalance all conditions (see *Procedure* for details), we aimed for a sample size of 32 participants. Moreover, this sample size yields a power of 0.8 to detect a medium-sized effect (d_z = 0.56) for a directional difference between two within-subject conditions at an α level of .017. Hence, we were able to test our three main behavioral hypotheses, which suggest that the aVE, the aCVAE and the aIVAE are reduced for low reliable visual stimuli at a global α level of .05 (assuming Bonferroni correction). The power analysis was conducted in G*Power 3.1 (Faul et al., 2009).

In an initial run, 32 participants were tested of which we had to exclude 7 participants due to insufficient localization accuracy during the baseline measurements (see *Baseline Measurements and Exclusion Criteria* for a detailed description of the exclusion procedure). The excluded participants were replaced yielding 39 participants in total with 32 remaining for further data analysis.

The remaining 32 participants (18 female, 14 male, mean age = 24.5 years, age range = 18-43 years; 30 right-handed) all reported normal or corrected to normal vision, no history of visual, auditory or neurological impairments and did not use any medication known to affect perception. The former information was collected via a questionnaire.

All participants were recruited through an online subject pool of the University of Hamburg. Written informed consent was obtained from all participants prior to taking part. The study was performed in accordance with the ethical standards laid down in the Declaration of Helsinki (revised from 2013). The procedure was approved by the ethics commission of the Faculty of Psychology and Human Movement of the University of Hamburg.

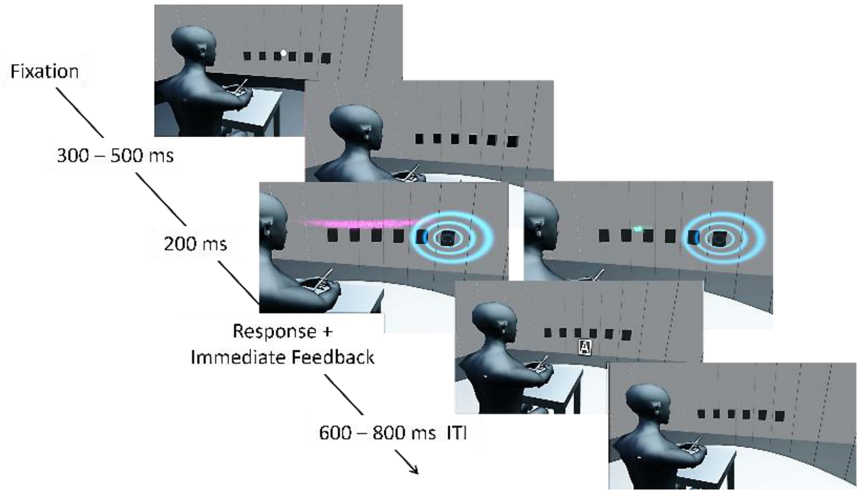


Figure 6. Localization trial. Participants initiated a trial by pointing towards a fixation point, subsequently a visual, auditory or audio-visual stimulus was presented (the figure shows an audio-visual trial), followed by a letter (A for auditory, V for visual) indicating whether to localize the auditory (symbolized by a blue wave pattern) or visual component (symbolized by a magenta random dot pattern). Avatar image adapted from "Low Poly Character" by TehJoran, 2011 (https://www.blendswap.com/blend/3408) licensed under CC BY.

Apparatus

The study was conducted in a sound-attenuated darkened room. Auditory stimuli were presented with six speakers which were mounted on a semicircular frame (90 cm radius) and covered by an acoustically transparent curtain. Participants were seated in the center of the frame and positioned their head on a chin rest at the level of the speakers. The speaker positions ranged horizontally from 22.5° left from straight-ahead (0°) to 22.5° right from straight-ahead in steps of 9° (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). A schematic illustration of the apparatus is shown in Figure 6. Visual stimuli were presented via RGB-LED-panels (APA 102, Shiji Lighting, Shenzhen, China) measuring 32 cm in width and 8 cm in height with a pixel width of 0.5 cm and a spacing of 0.5 cm (2.54 ppi). Overall, eight LED-panels were installed in a semicircular arrangement (84 cm radius) ranging from -87° to + 87° horizontally and covering 8 x 256 individual LED locations. The upper edge of the panels was adjusted to align with the bottom of the speakers in order to position them as close as possible to the speakers but nevertheless avoid acoustic shadows. An Arduino Due (Arduino SRL, Strambino, Italy) was used to interface between the experimental computer and the LED-panel. The content of the LED-panel could be updated similarly to a standard display by sending a whole frame to the panel. The frame duration could be varied individually.

Stimuli

The auditory stimuli were narrow-band filtered (1/3 octave) pink noise bursts with four different center frequencies (445, 890, 2000, or 4000 Hz) and were presented for 200 ms including 5 ms on- and off-ramps. The spacing of the center frequencies was chosen to assure, that at least one critical frequency bandwidth (Zwicker et al., 1957) lies between the borders of the frequency spectra of all tones. The stimulus intensity was randomly varied over a 4-dB range, centred at 70 dB(A) to minimize potential differences in the speaker transformation functions.

For visual stimulation in each trial, a monochrome random dot pattern was created across the whole 8x256 pixel LED-array, i.e., a random value between 0 and 1 was assigned to each pixel. To this random dot pattern, we applied a 2D gaussian amplitude envelope, whereby the mean of the gaussian envelope was taken as the position of the visual stimulus. The vertical standard deviation of the envelope was a 1.1° visual angle for all experimental conditions. For high reliable visual stimuli, we used a horizontal standard deviation of 2.5° visual angle whereas for the low reliable visual stimuli we used 21.6° visual angle. The resulting amplitude values were divided by four times the maximum amplitude value across the whole array, as the full amplitude range of the LED-panels was too bright in the darkened room and led to distracting afterimages.

We used four different colors (45°~orange , 135 °~ green, 225°~blue, 315°~magenta) defined in the plane spanned by the L-M and S-(L + M) axis of the DKL-Color-Space (Derrington et al., 1984). The choice of colors ensured that none of the colors solely stimulated the s-cone-channel and therefore might bypass the superior colliculus. The superior colliculus has been shown to be important for multisensory spatial orienting (Leo et al., 2008). The RGB-values (ranging from 0 to 255) of the colors were adjusted individually to provide perceptually isoluminant colors (see *Luminance Adjustment* for a detailed description). Finally, for each pixel we multiplied the RGB-values with the amplitude value for the corresponding pixel. In the case that any of the resulting RGB-values were below 1, we turned this pixel of, as this would have led to a selective shut-off of one of the three colored LEDs of that pixel, resulting in a distorted color.

Moreover, we used a fifth visual stimulus during baseline measurement that consisted of a white vertical bar (RGB-Value) with one pixel width (0.64° visual angle) and eight pixels height (5.09° visual angle). This stimulus (response noise stimulus) was used to estimate any noise in the pointing responses that was not due to perceptual noise (i.e. non-perceptual noise, NPN, see Tassinari, Hudson, & Landy (2006) for a similar procedure).

To indicate to the participants whether to localize the visual or the auditory component of the trial, another LED-panel was attached centrally, right below the LED-array for stimulus presentation. The letter “A” (6x6 pixel) surrounded by a square outline (8x8 pixel) in white (RGB-value) indicated to localize the auditory component of the trial. A black letter “V” (6x6 pixel) surrounded by a white (RGB-value) filled circle (4 pixels radius) indicated to localize the visual component of the trial. A custom-built pointing stick with a spatial resolution of 1° was used to collect localization responses of the participants.

Procedure

The study was split into three sessions on separate but not necessarily consecutive days, each session lasting about 3h. The first session was used to obtain individually adjusted isoluminant colors for subsequent visual stimulation in sessions 2 and 3. In addition, we assessed unimodal visual and auditory localization baseline accuracy and reliability. On the one hand this allowed us to check whether the manipulation of the visual reliability succeeded. On the other hand, we were able to estimate baseline distortion of auditory and visual perception, for instance biases to the left or right, as well as centrality or eccentricity biases. Sessions 2 and 3 were used to induce the actual experimental manipulations in intermixed blocks that contained unimodal visual and auditory trials as well as bimodal audio-visual trials. A schematic overview of the study design is given in Figure 7. We varied the reliability of the visual stimuli (low or high) between sessions. From the four possible colors, two were used for each session. Similarly, we grouped the auditory stimuli in two pairs (445Hz/2000Hz, 890Hz/4000Hz). One pair was used for each of the second and third Session in order to avoid carry-over effects between sessions (Bruns & Röder, 2019). For each session, two fixed audio-visual pairs were formed consisting of one of the colored visual stimuli and one of the auditory stimuli. We assumed that participants would primarily use interaural timing differences (ITD) to localize the 445Hz and 890Hz tones in contrast to the 2000Hz and 4000Hz tones which should primarily be localized via interaural loudness differences (ILD). As we always used one ITD and one ILD sound in each session, this factor was naturally counterbalanced with the condition of interest, i.e., the visual reliability. The same accounts for the audio-visual spatial discrepancy. Moreover, we counterbalanced the color used for the visual stimulus with the visual reliability. The visual reliability was also counterbalanced with session order (second or third session).

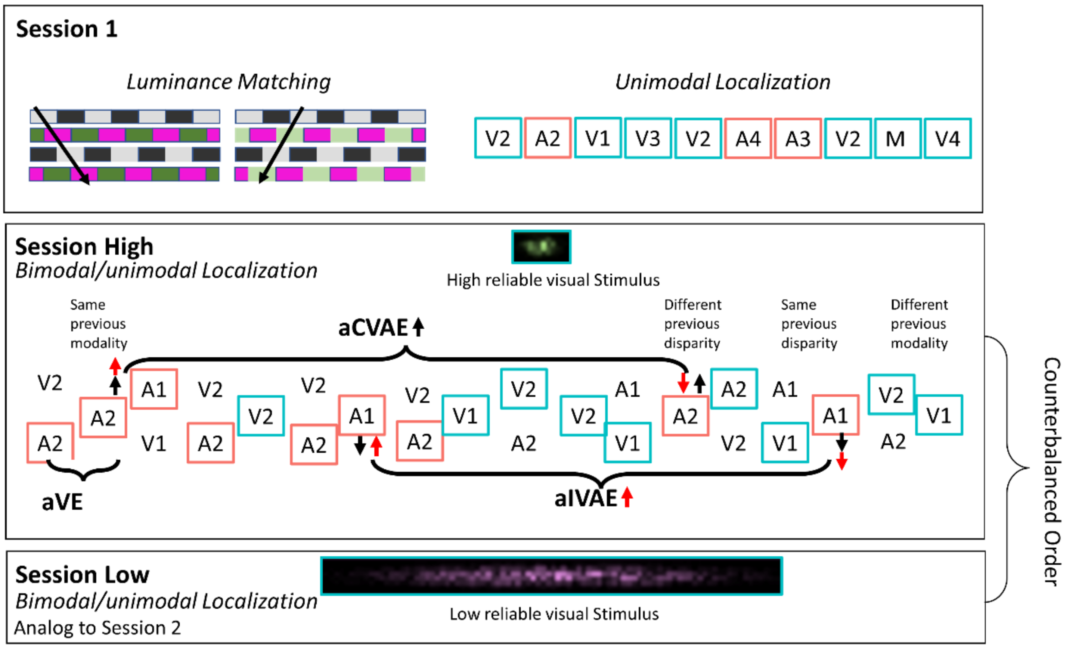

Session 1
Luminance Matching
Unimodal Localization
V2 A2 V1 V3 V2 A4 A3 V2 M V4
Session High
Bimodal/unimodal Localization
High reliable visual Stimulus
Same previous modality
aCVAE
Different previous disparity
Same previous disparity
Different previous modality
V2 A1 V2 V2 V2 V2 A1 A2 A1 V2 V1
A2 A2 V1 A2 V2 A1 V1 V2 A2 A1 A2
A2 A2 A2 V1 V2 V1
aVE
aIVAE
Session Low
Bimodal/unimodal Localization
Analog to Session 2
Low reliable visual Stimulus
Counterbalanced Order

Figure 7. *Experimental procedure.* The figure is chronologically ordered from left to right and top to bottom. The experiment started with a minimum motion paradigm to determine isoluminant colors for the subsequent blocks. Unimodal localization blocks were conducted in order to assess unimodal reliabilities and fit linear models between physical position and perceived position for all stimuli. The second and third session consisted of intermixed blocks were tones of different sound frequencies (A1 and A2) were paired with visual stimuli of different color (V1 and V2). Importantly, each pair (A1V1 and A2V2) was presented with an audio-visual discrepancy in opposite direction inducing a CVAE in opposite directions. On a trial-by-trial basis, the direction of the preceding discrepancy changed allowing to evaluate the IVAE. Shifts in audio-visual trials were induced by VE, CVAE and IVAE in common. Throughout the second and third session the visual reliability was either high for all visual stimuli or low (the assignment of session order was counterbalanced). Parts of this figure has been adapted from Bruns (2019).

First Session

Luminance Adjustment

The first session started with a minimum motion paradigm (Logothetis & Charles, 1990) to obtain perceptually isoluminant colors individually for each participant. We defined magenta (315° in DKL-Color-Space) at -35° luminance as the reference color and adjusted the remaining three colors with respect to magenta. Hence, three minimum motion blocks were performed preceded by a short training block in which magenta was used as reference and adjusted color. The stimuli consisted of horizontal, rectangular (6x64 pixel, horizontally centered at 0 °, i.e., straight ahead) square wave gratings with a period of eight pixels (5.09° visual angle). In each uneven frame, the color of the grating alternated between the reference color and the color to be adjusted, whereas in each even frame the square grating was monochrome (origin of the color-space), contrast altered between -0.5 (dark grey) and 0.5 (light grey) in the direction of the luminance axis (90° in the luminance plane). In each frame the square grating was offset by a quarter period. The high luminance period of the monochrome grating was identified with the reference color period of the colored grating. For each trial the shift was either consistently to the right (rightward trials) or consistently to the left (leftward trials). Hence, if the perceived luminance of the adjusted color was higher than the luminance of the reference color a leftward movement was likely perceived in leftward trials, whereas a rightward movement was likely perceived in rightward trials (Figure 6, B). The initial frame always had zero offset and the colored grating started with the reference color at the first half period (counted from left to right). A frame was presented for 80ms. Overall, a trial consisted of 12 frames. Participants had to indicate via button press whether they perceived a motion to the left or to the right. If no clear motion was perceived participants were instructed to choose intuitively.

For each minimum motion block, the reference color was used, and one of the three colors was adjusted. The order of colors that had to be adjusted was randomized across participants. The luminance of the adjusted color was varied according to a one-up-one-down staircase procedure (Macmillan & Creelman, 2004). Two staircases were run in parallel, randomly intermixed. One staircase starting with a presumably lower luminance value (-60°) the other starting with a presumably higher luminance value (-10°) for the adjusted color compared to the reference color. At the beginning of each trial, the offset direction was chosen randomly (leftward vs. rightward). Leftward responses for leftward offset direction were counted as correct, analogously for rightward offsets and rightward responses. In case of correct responses, the luminance of the adjusted color was reduced whereby the step size started with 6° and was adjusted after each reversal according to the following sequence (6.0, 3.0, 3.0, 1.5, 1.5, 0.75, 0.75, 0.375). After at least 11 reversals in each staircase, the procedure was stopped. The luminance values of the last 11 reversals of both staircases were averaged to calculate the point of subjective equality (PSE) separately for each staircase. When the staircases diverged (PSE difference larger than 15°), the block was repeated up to two times. The four resulting DKL-colors were then transformed into RGB values. Those values were used as maximum amplitude colors across the rest of the experiment for the visual stimuli.

Baseline unimodal Localization

In this experimental part we evaluated the initial accuracy and reliability of localization responses to unimodal visual and auditory stimuli. Nine different stimuli, i.e., four visual stimuli, four auditory stimuli and the response noise stimulus were presented from six positions (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°), 16 times each. In total, this part consisted of 864 trials in randomized order which were split into four equally sized blocks à 216 trials (baseline blocks). Between each block a pause of approximately 5 minutes was introduced.

Each trial started with the presentation of a fixation point at 0° azimuth. Participants had to direct the pointing stick towards the fixation point and press a button attached to the stick to start a trial. Only if the pointing direction did not deviate more than 10° from the direction of the fixation point the trial started. After a random delay (400ms to 600ms) the experimental stimuli were presented for 200ms, and a letter appeared (A for auditory experimental stimuli and V for visual experimental stimuli). Responses were allowed immediately after the letter appeared. Hence participants did not know beforehand whether the auditory or visual component was task-relevant and had to remember both. Thereby attendance to only one of the senses is avoided, which is known to attenuate integration and recalibration (Badde et al., 2020). Participants were instructed to respond as accurate as possible but nevertheless prompt by directing the pointing stick to the location where they had perceived the stimulus and confirm by button press. Moreover, we informed participants that all stimuli would be presented at the same height and that they should focus on horizontal pointing accuracy. The procedure diverged for response noise stimuli in so far as the vertical bar was displayed persistently until the participant confirmed the response.

Second and Third Session

During this experimental part we induced the VE, CVAE and IVAE and quantified these effects by intermixing bimodal and unimodal trials and measuring localization responses (intermixed blocks). We made use of an adapted version of the paradigm of Bruns & Röder (2015), which relies on the assumption that the CVAE is frequency-specific and therefore could be induced in different directions for different tones. Two audio-visual stimulus pairs were used across this part and a fixed audio-visual discrepancy to the left (-13.5°) was assigned to one pair (e.g., A1V1), whereas a discrepancy to right (13.5°) was assigned to the other pair (e.g., A2V2). The audio-visual discrepancy was defined as the relative displacement of the visual stimulus to the auditory stimulus. Hence, VEs, CVAE and IVAE in the according directions were induced during bimodal stimulation. Auditory positions were equal to the positions used in baseline blocks (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°), visual positions however were shifted on the audio-visual disparity (-9°, 0°, 9°, 18°, 27°, 36° for rightward and -36°, -27°, -18°, -9°, 0°, 9° for leftward). In addition, all four unimodal components (A1, A2, V1, V2) of the A1V1 and A2V2 were presented as unimodal trials (analogously to the baseline blocks) intermixed with the bimodal trials (see Figure 7). Positions in unimodal trials were equal to the positions used in baseline blocks (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). The procedure of a bimodal or unimodal trial was the same as in the baseline blocks. However, whereas for unimodal trials it was unambiguous which stimulus to localize, during bimodal trials participants could either localize the visual stimulus or the auditory stimulus. Therefore, we instructed participants to memorize both positions until a letter appeared (A for auditory, V for visual) that indicated whether to point to the auditory stimulus (Audition task-relevant) or to the visual stimulus (Vision task-relevant). We presented each of the two bimodal stimuli 32 times for each auditory position, i.e., 192 times in total. In half of the bimodal trials (96 per stimulus) vision was task-relevant whereas in the other half audition was task-relevant. Each unimodal stimulus was presented 16 times per position, i.e., 96 times in total. The order of the trials was pseudo randomized in a way that assured an equal number of unimodal trials (4) where the one, two, three or four preceding audio-visual trials had the same unimodal component or the different unimodal component, resulting in eight types of possible sequences of differing length. Moreover, for each of the preceding sequences in half of these vision was task-relevant during bimodal trials and audition in the other half. This design allowed to estimate the CVAE and IVAE in common based on localization response in unimodal trials (see Bruns & Röder, 2015 for details). Overall, this part contained 768 trials per session. These trials were split into two blocks of approximately equal length. The length differed slightly as pauses were set between sequences in order to not split them across two blocks.

Modelling Methods

Competing Models of Integration

Overall, we compared one family of models. The models differed with respect to the weights used for integration i.e., models with optimal weighting (optimal), suboptimal estimation of visual reliability and suboptimal weighting.

Model Parameter Integration

The responses for each audio-visual pair (A1V1, A2V2) for each visual reliability condition (high vs. low) were modelled with a separate mixture filter $\boldsymbol{MF}_{\text{AVp,VRq}}$. We use AVp,VRq with p in {1,2} and q in {high, low} as an index for pair p in reliability condition q. We only write A or V when referring to unimodal components of the audio-visual pairs. Hence, we had to estimate eight variances σ_{mpq}^2 (m in {A,V }) to calculate **R** for each pair ($\boldsymbol{R}_{\text{AV1,VRhigh}}$, $\boldsymbol{R}_{\text{AV2,VRhigh}}$, $\boldsymbol{R}_{\text{AV1,VRlow}}$, $\boldsymbol{R}_{\text{AV2,VRlow}}$). We fitted linear models to the baseline data of each unimodal stimulus and the non-perceptual noise stimulus. Then we calculated the variance of the residuals as $\hat{\sigma}_{res,\ mpq}^2$ and $\hat{\sigma}_{NP}^2$. The latter was directly used as an estimate for σ_{NP}^2across all filters $\boldsymbol{MF}_{\text{AVp,VRq}}$, assuming that the perceptual noise in this condition was neglectable and that the non-perceptual noise did not differ across all stimuli. If $\hat{\sigma}_{NP}^2$ exceeded one of the sensory variances $\hat{\sigma}_{res,\ mpq}^2$, we set $\hat{\sigma}_{NP}^2$ = Min($\hat{\sigma}_{res,\ mpq}^2$) – 0.46, whereby 0.46 is the squared angular size of a single pixel of the LED-matrix, which we assumed to be a reasonable lower bound for sensory resolution. In a further step we subtracted $\hat{\sigma}_{NP}^2$ from $\hat{\sigma}_{res,\ pq}^2$ to correct for the non-perceptual noise and divided by $\hat{h}_{pq}(m,m)$ to correct for the observation model (see *Approximation of R and H*). Note that $\hat{h}_{pq}(m,m)$ is a parameter that corresponds to the element in the mth row and column of $\boldsymbol{H}_{\text{AVpVRq}}$.

$$\hat{\sigma}_{mp}^2 = \left(\hat{\sigma}_{res,\ mp}^2 - \hat{\sigma}_{NP}^2\right)\hat{h}^2{}_{mp} \tag{40}$$

The diagonal elements $\hat{h}_{pq}(m,m)$ of $\boldsymbol{H}_{\text{AVpVRq}}$ were free parameter. The weighting factors $\omega_{SW,V}$ and $\omega_{SE,V}$ were free parameter for the suboptimal weighting and suboptimal estimation model and set to one for the optimal weighting model. We also allowed auditory overweighting or overestimation ($\omega_{SW,V} > 1, \omega_{SE,V} > 1$). As we centered the data block-wise for modelling integration we set $\hat{\beta}_{AVp,VRq,k=0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We set $\hat{\sigma}_q = 45$, note that although this choice is arbitrary it does not affect the model fits and predictions. If σ_q is chosen large enough, there exists always an observation model H to assure that the predicted responses follow the same linear trend (as a function of physical position) as the empirical data. Finally, the prior probability of a common cause $p(C = 1)$ was a free parameter in all models.

Bias Correction

Equation (21) allows to estimate the constant biases on grouped data. The upper bar indicates the average over trials.

$$\widehat{\boldsymbol{K}}_{k,C=2}{}^{-1}\bar{r} - \bar{y} \quad \approx \hat{\beta} \tag{41}$$

We calculated these estimates block-wise for each participant, session and stimulus. This allowed to fit models for integration and immediate recalibration without explicitly modelling cumulative recalibration. Note, that we had an overall of 3 integration models, 18 models for cumulative integration and 34r models for immediate recalibration. Modelling all combinations would yield 1.836 competing models which is not feasible.

Competing Models of Recalibration

In total, we had three families of models for recalibration. The first model factor differed with respect to the used errorterm ($\hat{e}_{UniDiff}$, $\hat{e}_{MeasureDiff}$, $\hat{e}_{VEDiff}$, $\hat{e}_{MultDiff}$) and thereby had four levels. The second factor differed with respect to recalibration mechanism (Kalman vs. Exponential) and the last factor differed whether posterior weighting was used $((C = 1|y_k) - Weighting)$ or not (No Weighting). Additionally, we added PRI model and the RBA model, the latter refers to the combination $\hat{e}_{VEk}$, No Weighting and Exponential learning but with the constraint $\sigma^2_{\beta Aleft} = \sigma^2_{\beta Aright} = \sigma^2_{\beta V}$, i.e., one supramodal learning rate. All combinations of model factor levels were realized, resulting in 18 competing models for cumulative recalibration.

Competing Models of Instantaneous Recalibration

We used the same model factors for immediate recalibration as for cumulative recalibration for all perceptual IVAE models (except for the RBA model). Moreover, we added eight models where the IVAE occurs on the response level. All the latter models made use of the MultDiff errorterm and differed with respect to timepoint of the IVAE update (instant vs. delayed) additionally to the same factors as for the perceptual level IVAE models (posterior weighting, learning mechanism). Overall, we tested 24 models.

Model Parameter Recalibration

Additionally, to the parameter used for the corresponding integration model, we introduced six parameters. First, the elements of $\widehat{P}_{\beta}$ (the spatial prior for the auditory and visual biases) were free parameter. As we observed generally larger auditory aftereffects to the right, we defined $\hat{P}_{\beta}$ separately for the audio-visual pairs based on, whether the audio-visual discrepancy was to the left ($\hat{P}_{\beta left}$) or to the right ($\hat{P}_{\beta right}$). This resulted in three different parameters $\sigma^2_{\beta Aleft}$, $\sigma^2_{\beta Aright}$ and $\sigma^2_{\beta V}$. Importantly we used the same variances for spatial bias priors across session, i.e., regardless of whether the visual reliability was high or low. Finally, the transfer factor τ_A^{CVAE} was a free parameter, whereas τ_V^{CVAE} was set to zero as we did not observe any consistent visual aftereffects across participants. To account for potential generalization across frequencies for the IVAE (Bruns & Röder, 2015), we set $\tau_V^{IVAE} = \tau_A^{IVAE} = 1$, indicating that the IVAE transfers completely across stimuli within one sensory modality but not across. In addition to the CVAE models, the IVAE models further included the decay factor d_{IVAE} as free parameter.

Model-free Analysis

Luminance Adjustment Block

Minimum motion blocks were analyzed immediately after completion to derive isoluminat colors that were used throughout the rest of the experiment. For each color two staircases were presented intermixed and stopped, when at least 11 reversals had occurred in each of them. The first 5 reversals from each staircase were discarded. The PSE was estimated as the average of all remaining reversals across staircases of the same color.

Baseline Measurements and Exclusion Criteria

For all experimentally relevant unimodal stimuli we fitted separate linear models to the localization responses with the actual stimulus positions as predictor. The obtained estimates of slope and intercept served as indicators of localization accuracy with a slope of one and an intercept of zero indicating perfect accuracy. Unimodal reliabilities were assessed for each unimodal stimulus separately based on the trial-wise residuals ($Res_{\ Base}$), i.e., the difference between empirical response and the response predicted by the linear model. After the initial sample of 32 participants was collected, we calculated the sample mean and standard deviation for the individual slopes and intercepts. Participants whose intercept or slope differed 2.5 times the standard deviation from the sample mean were excluded from further analysis. Overall, 7 participants were excluded due to these criteria. We refilled the sample with 7 additional participants and repeated the outlier detection procedure based on the criteria from the initial sample. In case any of the new 7 participants would not have met the criteria we would have iteratively repeated this procedure until a full sample of 32 datasets would have been acquired. However, no more participants were excluded after the second sampling.

Intermixed Block - Quantification of VE

First, we again fitted linear models separately for each unimodal stimulus to the localization responses with veridical stimulus position as predictor. From these linear models we predicted localization responses for the responses in bimodal trials. Residuals were obtained by subtracting the predicted responses from the actual responses in each bimodal trial. For each response modality in bimodal trials, we calculated the mean of the residuals as a measure for the VE (Res_{VE}). This was done for two reasons. First, not only the VE shifts the percepts of the auditory and visual component in a trial consistently towards each other but also the IVAE and CVAE. However, as both aftereffects also manifest in the unimodal trials as consistent shift, subtracting the predicted response (based on unimodal trials) from bimodal responses also subtracts the estimated aftereffects. Secondly, we did not use the same visual position during bimodal and unimodal trials for visual stimuli. As the VE mainly manifests as shift in the localization responses form unimodal presentation to bimodal presentation, we had to predict unimodal responses at the locations we used during bimodal stimulation.

The VE was statistically assessed based on the trial-wise residuals with a linear mixed model (Bates et al., 2015).

$$Res_{VE} \sim disparity * visual\ reliability * previous\ disparity + (1 \mid participant) \tag{42}$$

Quantification of IVAE and CVAE

To quantify the CVAE and IVAE we first predicted individual localization responses for each unimodal trial in the intermixed blocks based on the linear model derived from the baseline measurement. The residuals (Res_{VAE} , we use VAE since they were used for both aftereffects) were calculated as the difference of the predicted responses and the actual responses given by the participants. Hence, these residuals were corrected for any biases (intercepts different from zero and slopes different from one) present at baseline. These trial-wise residuals were analyzed with two linear mixed models, one including only unimodal auditory trials, the other including only visual trials. Fixed effects were visual reliability (low or high), previous disparity (same or different) and previous task (audition or vision).

The IVAE appears as a modulation of shifts in unimodal trials depending on the disparity in the previous bimodal trial. For a stimulus pair, e.g. A1V1, the average shift is in direction of the audio-visual disparity in bimodal A1V1 trials due to the CVAE. The difference between A1 trials in which the previous trial included the A1V1 pair and thereby a disparity in the same direction as the expected CVAE (previous disparity: same) and A1 trials in which the previous trial included the A2V2 pair and thereby a disparity in a different direction as the expected CVAE (previous disparity: different) serves as a measure of the IVAE (Figure 6, B). With respect to the direction of shifts opposite results are expected for A2, hence the IVAE should result in an interaction of disparity and previous disparity.

The main reason to include task-relevancy as a manipulation was to avoid an attenuation of integration and recalibration by modality-specific attention. However, the task-relevant stimulus component might have exerted an increased influence on subsequent trials via memory related processes (Park & Kayser, 2020), which is why we included this factor in the analysis. Moreover, we allowed random intercepts for each participant nested within participants, the final LMM is given by:

$$Res_{\,VAE} \sim disparity * visual\ reliability * previous\ disparity + (1 \mid participant) \tag{43}$$

All LMMs were analyzed with type II ANOVAs using Wald chi-square tests (Fox & Weisberg, 2019).

Model-based Analysis

Data Preparation

The same data preprocessing steps as for the behavioral data analysis were performed in advance of the model-based analysis. However, with did not exclude outliers from the dataset but simply labelled them as outliers. Due to the sequential nature of our study design we had to include these trials to properly simulate trial-by-trial sequential effects. Importantly we did not use the responses in those trials to compute model-fit measures. We pooled the data of all three session and then split this dataset for each participant based on whether the auditory and visual stimuli were used in the visual reliability high or low condition. Hence one of these two datasets included all trials of the second session plus all visual and auditory baseline trials that had the same color or pitch respectively as used in the second session (analogously for the third session). Hereby, the relative order of the trials was preserved, and we assumed that the influence of acquiring baseline data not at the same day as intermixed data is neglectable.

For IVAE models and integration models, we corrected these two datasets block-wise for the CVAE. This was done separately for each unimodal stimulus based on the responses in unimodal trials in each block following the procedure in *Bias Correction* in Chapter II. We treated the baseline session as one block, as we did not expect any constant biases to vary across this session. Otherwise, we preserved the block indices for intermixed blocks.

Inference Scheme

Since no analytical solution for the likelihood function for any of the considered models exists and Monte-Carlo methods are too computationally expensive, we used an approximate Bayesian Inference paradigm that does not depend on an explicit formulation of likelihood and approximates the Bayesian evidence via simulations (see Chapter II for a detailed explanation).

For the distance measures we partitioned the data based on the experimental conditions disparity, visual reliability, the type of stimuli (bimodal or unimodal), the task-relevant modality (vision or audition), the physical position of the task-relevant stimulus component (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) and whether the trial occurred in the first or second half of a given experimental block. For parameter inference with respect to the IVAE we added the factor previous disparity. For each of the combinations of these factors we split the data randomly and used 3/8 of the data for the prior calculation (prior data) and the remaining 5/8 for model comparison (model data). We used the Wasserstein-Distance for each partition as distance measure between observed and simulated data.

For all effects (VE, IVAE, CVAE) we used a three-step inference scheme. The marginal likelihood is highly sensitive to the choice of priors, therefore we decided to use partial bayes factors to remove any unspecified constant in the marginal likelihood due to the use of weakly informative priors (T. O. Berger & Pericchi, 2004). First, from the population priors (see *Priors*) we inferred individual priors for each participant and model based on the prior data via an ABC-SMC method with quantile based threshold (Beaumont et al., 2002). The initial quantile was set to 0.1 and all following to 0.5. Second, we set an acceptance ratio of 0.0025 or maximum number of 40 iterations (Batchsize = 256000) as stopping criterion to ensure comparable runtimes across models and participants. Each run consisted of 8000 samples.

Importantly all trials were used to simulate datasets but only responses in trials from the prior data were used to calculate distances.

Algorithm 1

Model Comparison ABC-SMC sampler

```
Initialize $t = 1$ and $t_{max}$, i = 1, ..., $N_s$, k= 1, ..., $N_m$ and $q_{init}$

At iteration t = 1

1.  Sample $m_i \sim \pi(m)$
2.  Sample $s_i \sim \pi(\theta)$
3.  Simulate sample $s_i$ from generative model $m_i$
4.  Set $\omega_{t,i} = {}^{1}/_{N_s}$
5.  Repeat 1-4 until ${}^{N_s}/_{q_{init}}$ initial samples are obtained and
6.  Set $\in_{t+1}$ $as\ q_{init}\ quantile\ of\ \|s_i, s_{observed}\|$ of t and $S_t = \{s_i;\ \|s_i, s_{observed}\| \leq \in_{t+1}\}$
7.  Set t=t+1

At iteration t > 1

1.  Sample $m_i \sim q_t\ (m)$,
        $q_t(m_i) = \left(\sum_{j\in\{j; m_i=m_j\ for\ s_j \in S_t - 1\}} \omega_{t-1,j}\right) \Big/ \left(\sum_{j=1}^{N_s} \omega_{t-1,j}\right)$
2.  Sample $\theta_i \sim q_t\ (\theta|m = m_i)$,
        $q_t\ (\theta|m = m_i) = q_i(m_i) K_t(\theta_t|\theta_{t-1}, m_i\ )$
3.  Set $\omega_{t,i} = {}^{\pi(m_i)\pi(\theta_i)}\big/_{q_t\ (\theta_i|m = m_i)}$
3.  Simulate sample $s_i$, if $\|s_i, s_{observed}\| \leq \in_t$ accept $s_i$
4.  Repeat 1-3 until $N_s$ samples are obtained
5.  Set $K_t(\theta_t|\theta_{t-1}, m_i\ ) = N(\theta_{t-1}, 2\Sigma_{t-1,m_i})$, with $\Sigma_{t-1,m_i}$ being the weighted covariance from all samples $s_j$ with $m_i = m_j$
7.  Set t=t+1; repeat until t $> t_{max}$
```

In a second step we ran an ABC-SMC sampler on the joint model space (Beaumont, 2019; Marin et al., 2012; Toni & Stumpf, 2010), i.e. we treated the type of model as a discrete parameter (**Algorithm 1**). The posterior probability of a model m_i can then be estimated based on the last generation of Samples S_{tmax} as $\tilde{p}(m_i|s_{observed}) = \sum_{j\in\{j;m_i=m_j\ for\ s_j\in S_{tmax}\}} \omega_{tmax,j}$. Importantly, if a model is a poor description of the data, no samples for this model might remain in the final sample generation and the approximated log probability becomes infinite. For numerical stability we therefore set the final estimate of the posterior model probability to $\hat{p}(m_i|s_{observed}) = min\ (\tilde{p}(m_i|s_{observed}), 10^{-5})/\sum_{i=1\ldots\mathrm{N_m}} min\ (\tilde{p}(m_i|s_{observed}), 10^{-5})$. All quantiles were set to 0.5. We set an acceptance ratio of 0.0025 or maximum number of 1400 iterations (Batchsize = 256000) as stopping criterion to ensure comparable runtimes across models and participants. The number of samples was set to 64000 for the VE and 128000 for the IVAE and CVAE. Importantly, we first identified the best VE model on the group level and used this model as integration model for the IVAE and CVAE run.

Based on the estimates of the model probabilities for each participant we performed a Bayesian model selection analysis on the group level and calculated protected expected exceedance probabilities (Rigoux et al., 2014; Stephan et al., 2009). We repeated these two steps five times and averaged marginal likelihoods per model and participant over these runs.

Model evaluation was based on protected exceedance probabilities and estimated model frequencies. Note, that since models can die out throughout the sampling procedure on participant level, it does not make sense to calculate global bayes factors as global model probabilities would become zero a soon as a model dies out for a single participant. Protected expected exceedance probabilities are based on the maximum a posteriori model estimate and hence, do not suffer from this problem (Stephan et al., 2009).

Finally, we inferred parameters for each participant and each model with an estimated frequency > 0 based on the whole data, but otherwise analog to the prior data run. The resulting posterior samples were used for posterior simulations and to calculate maximum a posteriori estimates of the model parameter. We took 256000 samples of the approximate posterior per participant and run simulations for all samples. Based on the trial-wise means of the simulated responses posterior simulation plots were created.

Priors

We used uniform priors over the interval (0,1) for $p(C = 1)$, d_{IVAE} as well as τ_A^{VAE} and τ_V^{VAE}, thus covering all possible values for these parameters with equal probability. Note that if the slope is parameterized in terms of the angle between X and Y axis a log-uniform prior yields equal probabilities for all angles, whereas a uniform prior is strongly biased towards large angles. The weighting factors $\omega_{SE,V}$ and $\omega_{SW,V}$ determine the ratio of auditory ventriloquism effects compared to visual ventriloquism effects, therefore a logarithmic uniform prior is more appropriate than a uniform prior as it is uniform across ratios. We choose a logarithmic uniform prior over the range $(1/100, 100)$ whereby 100 denotes de facto auditory dominance and 1/100 denotes de facto visual dominance.

So far, all parameters have the same effects across all competing models and act on the same scale, but this is not true for $\sigma^2_{\beta Aleft}$, $\sigma^2_{\beta Aright}$ and $\sigma^2_{\beta V}$. For instance, the sizes of the errorterms differ on average across models. If a model has a smaller errorterm on average than all other models, the parameter estimates and the variance of these estimates for $\sigma^2_{\beta Aleft}$, $\sigma^2_{\beta Aright}$ will be larger on average. Given the same prior across models, the posterior will then also be wider and thereby the marginal likelihood will be larger. Roughly speaking any weakly informative prior (this is however also true for any subjective prior) would a priori favor some models over others which is why we used partial bayes factors to minimize subjectivity. For the initial run we used log-uniform priors for $\sigma^2_{\beta Aleft}$, $\sigma^2_{\beta Aright}$ and $\sigma^2_{\beta V}$ with 0.001 as lower bound and max (σ^2_{mpq})*8 as participant-wise upper bound.

Results

Model-free Analysis

In the first experimental block isoluminant colors were derived via an adaptive staircase procedure, thereby equalizing the saliency of the colors. The luminance value of the reference color 45° was fixed to -20°. The average luminance *PSEs* (Figure 8,D) for colors -45° (-34.404, *SEM* = 0.249), 135° (-3.523, *SEM* = 0.468) and 225° (-26.464, *SEM* = 0.345) differed significantly ($\chi^2(1)$ = 4962.8, p < 0.001), reflecting the usage of an uncalibrated display. However, only two staircase blocks had to be repeated due to diverging staircases (17.6° and 55.63° difference) and all final staircases (Figure 8, A-C) converged. Figure 8 (D) shows that individual thresholds were narrowly distributed around the mean and no outliers were present, indicating that perceptually isoluminant colors were indeed derived for all participants.

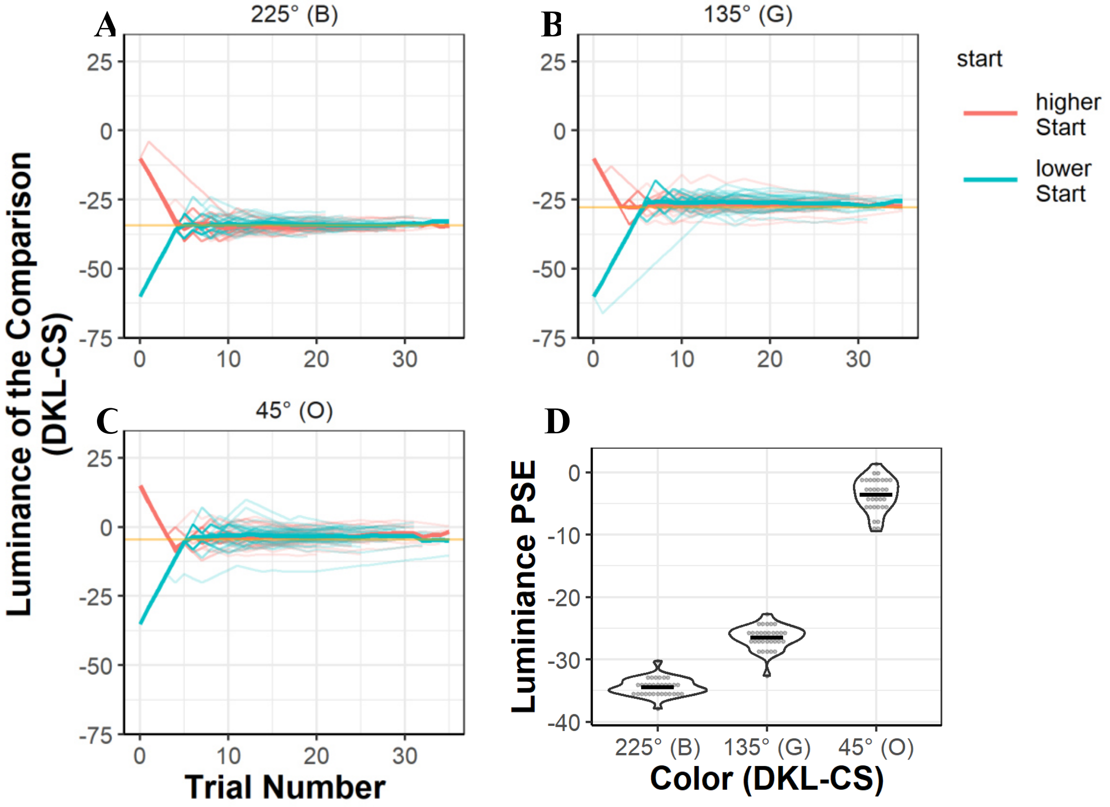


Figure 8. Individual and Group-level Results of the Luminance Adjustment. **A**, **B**, **C**: Individual (low alpha) and group-wise averaged response sequences of the one-up-one-down staircases for the luminance adjustment of each color (-45, 135 and 225°) relative to the reference color (45°). On average 10 reversal were accomplished after 28 trials. The first 5 reversal points were discarded, all remaining were used to calculate the PSEs. All individual staircases follow the general pattern of the group-wise averaged response sequences. Orange horizontal lines mark the group-wise PSE. **D**: Individual (low alpha dots) and group-wise averaged (black horizontal bars) PSEs indicating equiluminant colors relative to the reference color.

In a second step we examined whether stimuli differed in their reliability (Figure 9, A) based on the sensory modality (auditory vs. vision) and visual reliability (low vs. high). For auditory trials, the visual reliability of the paired visual stimulus was used as factor level. This allowed to test, whether auditory reliabilities differed a-priori across visual reliability conditions. We analyzed the absolute Res_{Base} in trials from the baseline measurements corrected for the slope of the underlying linear model (see Chapter II *Approximation of R and*

H for a theoretical justification and details of the methods). Corrected absolute Res_{Base} were submitted to a linear mixed model. Fixed effects were visual reliability and sensory modality. Moreover, we allowed random intercepts for each participant to account for our repeated measures. We found a main effect for visual reliability $\chi^2(1) = 563.16$, $p < 0.001$, for sensory modality $\chi^2(1) = 815.89$, $p < 0.001$ and an interaction effect of visual reliability and sensory modality $\chi^2(1) = 922.72$, $p < 0.001$. Most importantly, pairwise comparisons revealed that Res_{Base} were significantly larger when the visual reliability was low compared to when it was high only for visual trials $\chi^2(1) = 88.63$, $p < 0.001$ and not for auditory trials $\chi^2(1) = 1.74$, $p = 0.19$ indicating that the manipulation of the visual stimulus did indeed change the visual reliability and auditory stimuli did not differ in their reliability across visual reliability conditions. The results in Figure 9 (B) indicate that the reliability was highest for visual stimuli with high reliability and of similar size for auditory stimuli and visual stimuli with low reliability. In line with this observation, post-hoc pairwise comparisons revealed that Res_{Base} only differed significantly between auditory and visual trials when visual reliability was high χ^2 $(1) = 124.81$, $p < 0.001$, and not when visual reliability was low $\chi^2(1) = 0.19$, $p = 0.66$.

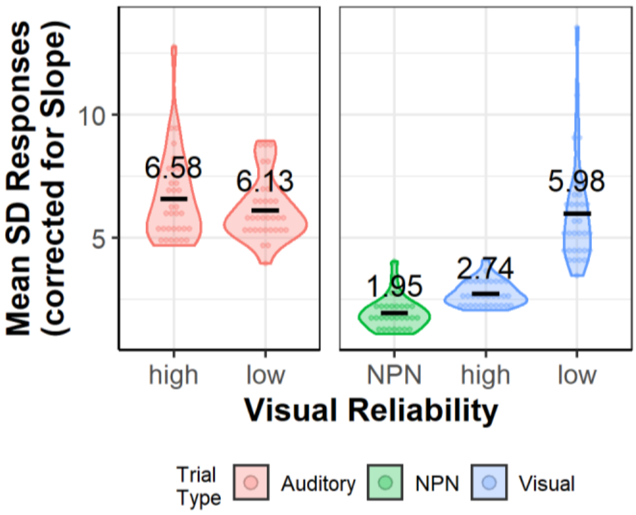


Figure 9. Sensory uncertainty estimates. Sensory uncertainties were estimated as standard deviations of the Res_{Base} from baseline blocks. The Res_{Base} were participant wise normalized by division through individual slopes. Points with low alpha indicate individual data points whereas black bars and text labels indicate the sample mean. Standard deviations are grouped by trial type and visual reliability. NPN refers to non-perceptual noise trials. For auditory trials, the visual reliability refers to the paired visual stimulus in bimodal trials.

Ventriloquism Effect

The auditory VE was apparent in a shift of auditory localization towards the accompanying VS (Figure 10, B left). This shift was of similar size for low reliable visual stimuli compared to high reliable visual stimuli. Similarly, a visual VE was apparent in a shift of visual localization towards the accompanying AS (Figure 10, B right). However, this shift was remarkably increased for low reliable visual stimuli compared to high reliable visual stimuli. Moreover, the size of the auditory VE depended on whether the same or a different disparity was used in the preceding trial (Figure 10, D left).

Statistically, the ventriloquism effect was analyzed based on the residuals Res_{VE} (see *Intermixed Block - Quantification of VE*) in bimodal trials in the intermixed blocks. Residuals were analyzed in separate linear mixed models for auditory and visual trials.

Auditory bimodal residuals were submitted to a linear mixed model. Fixed effects were visual reliability (low or high), previous disparity (same or different) and disparity (left or right). We found a significant main effect for disparity $\chi^2(1)$ = 2111.2 p < 0.001 indicating that auditory localization was indeed shifted in the direction of the visual stimuli. Post-hoc comparisons revealed that the aVE different from zero for both levels of visual reliability, *EMM* = 5.84, *z* = 32.15, p < 0.001 for visual reliability low and *EMM*= 5.75, *z* = 31.629, p < 0.001 for visual reliability height. Importantly no significant interaction of reliability and disparity was found $\chi^2(1)$ = 0.1, p = 0.81 indicating that the size of the aVE did not differ as a function of the visual reliability. Finally, a significant interaction of previous disparity and disparity was found, $\chi^2(1)$ = 6.4, p =0.01, indicating that the size of the aVE was lower when the previous disparity direction was different compared to when it was the same (Figure 10, D left).

Analogously visual bimodal residuals were submitted to a linear mixed model. Fixed effects were visual reliability (low or high), previous disparity (same or different) and disparity (left or right). We further included random intercepts for each participant. We found a significant main effect for disparity $\chi^2(1)$ = 601.7, p <0.001 indicating that visual localization was shifted in the direction of the auditory stimuli. Post-hoc comparisons revealed that the vVE was different from zero and in the direction of the auditory stimulus for both visual reliability low $\chi^2(1)$ = 1224.5, p < 0.001 and even high condition $\chi^2(1)$ = 12.1, p < 0.001.

Importantly a significant interaction of reliability and disparity was found $\chi^2(1)$ = 678.9, p < 0.001 indicating that the size of the vVE differed as a function of the visual reliability.

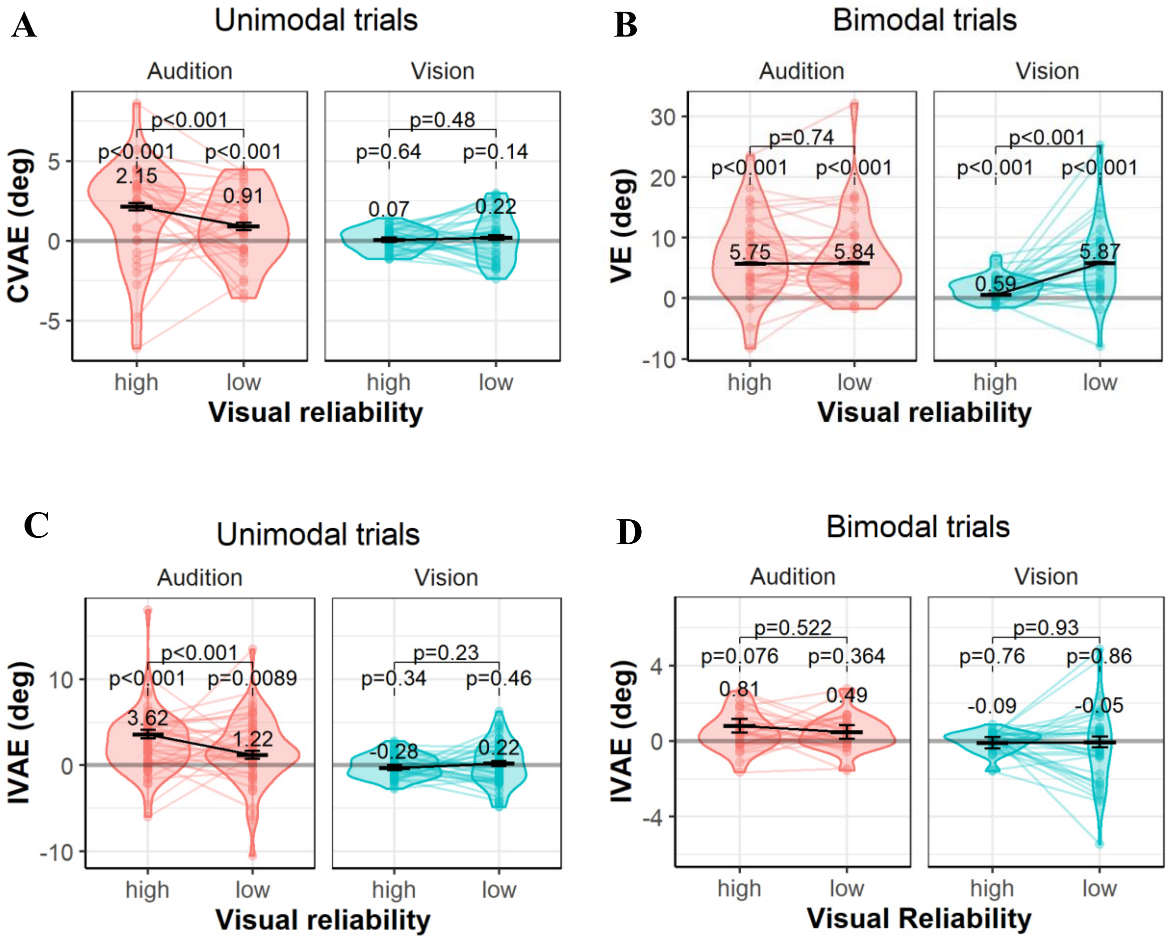


Figure 10. Effects of visual reliability on the CVAE, IVAE and VE. Estimates of the CVAE, IVAE and VE are grouped by the task-relevant modality (audition or vision) and the reliability of the (accompanying) visual Stimulus. Low alpha dots and lines indicate individual data points. *EMMs* are shown with black lines and dashes. Errorbars show the SE. P-Values are holm-adjusted whenever a significant main effect or interaction was present, all other p-values are not adjusted and presented only for completeness. **A**: Intercept shifts of unimodal trials in intermixed blocks compared to baseline as estimates of the CVAE. **B**: *EMM* differences of bimodal trials in intermixed blocks and unimodal trials in intermixed blocks as estimates of the VE. **C**: Intercept difference of unimodal trials with same vs. different previous disparity in intermixed blocks as estimates of the IVAE. **D**: *EMM* difference of bimodal trials with same vs. different previous disparity in intermixed blocks as estimates of the IVAE.

Finally, no significant interaction of previous disparity and disparity was found $\chi^2(1) = 0.12$, $p = 0.73$ indicating that the size of the vVE did not significantly differ for different previous disparities compared to equal previous disparities. Complete LMM results for visual and auditory Res_{VE} are presented in the Appendix A (Table A. 1 and Table A. 2).

Cumulative and immediate Aftereffect

In a next step we analyzed the unimodal residuals Res_{VAE} to examine the effect of visual reliability on the CVAE and IVAE. Consistent auditory aIVAEs as well as aCVAEs were observed regardless of the visual reliability (Figure 10). Most importantly, the aCVAE as well as the aIVAE were modulated by the visual reliability. Low visual reliability led to a reduced IVAE (*EMM* = 1.22, *SE* = 0.47) compared to high visual reliability (*EMM* = 3.62, *SE* = 0.47) and a reduced aCVAE (*EMM* = 2.149, *SE* = 0.233) compared to high visual reliability (EMM= 0.907, *SE* = 0.233).

Neither a vCVAE nor an vIVAE was found for visual unimodal stimuli. Again, visual and auditory trials were statistically analyzed separately.

The linear mixed model analysis of the Res_{VAE} revealed a significant main effect for disparity χ^2 (1) = 86.48, p < 0.001 indicating that auditory unimodal localization was shifted distinctively for leftward and rightward adaptation in comparison to baseline localization. However, post-hoc comparisons revealed, that unimodal shifts were to the right for both rightward adaptation (*EMM* = 2.552, *z* = 12.138, *p* < .0001) and leftward adaptation (*EMM* = 0.450, *z* = 2.146, *p* = 0.064), although not significant after Bonferroni correction in the latter case. Importantly, rightward shifts were significantly lower for leftward adaption (contrast = “left-right”, *estimate* = -1.53, *z* = -9.28, *p* < 0.001). This indicates, that the aCVAE induced by rightward adaptation might have partially generalized to the left or at least lead to a suppression of leftward adaptation. Importantly, we observed a significant interaction between disparity and reliability χ^2(1) = 14.25, p < 0.001. A post-hoc contrast revealed that the difference between shifts for leftward adaptation and rightward adaptation increased when visual reliability was high (*EMM* = 2.15, *SE* = 0.23) compared to low (*EMM* = 0.91, *SE* = 0.23), *z* = 3.77, *p* <0.001. Importantly, for rightward adaptation also the absolute aCVAE increased when visual reliability was high compared to low contrast = “high - low”, *estimate* = 1.30, z = - 5.59, *p* < 0.001. No effect was found for leftward adaptation contrast = “high - low”, estimate= 0.06, *z* = 0.27, p = 0.79. Hence, reliability of the visual stimulus indeed affected the size of the aCVAE for rightward adaptation.

A significant interaction of disparity and previous disparity, χ^2(1) = 53.91, p < 0.001 confirmed the occurrences of the aIVAE. Post-hoc comparisons showed that shifts in direction of the disparity were significantly smaller for previous disparity same trials compared to previous disparity different trials for leftward (contrast = “same – different”, *estimate* = 1.3, z = 5.68, p < 0.001) and rightward (contrast = “same – different”, *estimate* = 1.10, z = 2.37, p < 0.001) disparities.

The effect of reliability on the aIVAE was confirmed by a threefold interaction of disparity, previous disparity, and reliability, $\chi^2(1)$ = 13.36, p < 0.001. In fact, aIVAEs were larger for high visual reliabilities for both leftward (contrast = “high – low”, *estimate* = 1.3, z = 2.80, p = 0.005) and rightward (contrast = “high – low”, *estimate*=1.10, z = 2.37, p = 0.0179) disparity. No effects including the previous task were found.

Visual $Res_{\ VAE}$ were submitted to an otherwise identical linear mixed model. Importantly neither a main effect of disparity $\chi^2(1)$ = 1.90, p = 0.168 nor an interaction between disparity and reliability χ^2 (1) = 0.50, p = 0.478 or disparity and previous disparity $\chi^2(1)$ = 0.02, p = 0.884 were found. Hence, neither a significant vCVAE nor a significant vIVAE was observed.

Complete LMM results for visual and auditory $Res_{\ VAE}$ are presented in the Appendix A (Table A. 3 and Table A. 4).

Model-based Analysis

Multisensory Integration shows optimal Weighting

The standard Causal Inference model cannot account for visual overweighting during multisensory integration which is reported with increased regularity (Battaglia et al., 2003; Meijer et al., 2019). However, we found that only a small subset of the participants, estimated frequency(*eF*) = 0.09 but *PEP* = 0.0, is better explained by suboptimal weighting (not necessarily visual overweighting). Audio-visual integration was best explained by models assuming optimal weighting, *eF* = 0.76 and *PEP* = 1.0, compared to models assuming suboptimal weighting or suboptimal estimation. These results justify the selection of CI-model with optimal weighting as integration model for subsequent parameter estimation and model selection for the IVAE and CVAE.

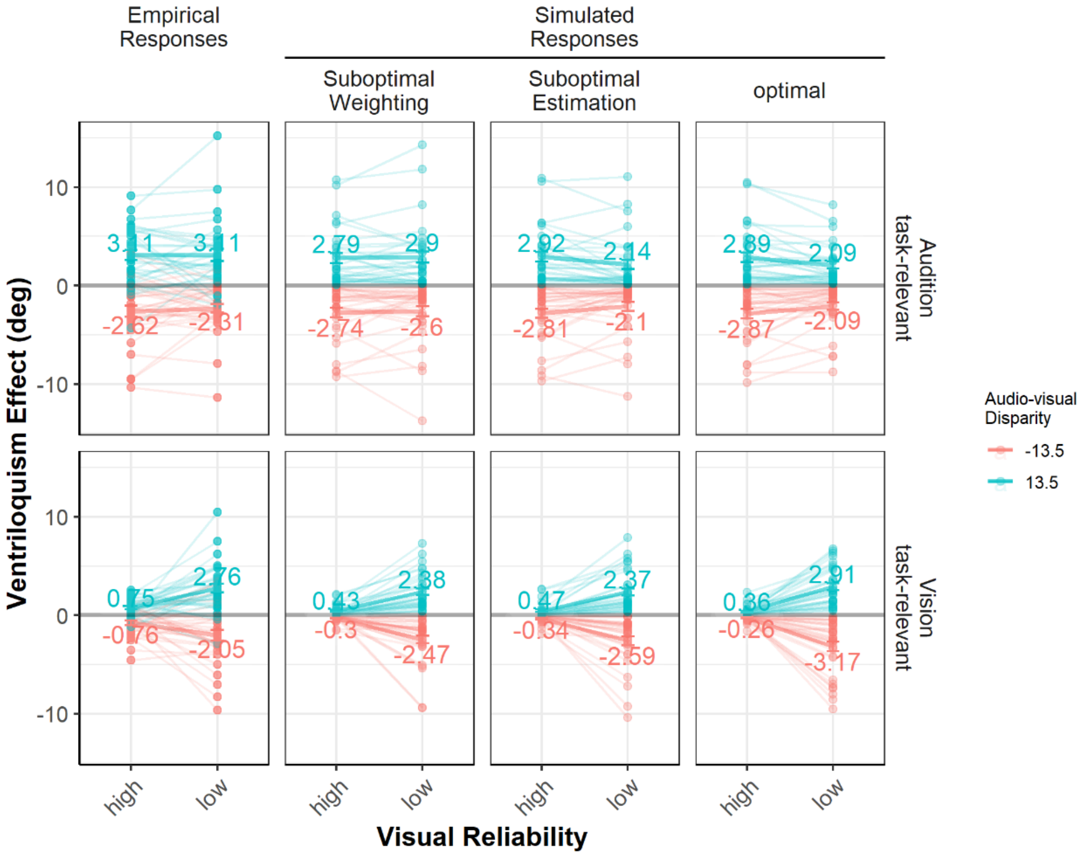


Figure 11. Empirical results and posterior simulations of the best fitting VE models. Auditory and visual average VEs as a function of visual reliability (**top** vs. **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left).

The size of the VE is better recovered in posterior simulations, when visual overestimation is considered (Figure 6), avoiding an underestimation of the auditory VE, as observed for the optimal model and the overweighting model. However, this improvement is not sufficient to justify the increased model complexity according to our Bayesian model selection approach, that implicitly incorporates the parsimony principle. This is in fact strong evidence in the direction of optimal integration, since the suboptimal estimation factor can essentially reproduce an arbitrary weighting between vision and audition and nevertheless fails to substantially improve the model fit on individual level.

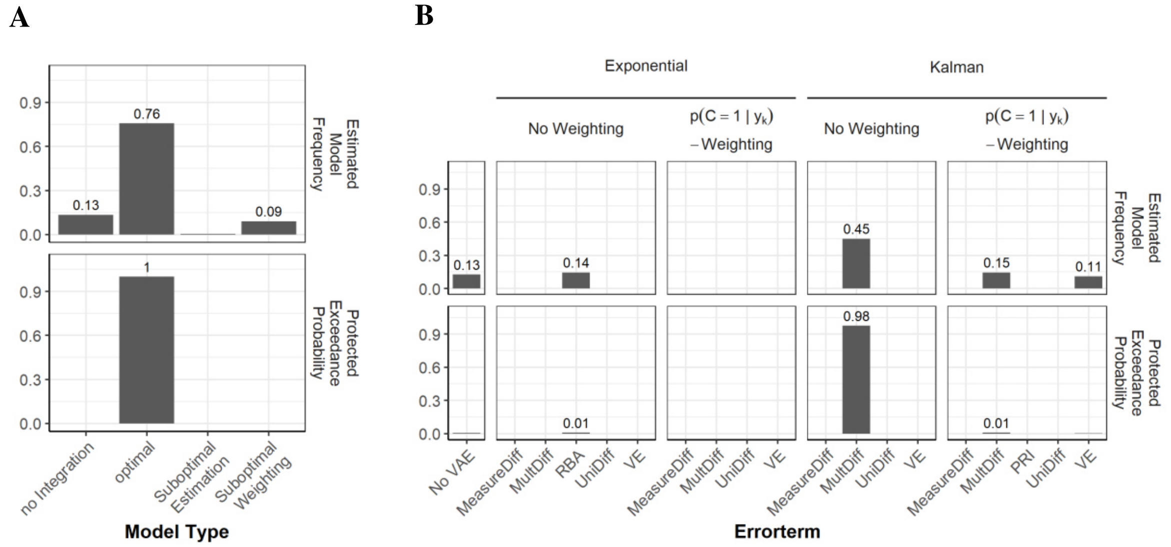


*Figure 12. Estimated model frequencies (**Top**) and protected exceedance probabilities (PEP, **Bottom**) from random effects Bayesian group model comparison [11] for the VE and CVAE.* The protected exceedance probability denotes the probability that a specific model is more frequent than all other models, protected against the Bayesian Omnibus Risk, that all models are equally frequent. Participants that were best described by a model assuming no integration or CVAE in a previous run were excluded. **A:** Adapted CI-models (remaining N=23). Participants can either make optimal usage of their sensory reliabilities, or suboptimal estimation or weighting can lead to suboptimal integration. **B:** Cumulative VAE-models (remaining N=25). Models are categorized by the errorterm, learning mechanism and type of CI-Weighting. Factors of the model taxonomy are described in detail in Chapter II.

To further clarify the role of suboptimal weighting we show parameter estimates derived from model averaging (Figure 13). Hereby, the uncertainty about which model is correct is considered in the parameter estimation process, by calculating the weighted average of the parameter estimates across models, with the model evidence as weights. This approach avoids inconsistencies between model comparison and parameter estimation procedures (Campbell & Gustafson, 2022). The distribution of $\omega_{SW,V}$ is approximately normal (Shapiro-Wilk normality test, *W*= 0.963, *p*= 0.325) with *M*=0.01 and the *skewness*= -0.006 not different from 0 (D'Agostino test for skewness, *z* = -0.015, *p* = 0.988), indicating that the distribution is symmetric around the mean. Parameter estimates for all models with *eF* > 0 are given in Table A. 5.

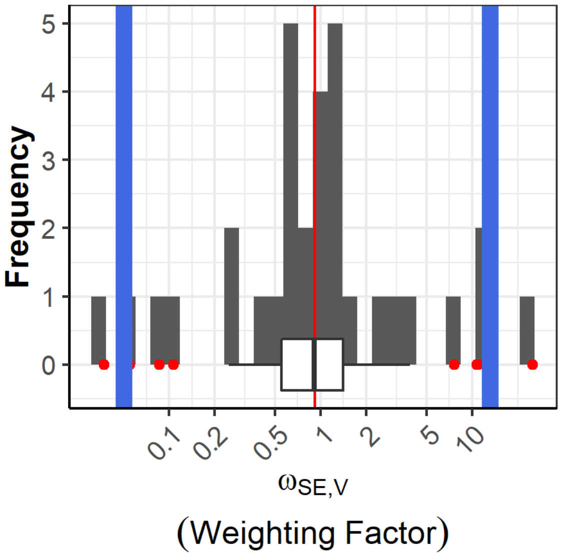


Figure 13. Histogram of the parameter estimates for $\omega_{SE,V}$ based on model averaging. Blue vertical bars indicate 0.95-credible intervals (Equal-tailed Interval). Red vertical line indicates the sample mean. Boxplot depicts 0.25 and 0.75 quartiles as well as median. Outliers are show as red dots (0.25/0.75 quartiles –/+ 1.5 * inter-quartile range).

Cumulative Recalibration depends on Multisensory Integration

The model comparisons analysis as well as posterior simulations identify the MultDiff model, based on a Kalman learner and with no posterior weighting as the most likely (*eF* = 0.45, *PEP* = 0.98). The model can recover the main pattern -a decreasing aftereffect with decreasing visual reliability, of the auditory CVAE (Figure 14). Importantly this is achieved without negatively affecting the predictions of the visual CVAE (Figure 15) as well as visual and auditory VEs (Figure 17). Other estimated frequencies different form zero were only acquired for the RBA model (*eF* = 0.14, *PEP* = 0.01), for the MultDiff Model, based on a Kalman learner and with $p(C = 1|y_k)$ – Weighting (*eF* = 0.15, *PEP* = 0.01) and the VEDiff Model, based on a Kalman learner and with $p(C = 1|y_k)$ – Weighting (*eF* = 0.11, *PEP* = 0.0). Importantly, posterior simulations demonstrate, that Reliability-based adaptation is not capable to reproduce our results but remarkably underestimates the CVAE ((Figure 14). Moreover, RBA predicted visual aftereffects when the visual reliability was low (Figure 15). The MultDiff Model, based on a Kalman learner and with $p(C = 1|y_k)$ – Weighting reproduces CVAEs of similar size as in the empirical data, but fails to predict the decrease of CVAE with decreasing visual reliability. Moreover, auditory VEs are underestimated to greater extend compared to the best fitting model (Figure 15).

The VEDiff Model, based on a Kalman learner and with $p(C = 1|y_k)$ – Weighting can reproduce the general trend of the CVAE similarly well as the best fitting model. Importantly, this model also underestimates the VE to greater extend then the best fitting model (Figure 15).

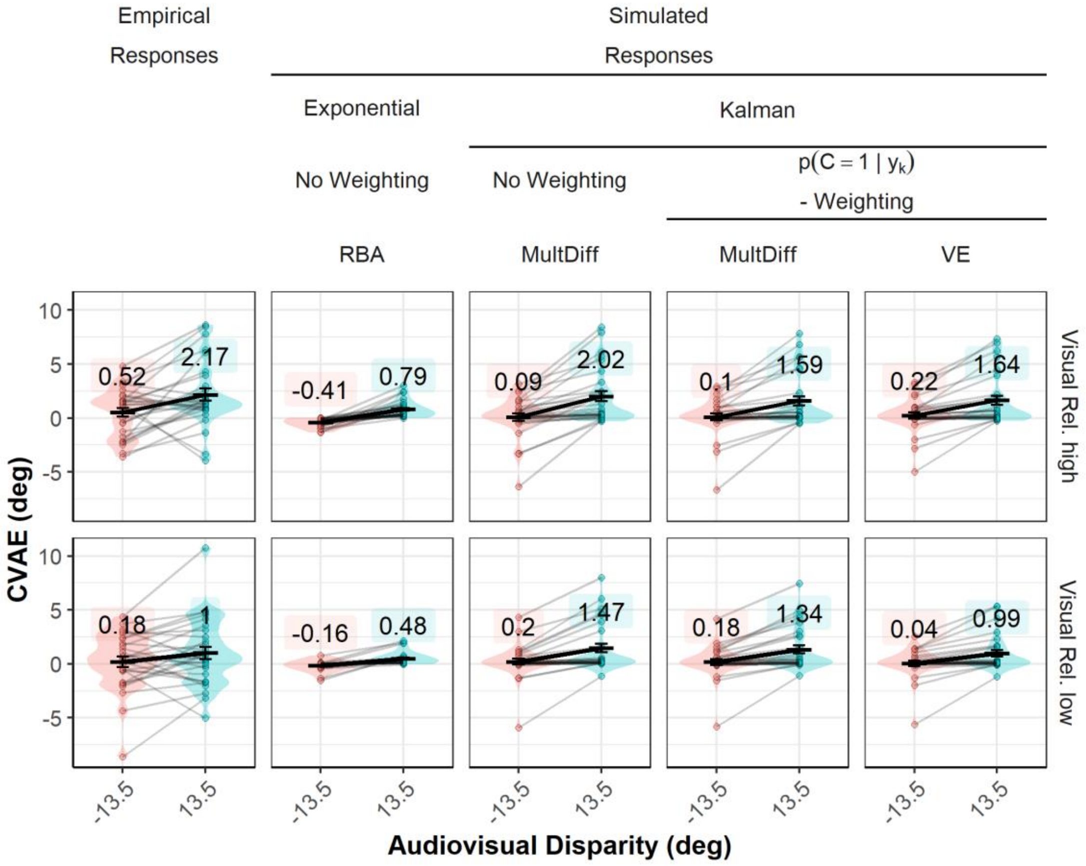


Figure 14. Empirical results and posterior simulations of the best fitting CVAE models for the auditory CVAE. Auditory average CVAEs as a function of visual reliability (**top** vs. **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Factors of the model taxonomy are described in detail in Chapter II.

An estimated 13% of our sample is best described by an integration model assuming no CVAE at all. Parameter estimates for all models with *eF* > 0 are given in Table A. 6.

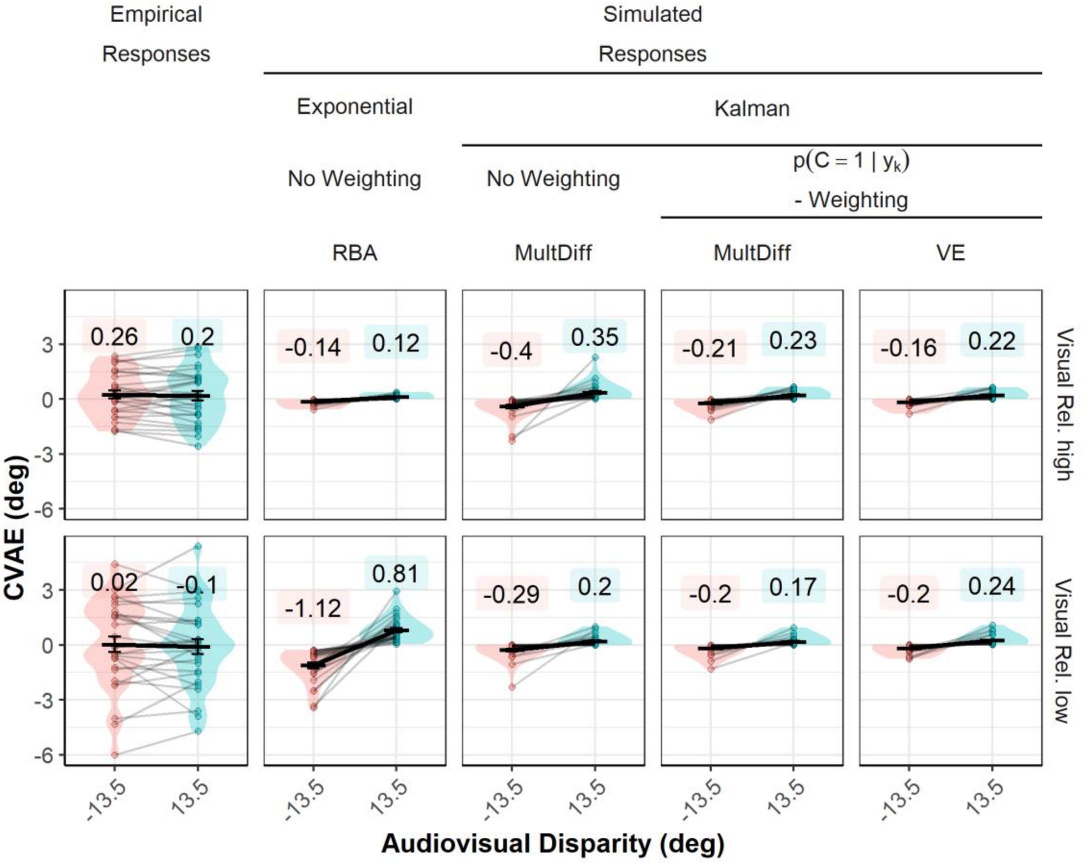


Figure 15. Empirical results and posterior simulations of best fitting CVAE models for the visual CVAE. Visual average CVAEs as a function of visual reliability (**top** vs. **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Factors of the model taxonomy are described in detail in Chapter II.

Several models proposed in the literature (Bosen et al., 2018; Burge et al., 2010; Ghahramani et al., 1997; Hong et al., 2021; Sato et al., 2007; Watson et al., 2019). failed to account for our result pattern and their estimated frequency was zero (Figure 18). Posterior simulations (Figure A. 1- Figure A. 4) provide conclusive insights, why these models were deprecated throughout the model comparison. When different learning rates for vision and audition are considered, RBA (Exponential – No Weighting – VE) can account for the observed auditory CVAEs (Figure A. 1) and the lack of visual CVAEs (Figure A. 2), yet in exchange RBA even further underestimates the auditory VE (Figure A. 3) compared to the best model.

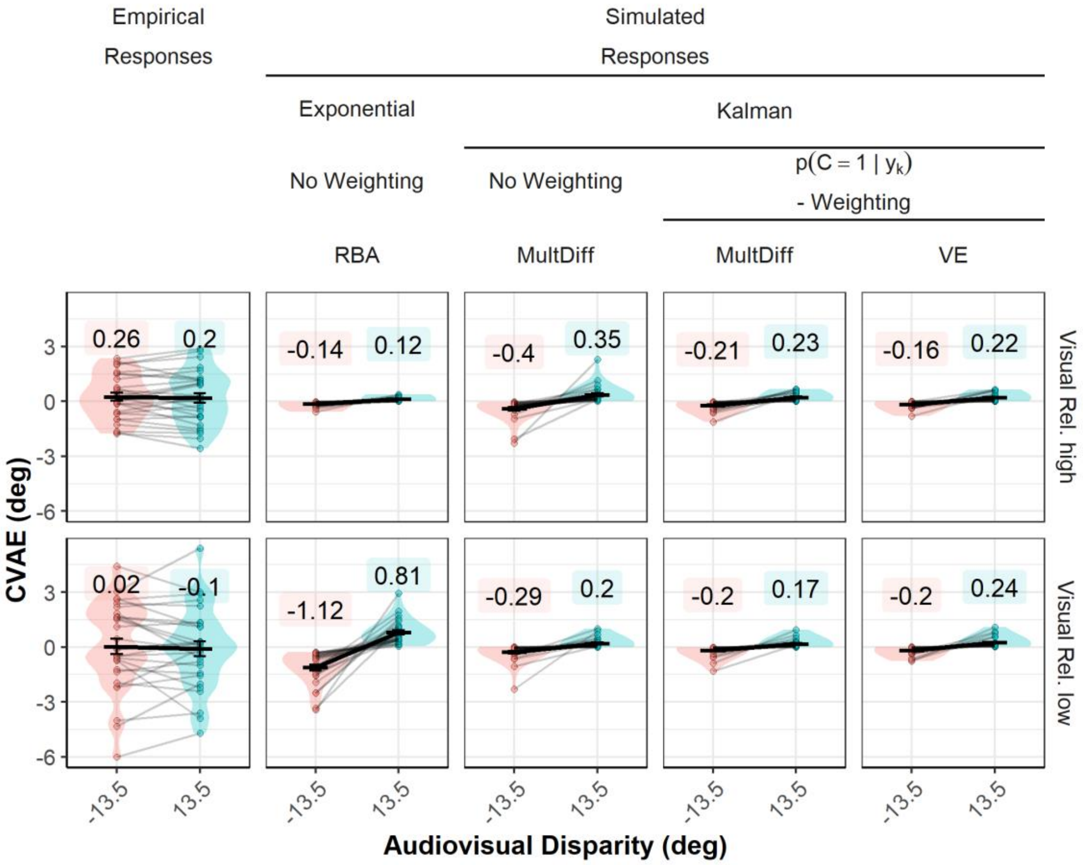


Figure 16. Empirical results and posterior simulations of best fitting CVAE models for the visual CVAE. Visual average CVAEs as a function of visual reliability (**top** vs. **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Factors of the model taxonomy are described in detail in Chapter II.

The recently proposed model of Hong (2021) would predict an increase of the auditory CVAE with decreasing visual reliability, since the auditory VE did not decrease with decreasing visual reliability but $p(C = 1|y_k)$. Again, this conflict is resolved by underestimating the auditory VE (Figure A. 3). This underestimation is less pronounced when the model of Hong (2021) is modified with a Kalman learner. Fixed ratio adaptation (Exponential – No Weighting – UniDiff) underestimates the auditory CVAE in the visual reliability high condition and overestimates adaptation in the visual reliability low condition (Figure A. 1).

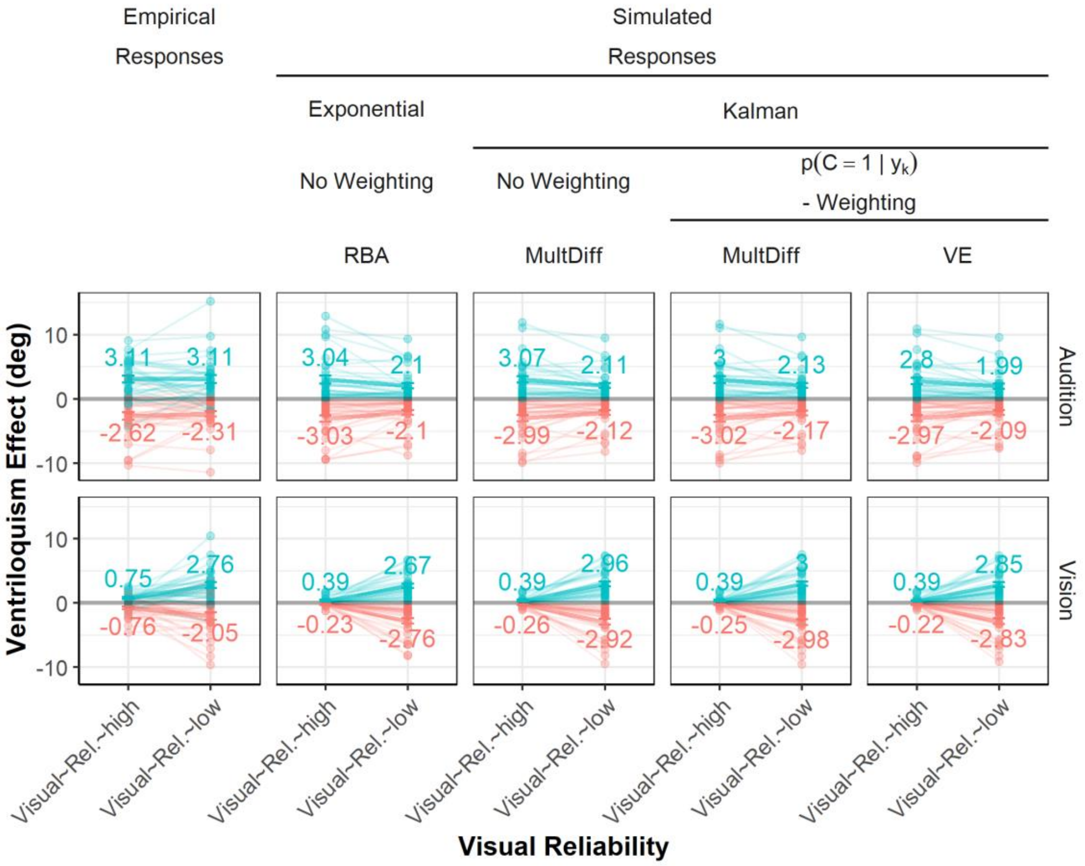


Figure 17. Empirical results and posterior simulations of the best fitting CVAE models for the auditory and visual VE. Auditory and visual average VEs as a function of visual reliability (**top** vs. **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Factors of the model taxonomy are described in detail in Chapter II.

Instantaneous Recalibration is incorporated in Multisensory Integration

In the initial run a total of 2 participants were attributed to the no-IVAE-model with a probability > 0.75. We decided to exclude these participants and continued with an exploratory analysis of the remaining 30 participants to investigate whether insights regarding the computational principles of the IVAE can be made based on the data of participants that likely showed an actual IVAE.

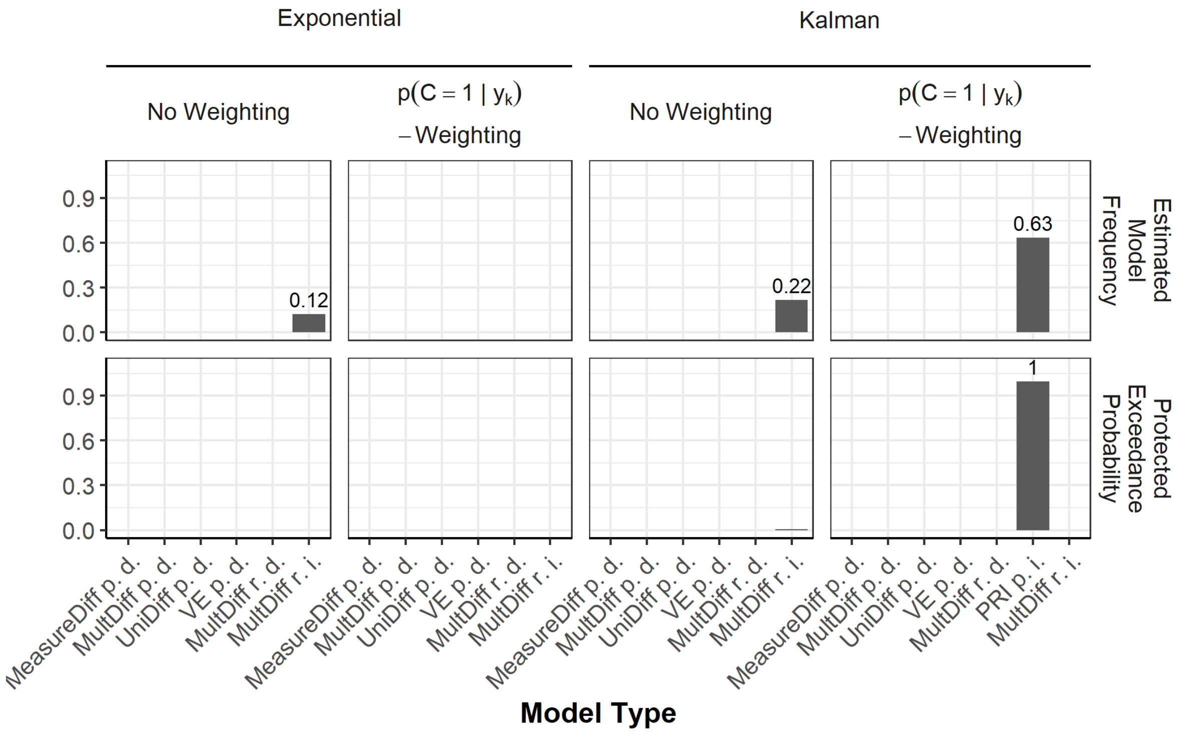


*Figure 18. Estimated model frequencies (**Top**) and protected exceedance probabilities (PEP, **Bottom**) from random effects Bayesian group model comparison for the IVAE.* Participants that were best described by a model assuming no IVAE in a previous run were excluded. IVAE-models are depicted (remaining N=30). Models are categorized by the errorterm, learning mechanism, type of Weighting as well as the timepoint at which the IVAE is updated (d. = delayed, i.e., after a bimodal trial, i. = instantaneous, i.e., within a bimodal trial) and the processing stage (r. = response level, p.= perceptual level). Factors of the model taxonomy are described in detail in Chapter II.

In the second run, only three models yielded estimated frequencies different from zero. At the response level, for the MultDiff model, No Weighting and Kalman learning an estimated frequency of 0.22 with a PEP of 0.0 was obtained. For the MultDiff, No Weighting and exponential learning model an *eF* of 0.12 with a *PEP* of 0.0 was obtained. Whereas on the perceptual level for the PRI model an *eF* of 0.63 with an *PEP* of 1.0 was obtained.

Posterior simulations of all models provide similar estimates of the IVAE (Figure 19), i.e., the models tend to underestimate the IVAE in the unimodal condition. Apart from that, the models are capable to reproduce the decreased IVAE in bimodal trials compared to unimodal trials. All models assume that the IVAE is updated instantaneously during bimodal trials, i.e., the updated IVAE immediately affects the response in bimodal trials.

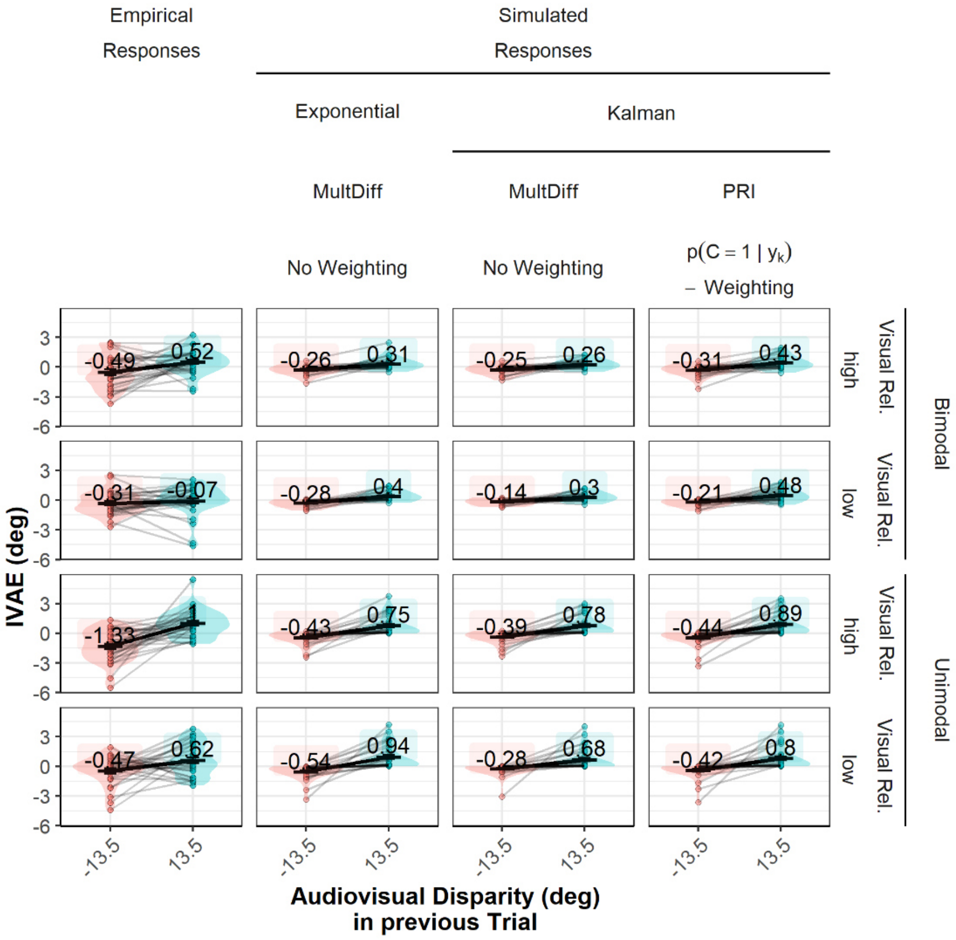


Figure 19. Empirical results and posterior simulations of the best fitting IVAE models. Models are categorized by the errorterm, learning mechanism and type of CI-Weighting. All models apply instant updating of the IVAE. Auditory average IVAEs as a function of visual reliability and trial type (**top** to **bottom** row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Factors of the model taxonomy are described in detail in Chapter II.

Posterior simulations from the best fitting models that do not assume instantaneous update are shown in Figure A. 4. Hereby, for all models the IVAE is of roughly the same size in bimodal trials compared to unimodal trials. To control whether this is a general pattern, we simulated datasets (*Figure 20*) based on the sample mean of the best fitting parameters over a wide range of σ_β^2 (Variance of the bias prior). The simulations demonstrate that only models assuming an instantaneous update can predict large discrepancies between IVAEs in bimodal and unimodal trials.

A PRI
$p(C = 1|y)$-Weighting
Kalman learner
perceptual level
instantaneous update

B MultDiff
no Weighting
Kalman learner
response level
instantaneous update

C MultDiff
$p(C = 1|y)$-Weighting
Kalman learner
response level
delayed update

D MultDiff
$p(C = 1|y)$-Weighting
Kalman learner
response level
delayed update

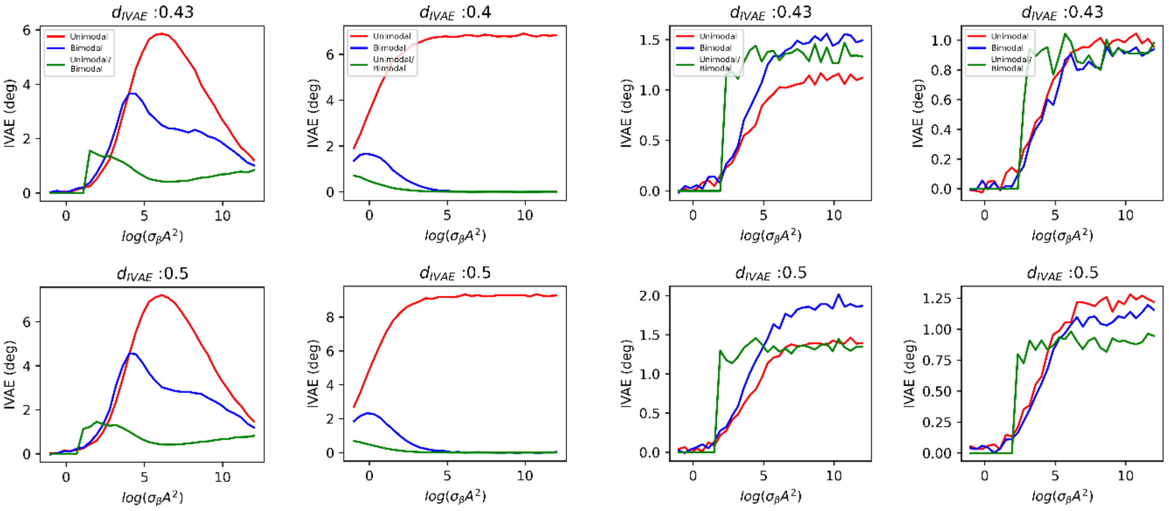


*Figure 20. Simulated IVAE in bimodal trials for the best fitting IVAE models (**A** and **B**) as well as the two best fitting IVAE models with delayed update (**C** and **D**).* Simulations are based on the sample mean of the participant-wise best fitting parameter (including auditory and visual reliabilities). Unimodal IVAEs are shown in red, bimodal IVAEs are shown in blue and the ratio is depicted in green. The upper row shows simulations for the average reliability in the visual reliability high condition, whereas the lower row shows the results for the visual reliability low condition. Factors of the model taxonomy are described in detail in Chapter II.

Going further the simulations show, that a delayed update of the IVAE essentially leads to IVAEs of similar size in bimodal and unimodal trials, which is clearly not present in our data. Parameter estimates for all models with *eF* > 0 are given in Table A. 7.

Comparison of IVAE and CVAE

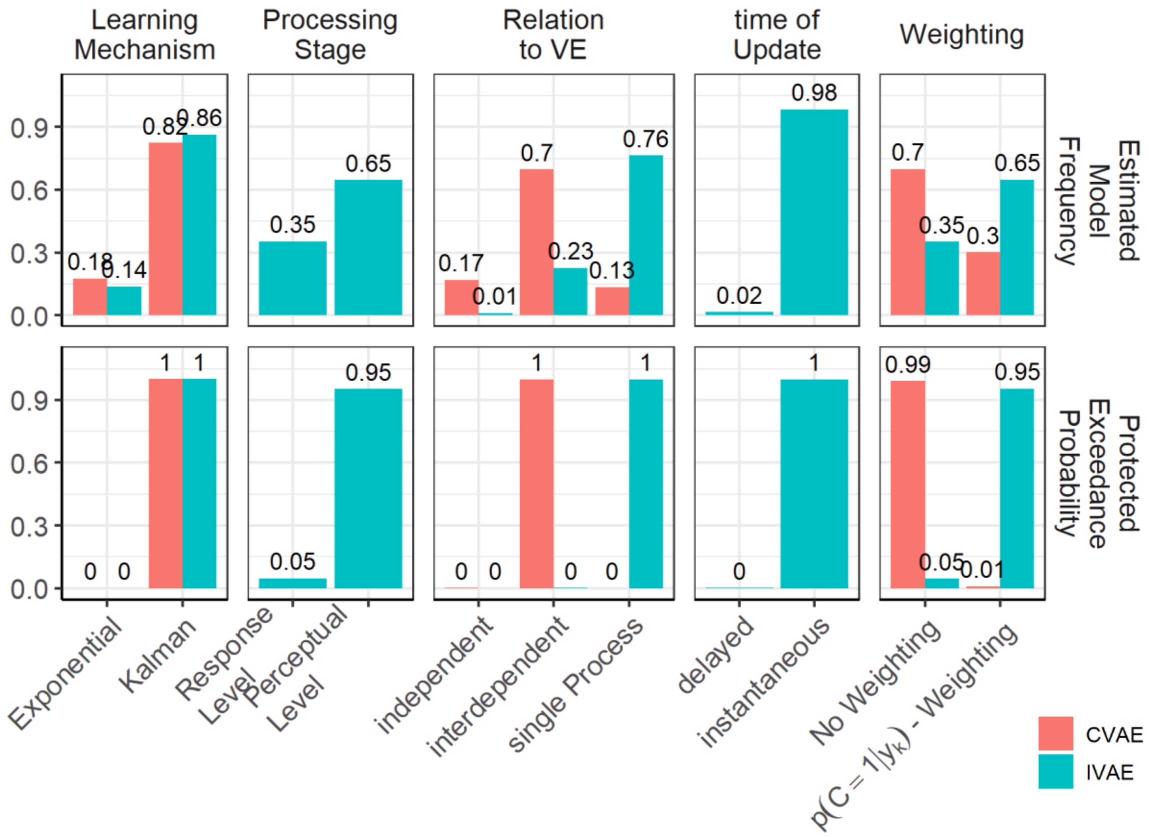


*Figure 21. Estimated model frequencies (**Top**) and protected exceedance probabilities (PEP, **Bottom**) using model averaging.* Participants that were best described by a model assuming no IVAE (remaining N=30) in a previous run were excluded. Models indicating no IVAE or no CVAE were not excluded from analysis. Models are categorized by the model factors learning mechanism, processing stages, their relation to the VE and the type of Weighting (columns from **left** to **right**). For each model factor Bayesian model averaging was applied over all other factors. Factors of the model taxonomy are described in detail in Chapter II.

To compare the IVAE and the CVAE based on our model taxonomy we used model averaging to evaluate each model factor separately (see Figure 21 for an overview). Both processes seem to rely on Kalman learning (*PEP* = 1 for CVAE and IVAE). Previous studies suggested the aCVAE is based on early perceptual stages, our results do also favor a perceptual processing stage for the aIVAE (Figure 21, Panel Processing Stages). Moreover, the aCVAE is likely interdependent to multisensory integration (*PEP* = 1, Figure 21, Panel Relation to VE) indicated by an errorterm that includes estimates provided by multisensory integration, whereas the aIVAE is likely a single process with multisensory integration (*PEP* = 1, Figure 21, Relation to VE). Finally, more participants perform $p(C = 1|y_k)$-Weighting throughout instantaneous recalibration compared to cumulative recalibration (Figure 21, Weighting).

Discussion

Behavioral Findings

To evaluate IVAE, CVAE and VE in common, we adapted the paradigm of Bruns et al. (Bruns & Röder, 2015) by adding two levels of visual reliability and a unimodal baseline measurement. This allowed to measure IVAE, CVAE and VE by comparing localization during intermixed stimulation and unimodal stimulation at baseline. The behavioral results clearly indicated a modulation of the IVAE and CVAE by the reliability of the visual stimulus. Surprisingly, we did not find a modulation of the auditory VE by visual reliability. The latter finding can be explained by the non-linear relation between reliability and the magnitude of the VE described by the "Causal Inference"-model (Körding et al., 2007). Also, in accordance with the CI-model, we found a rarely reported visual ventriloquism effect.

The lack of visual recalibration excludes purely reliability-based recalibration mechanisms (Burge et al., 2010; Ghahramani et al., 1997). As the visual reliability in the low condition was on average at a similar level as the auditory reliability, these models would predict visual IVAEs as well as CVAEs. Secondly, the effect of reliability on CVAE and IVAE rules out simple fixed ratio mechanisms (Zaidel, Turner, & Angelaki, 2011), as these would have predicted equal auditory CVAEs or IVAEs across visual reliabilities.

In contrast to previous studies (Lewald, 2002; Recanzone, 1998; Woods & Recanzone, 2004) we observed a transfer of the CVAE across frequencies. It is an ongoing debate whether the CVAE transfers across frequencies or not. In systematic studies Frissen et al. (2003, 2005) found transfer even across several octaves. A more recent study (Bruns & Röder, 2019) suggests, that the CVAE might be partially frequency specific in a way that concurrent recalibration to the left and to the right leads to a selective suppression of leftward adaptation. Similarly, we observed mainly rightward adaptation in our study. However, our results are not only in line with suppression but also transfer since we observe a trend to rightward shifts even under leftward adaptation.

Importantly, we did not use sinusoidal tones (as e.g., Lewald, 2002; Recanzone, 1998; Woods & Recanzone, 2004) but narrow band noise. A more recent study (Ege et al., 2019) using also narrowband noise found similar transfer as we did.

Multisensory Integration

Most participants were best described by a model assuming optimal weighting, indicating that in line with previous studies (Alais, 2004; Beierholm et al., 2020; Meijer et al., 2019; Rohe & Noppeney, 2015), participants were able to update their internal estimates of the visual reliability. In contrast to previous studies, we found that Integration is optimal for the majority of our sample (Battaglia et al., 2003; Meijer et al., 2019).

Interestingly, none of the former studies that systematically investigated auditory and visual weights tested for or reported a visual VE (Alais & Burr, 2004; Battaglia et al., 2003; Hong et al., 2021; Meijer et al., 2019), although this is a fundamental prediction of maximum likelihood integration and Causal Inference. In fact, when responses to visual components in bimodal trials are not analyzed, scenarios might occur where it is not possible at all to differentiate between optimal and suboptimal weighting. For instance, a larger than optimal auditory VE can be fitted by simply assuming a higher a-priori probability of a common cause. In this case the CI-model would predict an overly large visual VE, but this goes unnoticed when the visual VE is not empirically measured. Hence, this misspecification must be avoided by measuring visual responses in bimodal trials. Hereby a false prior probability of a common cause would lead to an overestimation of the visual VE and an overestimation of the visual reliability would lead to an underestimation of the visual VE. In this respect, our study contributes to the literature by demonstrating that when parameter misspecification is less likely most participants optimally weight vision.

Meijer (2019) provide another explanation for suboptimal weighting, visual and auditory percepts might not have been fully fused in their study and vision could have been mistakenly interpreted as task relevant. As we explicitly told participants which sensory component to localize in each trial, this explanation is less likely for our experiment. This might actually explain, why Meijer (2019) find visual overweighting in 20 out of 36 participants, whereas we only find suboptimal weighting (i.e. visual overweighting and underweighting) in an estimated 9% of our sample. Importantly, participants did not systematically overweight vision, but the distribution of the suboptimal weighting parameter was symmetric around 1 (Figure 13), indicating that over and underweighting occurred approximately equally often. Hence, our results speak against a general tendency to overweight vision and we further conclude that the weighting of vision and audition might be closer to optimal than the results of Meijer et al. (2019) suggested.

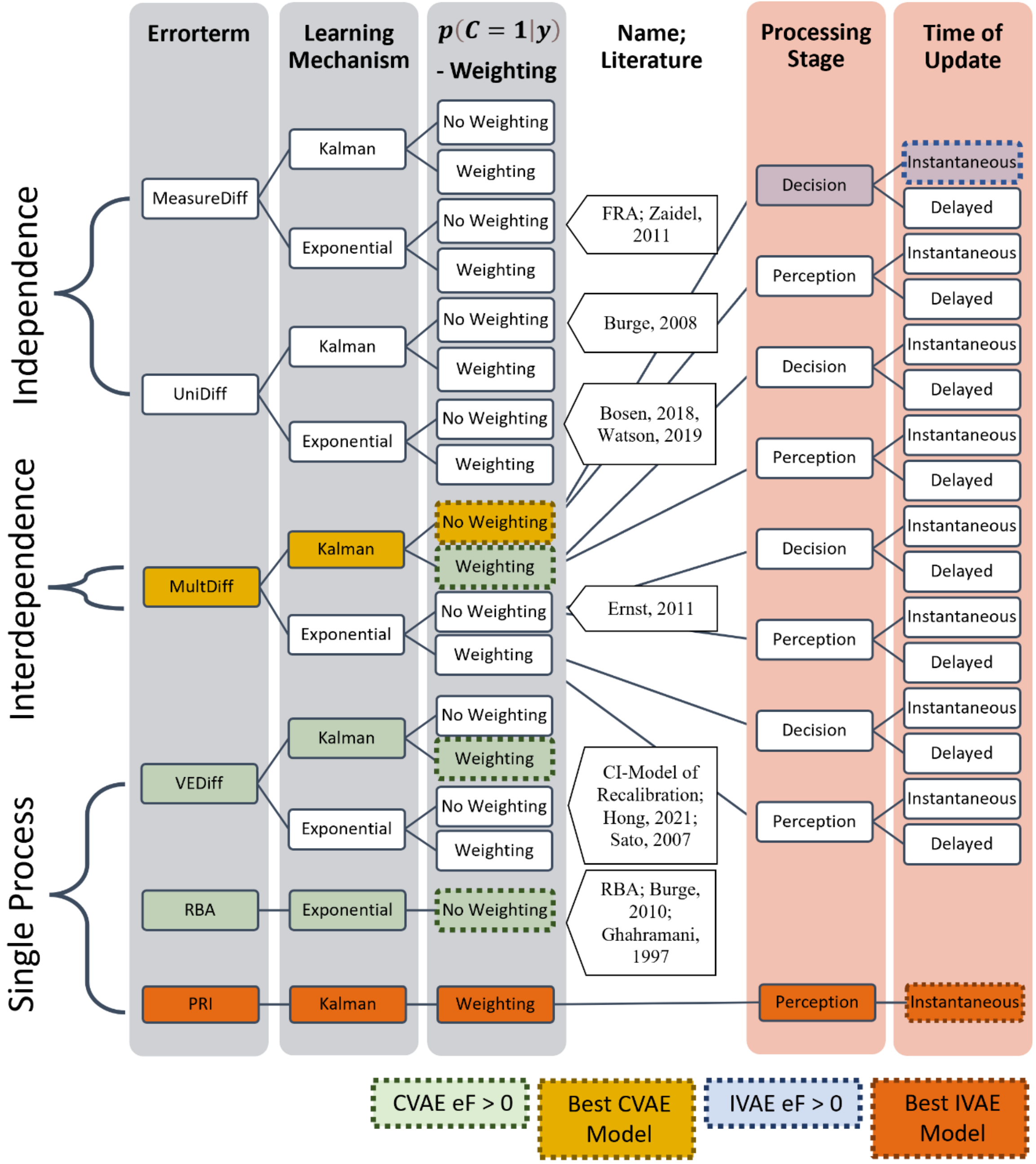


Figure 22. Overview of all tested CVAE and IVAE models. All levels of Errorterm, Learning Mechanism and Weighting were realized for the IVAE as well as the CVAE (grey background). For the IVAE additionally the levels of Processing Stage and Time of Update were realized (pale pink background), but only for the MultDiff Errorterm. Best IVAE and CVAE models are underlined in gold. Factors of the model taxonomy are described in detail in Chapter II.

Multisensory Recalibration

The most frequent CVAE model (Kalman learning, no Weighting, MultDiff errorterm; see

Figure 22 for a schematic overview of the modelling results) with the highest exceedance probability is a model that relies on the discrepancy between multisensory estimates of auditory and visual positions (MultDiff) similar to the model of Ernst & Di Luca (2011). However, our model assumes a Kalman Filter as learning mechanism with approximated error uncertainties. The reduced aftereffect when visual reliability is low, is thereby the consequence of two overlaying effects. On the one hand, the errorterm gets smaller as the increasing visual VE reduces the residual multisensory discrepancy. On the other hand, the uncertainty of the errorterm increases, reducing the trial-by-trial Kalman gain and thereby leading to slower adaptation. Moreover, the behavioral data as well as our modelling results indicate that neither RBA nor FRA can account for the CVAE. In the low visual reliability condition, RBA would have predicted large visual aftereffects in clear contradiction to the absence of any visual CVAEs in our data (Figure 15). Simple FRA on the other hand cannot account for the reliability dependence of the CVAE and even predicts an increased CVAE when the visual reliability is low (Figure A. 1).

Therefore, our results seem to be in conflict with a previous study that describes visual-vestibular recalibration as FRA (Zaidel et al., 2011) whereby the vestibular sense is consistently stronger recalibrated than vision. Although some studies suggest visual-vestibular heading perception is also based on Causal Inference (de Winkel et al., 2017; Dokka et al., 2019), it seems that in general the binding tendency is higher. De Winkel et al. (2015) show forced fusion integration in 5 out of 9 participants and the estimate of $p(C = 1)$ was above 0.94 for the remaining participants. For such a high degree of integration the residual discrepancy is certainly not a good errorterm. Due to the different principles of visuo-vestibular and audio-visual integration, it is possible that their respective forms of recalibration evolved differently. A clear indicator for this assumption is that Zaidel et al. (2011) observe visual aftereffects, which to our knowledge cannot be observed in audio-visual recalibration of healthy adults (Bruns et al., 2022), even if the visual reliability is heavily degraded (Figure 14) .

What is more important, is that we come to the same main conclusion as Zaidel et al. (2011). Recalibration rather relies on internal beliefs about the accuracy of the respective sensory modalities. In the Kalman Filter these are encoded in the widths of the priors for the bias estimates, whereby wider priors can be interpreted in the way that larger errorterms are more likely and it is thus more likely that this sense is inaccurate. Recalibration as a function of the ratio of these priors behaves in similar ways as FRA does in the study of Zaidel et al.

(2011). In fact, our model can be seen as an extension of FRA, that calculates the optimal ratio not only based on prior beliefs but also considering uncertainty in the sensory observations. Evidence for the importance of prior beliefs about sensory accuracies was also found by Di Luca, Machulla, & Ernst (2009), who showed that in audio-visual temporal recalibration, audition was recalibrated when the context indicated a high likelihood of biases in the auditory system (sound presentation via headphones.). In contrast, vision was recalibrated when visual and auditory stimuli were spatially collocated, arguing for the malleability of these prior beliefs, Burge et al. (2008) showed that visuo-motor calibration increases when the uncertainty of the visuomotor mapping is increased. They introduced a time varying bias between visual feedback and reach endpoint. Importantly, this variation was not random from trial to trial but correlated over time, which allowed a dissociation between mapping uncertainty and simple noise in the feedback signal. In our model, an increase in mapping uncertainty would correspond to a wider bias prior and similarly imply a larger CVAE. In summary, these results point to the interpretation that the perceptual system forms plastic estimates of uncertainty with respect to the accuracy of its sensory modalities.

To distinguish between $p(C = 1|y_k)$ – Weighting and no weighting, our study does not provide support for the former (*PEP* = 0 for $p(C = 1|y_k) -$ Weighting). A recent study that also manipulated visual reliability found support for the CI-model of recalibration, which refers to an exponential learner model with the VEDiff as errorterm and $p(C = 1|y_k) -$Weighting in our taxonomy (Hong et al., 2021). Our approach differs from the one in Hong (2021) in several aspects. Hong (2021) only considered a hand-selected subset of candidate models and most notably did not consider Kalman Filtering as learning mechanism and the MultDiff errorterm. Secondly, they did not explicitly model the adaptation phase, thereby the model estimations of the errorterm during adaptation phase were not constrained by actual data. Our results speak against $p(C = 1|y) -$Weighting which is a prerequisite for an increase of the CVAE as a function of visual reliability as found by Hong (2021).

Given that approximately 22 of our participants did not apply $p(C = 1|y_k) -$Weighting, whereas approximately 10 did in our study, and moreover 4 of 6 did apply $p(C = 1|y_k) -$Weighting in the study of Hong et al. (2021), the question arises whether there is actually one common model to explain the behavior of all participants under all conditions. Mahani et al. (2017) found that, at the beginning of their experiment, participants integrated audio-visual cues. However, as the experiment progressed, participants switched to a cue selection strategy, because a common cause was unlikely for their stimulus configuration, resulting in no recalibration. However, when auditory, visual, and tactile cues were used in a single session, recalibration occurred. A model-based analysis showed that this might have been due to an increased likelihood of a common cause in this condition overall. This flexibility lets us speculate whether the likelihood of a participant to apply $p(C = 1|y_k)$ −Weighting might depend on the probablility of different causal structures across experimental conditions. That means when differences in causal structures are salient for participants across conditions, they are more likely to apply $p(C = 1|y_k)$ −Weighting.

The auditory VE did not vary as a function of visual reliability in our study, in such a scenario the CI-model of recalibration (Hong et al., 2021; Sato et al., 2007) would have predicted an increase of the aCVAE, due to a higher posterior probability of a common cause, which we did not find. Surprisingly, the CI-Model provides nevertheless relatively good predictions for the aCVAE (Figure A. 1. Empirical results and posterior simulations of died out CVAE models. Auditory average CVAEs as a function of visual reliability (top vs. bottom row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left).. However, all VEDiff based models (including the CI-Model) underestimate auditory VEs to not overestimate the aCVAE when the visual reliability is low. Therefore, we want to stress the paramount importance to include bimodal trials in the fitting and model comparison procedures for multisensory recalibration. Only if the integration model throughout the recalibration phase is constrained by data, overfitting by tweaking the size of the VE is penalized by lower model evidence.

From a computational perspective the question, whether multisensory integration and recalibration are common, independent, or interdependent processes, is closely linked to the errorterm, with the MeasureDiff and UniDiff errorterm implying independence, the MultDiff errorterm implying interdependence and the VEDiff errorterm implying a common process. Since evidence was only found for errorterms based on multisensory integration (MultDiff and VEDiff) remapping on the level of sensory cues (MeasureDiff) and early unisensory perceptual estimates (UniDiff) is unlikely. This suggests that pre-existing biases in vision and audition have to be considered when investigating audio-visual recalibration and furthermore cumulative recalibration and integration are at least interdependent.

An interesting feature of the most likely MultDiff errorterm is that it predicts incomplete recalibration when the cue-conflict is resolved by integration whereas the Kalman

CI-model would always predict full recalibration. Thereby our model can even explain a total lack of recalibration when the reliability of the involved sensory cues is low (as observed in Vercillo, Burr, Sandini, & Gori, 2015). The idea that recalibration and integration are distinct but complementary processes, i.e. both in common resolve cue conflicts and the sum of both effects cannot exceed the size of the cue conflict, has been extensively discussed since its early proposals (Welch & Warren, 1980). Generally it has been assumed that in this case negative correlations between the magnitude of integration and recalibration should occur (Block & Bastian, 2011). Similarly, positive correlations were expected if recalibration depends on integration (Bruns et al., 2022). But in fact, our simulations demonstrate that for instance recalibration with the MultDiff errorterm and based on $p(C = 1|y_k) -$Weighting is a highly non-monotonic function of sensory reliabilities (skewed bell-shaped) i.e., it first increases, then ceils and finally decreases with decreasing reliabilities, whereas the VE constantly increases. Even though the processes are complementary in that respect, the type of correlation thus highly depends on the choice of experimental conditions. Depending on the experimental parameter negative, positive as well as zero correlations can occur. This non-monotonicity might further explain the lack of reliability dependence in a previous study (Rohlf et al., 2021). This highlights the importance of computational models to avoid misinterpretations of more descriptive statistics, such as correlations. Additionally, regardless of the CVAE model, the learning rates or bias priors are always individual parameters. Accordingly, to investigate the relation of VE and CVAE it is obligatory to manipulate their sizes within each participant. Otherwise, models can account for differences by fitting group wise differing learning rates and assuming that there is in fact a variance in the true population we will also be less likely to detect any meaningful correlations at all.

The present study found no evidence for any visual aftereffects, even if the visual stimulus was massively blurred along the azimuth. This points to that the superior reliability of the visual system might not be the reason for lack of studies reporting visual aftereffects. Rather, there might be hardwired limits to audio-visual spatial recalibration. A recent study investigating the CVAE in cataract reversal individuals however found visual aftereffects (Bruns et al., 2022) similar to those in visuo-vestibular recalibration. The authors argue that the lack of typical visual input during a sensitive period might have led to a strengthening of auditory influences in multisensory areas and after sight restoration, these additional connections might drive visual recalibration as a second mean to achieve internal consistency.

Similarly, additional functional or anatomical connectivity might allow for visual recalibration in visuo-vestibular recalibration.

The question of anatomical and functional prerequisites is tightly coupled to the developmental trajectory of multisensory integration and recalibration. There is an ongoing discussion, whether recalibration (Burr & Gori, 2012) or integration (Rohlf et al., 2020) occurs first throughout development. A step-wise developmental trajectory has been observed for audio-visual spatial perception, whereby integration reaches an adult-like state first, followed by instantaneous integration (the aIVAE) and finally cumulative recalibration around the age of 7 - 8 (Rohlf et al., 2020) occurs. Our finding that recalibration and integration are interdependent is well in line with this stepwise trajectory, since the MultDiff errorterm of the best fitting CVAE model is based on the output of multisensory integration suggesting that multisensory integration might be a prerequisite for recalibration. In consequence, it seems reasonable that integration might develop prior to cumulative recalibration.

We highlighted the importance of accurate cues for multisensory integration to be beneficial, which seems contradictory to the earlier emergence of integration. However, given that the CVAE can emerge after as little as 24 exposures (Frissen et al., 2012), we would argue that this type of rapid recalibration might not underly recalibration on a developmental trajectory. Assuming that the perceptual systems learning rates are fine-tuned to the volatility of specific sources of cue inaccuracies (Kording et al., 2007) the learning rate of the CVAE would be too high for developmental changes which rather emerge over month or years. In line with this argument, Knudsen (2002) presented evidence indicating that in juvenile owls, audio-visual spatial recalibration occurs alongside anatomical changes. However, in adult owls, this recalibration process is linked to physiological changes that involve GABAergic-dependent feedback interactions. Rohlf et al. (2020) argued, that similarly in children under the age of six, GABAergic-dependent circuits might not have been sufficiently developed. Kalman models of recalibration can account for these constraints for instance by setting narrow or even Dirac delta priors for biases, when functional or anatomical prerequisites for plasticity are lacking.

This highlights the semantic double role of priors in the Bayesian framework, especially if we want to relate computational models to the underlying neural circuitries. On the one hand, they operationalize the structural or anatomical connectivity, which is a prerequisite for functional connectivity (Friston, 2011) by for instance setting hard constraints whether functional connections exist or not, or setting upper limits for their efficiency. On the other hand, priors serve as a measure for plasticity of functional or effective connectivity. In our model, this plasticity is directly coupled with the uncertainty about the accuracy of a given sensory modality, i.e., high uncertainty implies high plasticity and vice versa.

With respect to the IVAE, the best fitting model assumes, that the perceptual system tries to estimate the true position of an audio-visual object in common with biases in the auditory and visual sensory system. This suggests a common process for instantaneous recalibration and multisensory integration. The IVAE seems to partially share neural resources (Park & Kayser, 2019) with the VE. Importantly, the regions that flexibly integrate auditory and visual information in bimodal trials (anterior parietal see Rohe & Noppeney, 2015), seem to be recruited for the IVAE. Hence, both the PRI model as well as the MultDiff model are in line with previous neurophysiological findings. The former because it assumes a single common process for the VE and the IVAE and the latter because it assumes that the IVAE is based on the output of integration.

The overall validity of IVAE models (Figure 19) is lower than for CVAE models (Figure 14), therefore we treat these results with caution. The lower validity is mainly driven by the fact, that although highly significant on a group level, the IVAE was inconsistent within participants across conditions. We would expect a higher variation of the IVAE within participants as it generally builds up and decays faster than the CVAE (Bosen et al., 2017, 2018). Hence, individual inconsistencies (e.g., aftereffects to the right, when the previous disparity was to the left) might be due to the volatile nature of the IVAE, and the model fitting procedure thereby prefers parameter sets that underestimate the auditory IVAE. Although the IVAE was underestimated, this was the case for all tested IVAE models - we considered a presumably exhaustive set of proposed models from the literature (

Figure 22), and the data was nevertheless clearly in favor of only two out of 34 models.

The close relation of the IVAE to the VE becomes especially apparent when contrasted with the CVAE (Figure 21). We observe a trend towards an interdependence of the VE and the CVAE, i.e., the VE provides the errorterms used for CVAE, but both processes are relatively independent, while on the other hand we observe a trend towards a single process underlying the VE and the IVAE. Moreover, our data are in favor of a CVAE, that is not based on $p(C = 1|y_k) -$Weighting, whereas the PEP is highest for models assuming $p(C = 1|y_k) -$ Weighting for the IVAE. In line with this result, Wozny et al. (2011a) found an increased IVAE when the probability of a common cause in a previous bimodal trial was high compared to low.

Especially early studies on the VE (Bertelson, Vroomen, et al., 2000; Bertelson & Aschersleben, 1998; Vroomen, 2001) made large efforts to confirm a perceptual basis of the VE. The following attempts to computationally describe the VE, IVAE and CVAE might therefore have focused on perceptual models. Just recently rather decisional contributions to the CVAE (Aller et al., 2022) and IVAE (Park et al., 2021) have been reintroduced to the discussion. Park (Park et al., 2021) found that although perceptual components seem to dominate the IVAE, previous responses of participants seem to contribute as well. Our data suggests that the IVAE is almost equally likely a response level or perceptual level phenomenon. In contrast to the assumption that the IVAE has contributions from both perceptual and response related processing stages, we only tested for one or the other. Otherwise, we would have needed to test each combination of perceptual and response related model leading to an unfeasible number of models. In fact, the indecisiveness with respect to the processing stage might be a hint, that both stages contribute to the IVAE in common. Our model comparison procedure might simply favor the dominant component for each participant. So far, perceptual models of the IVAE have always proposed two stage procedure, i.e., first multisensory estimates are calculated, then the IVAE is updated. A direct consequence of such two-stage models is however, that the IVAE should affect bimodal perception by a similar amount as unimodal perception (Figure 20). Although we found an effect of the IVAE on bimodal perception (Figure 19) it was strongly reduced compared to the effect on unimodal perception. The PRI model and response level models allow an instantaneous update of IVAE before a response is made in bimodal trials. This seems to be the main reason why these models outperformed any other suggested models, as they can explain the reduced IVAE in bimodal trials.

Conclusion

We confirmed a core prediction of the Causal Inference model: The occurrence of a visual VE when visual reliability is sufficiently degraded. Further we demonstrated that the weighting of visual and auditory cues is close to optimal in most of our participants. Our results provide a computational interpretation of the various neurophysiological findings implying partially overlapping networks underlying VE and CVAE (Bonath et al., 2007; Zierul et al., 2017), VE and IVAE (Park & Kayser, 2019) as well as IVAE and CVAE (Park & Kayser, 2021). All three phenomena are mutually interlinked as they make use of the information provided by multisensory integration. Although we found a reliability-dependence of the IVAE as well as of the CVAE, our computational approach allowed us to further dissociate these processes. We suggest that the CVAE might make use of multisensory information in order to update early auditory representations based on consistent evidence over time (Bruns, Liebnau, et al., 2011; Wozny & Shams, 2011a). Meanwhile, the best fitting PRI model of the IVAE, assuming IVAE and VE emerge in parallel as outcome of a single process, is well in line with a common neural substrate associated with later processing stages in anterior parietal regions (Park & Kayser, 2019). This common representation allows to flexibly adapt to external demands and contexts based on multisensory information, without rendering early auditory representations of space volatile over time. The closer computational link between IVAE and VE is compatible with the finding that the IVAE emerges prior to the CVAE during ontogenetic development (Rohlf et al., 2020). Moreover, as the IVAE is incorporated in the Causal Inference process, it becomes obvious that the ability to combine cues is a prerequisite for the IVAE, implying that multisensory integration should indeed emerge prior to trial-by-trial calibration (Rohlf et al., 2020). Finally, our results demonstrate that it is not simply the audio-visual physical discrepancy that drives audio-visual recalibration. Given that our modelling approach provides estimates of multisensory percepts, aftereffects and errorterms on a trial-by-trial basis, it provides a useful tool for future neurophysiological studies.

Chapter IV - Study 2

Causal Inference Differentially Affects Cumulative and Instantaneous Recalibration

Introduction

The process of segregating and integrating signals from an unknown number of sources across multiple sensory modalities is well described by Causal Inference (Körding et al., 2007). The perceptual system combines incoming sensory evidence with a-priori predictions into several perceptual estimates, each estimate reflecting a different potential causal structure of the observed event. These estimates are then combined based on their posterior probability. In its initial formulation the Causal Inference model assumed that sensory cues provide unbiased but noisy estimates of the external features. Importantly, these assumptions introduced two potential sources for cue discrepancies, auditory and visual spatial cues were suggested to be discrepant either due to noise, or due to having two spatially separable sources. The Causal Inference model only provided optimal estimates of external features and the causal structure of an event if sensory cues were unbiased. This simplification is faulty in general, but especially incorrect with regard to audio-visual spatial perception, since auditory and visual perceptual space are often distorted (Badde, Navarro, & Landy, 2020; Odegaard, Wozny, & Shams, 2015, also see Chapter II).

Accordingly, several extended formulations have been proposed to model how the perceptual system accounts for biased sensory cues (Badde et al., 2020; Sato et al., 2007, see also Chapter II). Empirical studies suggest that the perceptual system is in fact capable of recalibrating biased sensory cues over a wide range of sensory modalities (Badde et al., 2020; Bruns, Spence, et al., 2011; Mendonça et al., 2015; Zaidel et al., 2011). However, in healthy adults usually only auditory spatial cues are recalibrated in audio-visual spatial perception (Bruns, 2022, see also Study 1 and 3).

. Lewald et al (2002) observed visual response shifts but argued that these were not induced by audio-visual recalibration but rather due to the unbalanced spatial distribution of the visual stimuli during recalibration. While it is a long-established finding that, audio-visual spatial integration is based on Causal Inference, debates are ongoing how audio-visual spatial integration is related to recalibration and further whether recalibration follows the principles of Causal Inference. From a normative perspective, recalibration should only occur when an audio-visual discrepancy is observed, although a common cause is likely, and this discrepancy is not only due to noise. Study 1 (Chapter III) and a recent study (Hong et al., 2021) provide seemingly contradicting results with respect to whether the posterior probability of a common cause ($\mathrm{p(C = 1|y_k)}$) modulates the magnitude of recalibration (Hong et al., 2021) or not (Study 1).

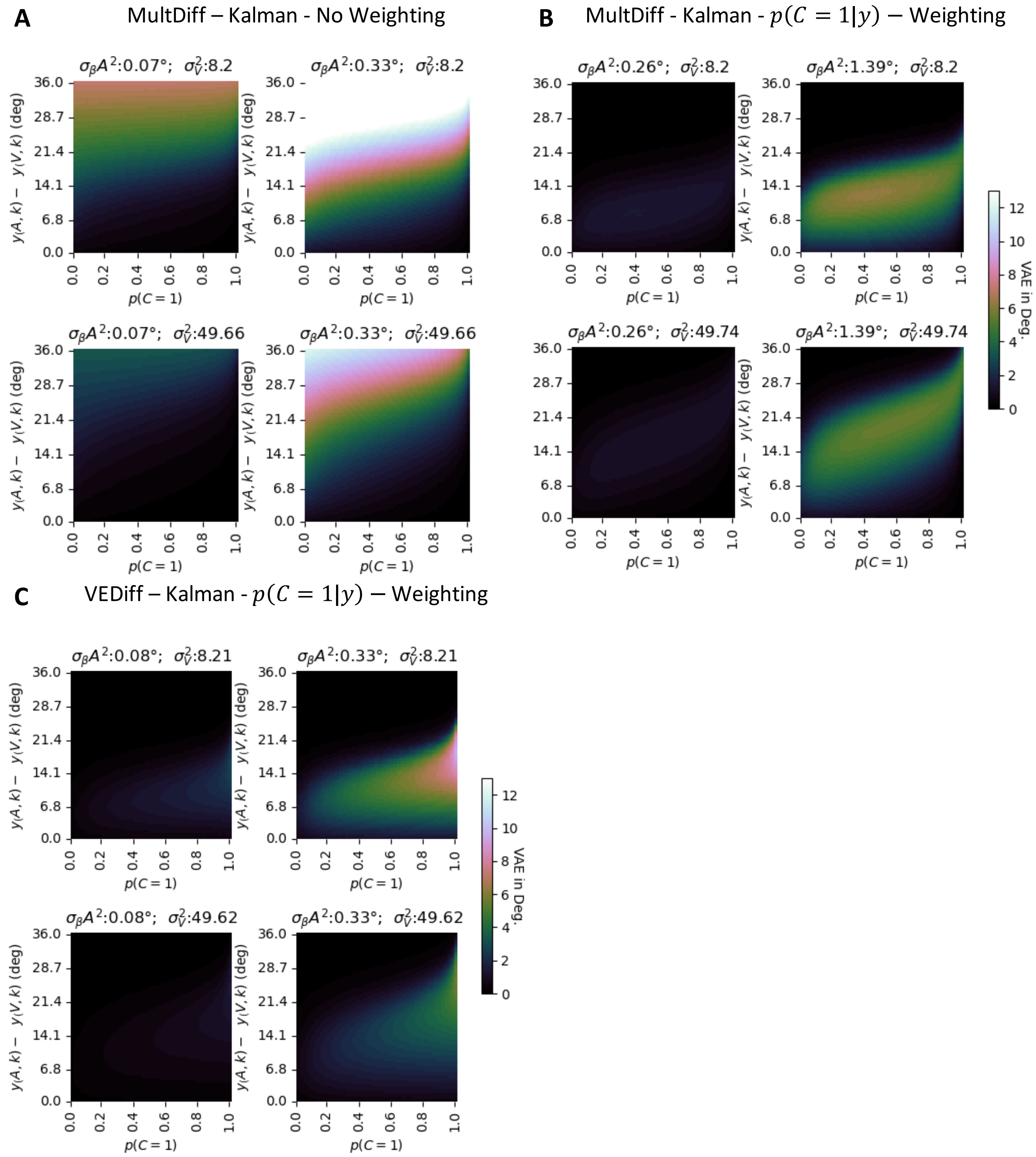


Figure 23. Simulation results for the most frequent aCVAE models of Chapter II. Simulations were based on the trial numbers of the behavioral experiment; the trial sequence was randomly chosen from one of the participants. Average parameters of Study 1 were used for simulations. Average auditory aCVAEs across 96 unimodal and 96 bimodal trials as a function of $p(C = 1)$ and physical audio-visual discrepancy. Graphs in the upper left corner from each panel show simulations for the average parameter estimates of $\sigma_{\beta A}^2$ and for the average σ_V^2 in the high visual reliability condition of Study 1. Graphs on the right show the results for $Mean(\sigma_{\beta A}^2) + 0.5 * SD(\sigma_{\beta A}^2)$. Graphs in the lower row show the results for the average σ_V^2 in the low visual reliability condition of Chapter II. Descriptions of model factors and parameter definitions are given Chapter II.

Both studies manipulated the visual reliability. Hong et al. (2021) found that with decreasing visual reliability, several participants showed increased recalibration, which is well in line with $p(C = 1|y_k)$ -Weighting and direct dependence between integration and recalibration. By contrast, in Study 1 with decreasing reliability the magnitude of recalibration also decreased, rather in line with bayes-optimal recalibration based on the residual difference between auditory and visual percepts (see Chapter I for an overview of possible models of recalibration).

However, none of the models could explain the behavior of all participants within each study or even across studies. This highlights the difficulties of model identification when the behavioral outcome is a product of multiple interacting processes. Even in Study 1 a minority of participants was best described by $p(C = 1|y_k)$ - Weighting. Additionally, Study 1 used a design that was capable to dissociate long-term recalibration, emerging after consistently perceived spatial misalignment, from immediate recalibration. The latter appears already after a single exposure (Wozny & Shams, 2011a). Both types of recalibration are likely distinct phenomena (Bruns, 2015; Park, 2021; Rohlf, 2020; as well as Chapter II). Furthermore, instantaneous recalibration was better described by $p(C = 1|y_k)$ - Weighting in Study 1.

This leads to the question: which factors influence the effect of different causal structures on recalibration? Picture a junction with an electric car waiting right next to a truck. When both cars accelerate, the motor of the electric car might emit a high frequency swoosh whereas the truck might emit low frequency noise and rattle. Although both auditory signals provide noisy spatial estimates and could thereby easily be attributed to the false car due to their spatial proximity, the likelihood of mixing up sound sources will be very small. This is due to the fact that the perceptual system makes use of its prior knowledge about the distinctive sound of each car. On the other hand, if there were two similar trucks standing next to each other it would be virtually impossible to associate each sound with the correct car. If they turned into different directions, however, the sounds would become separable, and each could be attributed to the correct car again. This example illustrates the two components determining the perceived causal structure of the scene: prior knowledge and sensory evidence. In this study we aim to clarify the importance of the causal structure in instantaneous and cumulative recalibration by manipulating the prior knowledge and sensory evidence concurrently.

We assumed that the prior knowledge of a common cause is malleable by experience and can be tuned to the true statistics of the world (Chapter I, *Learning Causal Priors).* To change the prior, we adopted an association paradigm (Tong et al., 2020) in which one stimulus pair is consistently presented at the same time and position whereas another stimulus pair is presented with temporal and spatial discrepancies as well as the auditory and visual components of this pair in isolation. Tong et al. (2020) showed that the audio-visual integration is increased for consistently associated stimulus pairs compared to inconsistently paired stimulus pairs, implying different causal priors. The sensory evidence is manipulated by using either a small audio-visual discrepancy or a large audio-visual discrepancy.

We expected that we can replicate the findings of Study 1, i.e., that instantaneous recalibration is based on $p(C = 1|y_k)$-Weighting and a single process with integration. Moreover, we hypothesized that if participants were able to flexibly switch between $p(C = 1|y_k)$-Weighting strategies and no Weighting strategies, based on the saliency of differences in the causal structures across conditions, most participants should apply $p(C = 1|y_k)$-Weighting also for cumulative recalibration in this study. Importantly, the interactions between the prior probability of common cause and the spatial discrepancy can be highly non-monotonous (Figure 23) e.g., participants with initially high values of p-common might almost fully integrate, leading to small errorterms and little recalibration. For these participants, reducing $p(C = 1)$ might lead to larger errorterms and thereby larger recalibration. Participants with initially low internal values of $p(C = 1)$, might on the other hand not recalibrate at all, as the errorterm is down weighted close to zero by the low $p(C = 1|y_k)$. In this case increasing $p(C = 1)$ would lead to larger recalibration. Hence the directionality of our manipulation can be interindividually different. Consequently, we relied on a participant-wise model comparison approach which we introduced in Chapter II to test our main hypothesis.

Methods

Participants

To counterbalance all conditions (see *Procedure* for details), we aimed for a sample size of 32 participants. Moreover, this sample size yields a power of 0.8 to detect a medium-sized effect (d_z = 0.56) for a directional difference between two within-subject conditions at an α level of .013. Hence, correcting separately for VE, IVAE and aCVAE, we were able to test for an effect of disparity, type of association and the interaction with a-priori comparisons on a global α level of .05 (assuming Bonferroni correction). The power analysis was conducted in G*Power 3.1 (Faul et al., 2009).

Localization accuracy was assessed directly after baseline measurements (see *Exclusion Criteria* for a detailed description of the exclusion procedure). Participants who did not fulfill the inclusion criteria did not proceed with the subsequent sessions and were immediately replaced. A total of 41 participants were recruited of which we had to exclude 9 participants based on our accuracy criteria. Another participant had to be excluded from further analysis due to data loss. The remaining 31 participants (14 female, 17 male, mean age=27.4 years, age range=20-42 years; 3 left-handed) all reported normal or corrected to normal vision, had no history of visual, auditory or neurological impairments and did not use any medication known to affect perception. The former information was collected via a questionnaire. All participants were recruited through an online subject pool of the University of Hamburg. Written informed consent was obtained from all participants prior to taking part. The study was performed in accordance with the ethical standards laid down in the Declaration of Helsinki (revised from 2013). The procedure was approved by the ethics commission of the Faculty of Psychology and Human Movement of the University of Hamburg.

Apparatus

The study was conducted in a sound-attenuated darkened room. Auditory stimuli were presented with six speakers which were mounted on a semicircular frame (90 cm radius) and covered by an acoustically transparent curtain. Participants were seated in the center of the frame and positioned their head on a chin rest at the level of the speakers. The speaker positions ranged horizontally from -22.5° (22.5 left from straight-ahead) to 22.5° (22.5 right from straight-ahead) in steps of 9° (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). Visual stimuli were presented via a virtual reality head mounted display (HTC VIVE, HTC Corporation, New Taipei City, Taiwan), with a resolution of 1080 ×1200 px, a field of view of 110° and a refresh rate of 90 Hz. Two HTC base stations were used for tracking providing submillimeter resolution and sufficient tracking accuracy for behavioral research (Niehorster et al., 2017). An HTC VIVE controller was used as pointing device for participants. Furthermore, three HTC trackers were used to control alignment between virtual and physical space. The trackers were positioned at three reference points in physical space. Based on the corresponding virtual points and the position of the trackers in virtual space, rotations and transformations were calculated via singular value decomposition, and the virtual space was rotated and translated accordingly. An eye tracker (Pupil Labs HTC Vive Add-On) was installed in the head mounted display (HMD) to control for open eyes. The virtual setup was a rectangular room with the same dimension as the physical room. A virtual spherical surface spanning 180° horizontally and 60° vertically served as projection plane for visual stimuli.

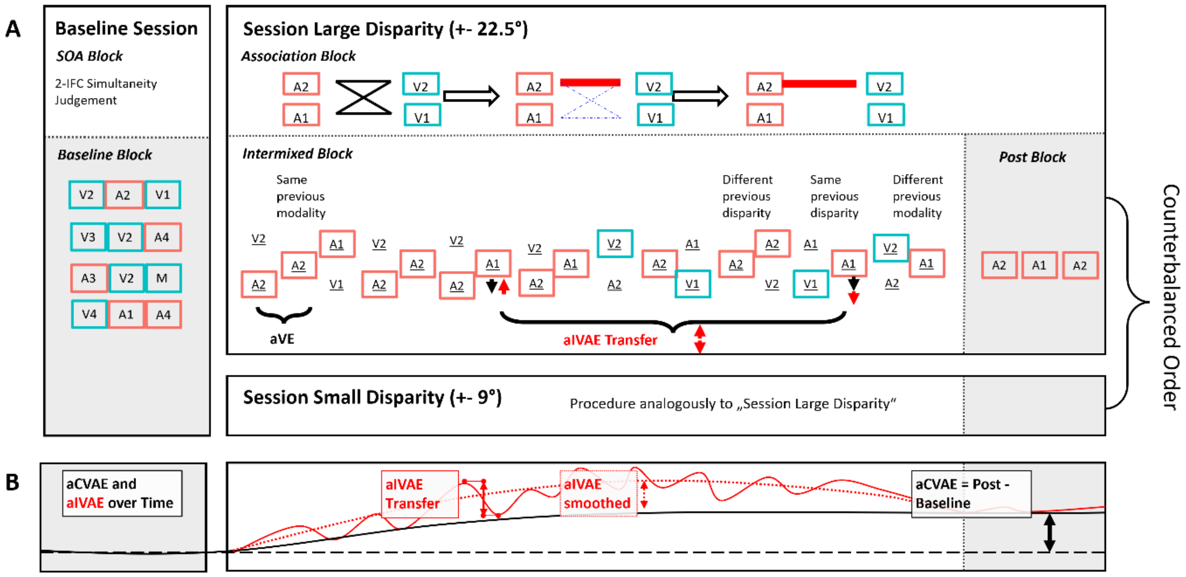


Figure 24. Experimental procedure and hypothetical time-courses of aCVAE and aIVAE across Blocks. **A**: Experimental Procedure: The figure is chronologically ordered from left to right and top to bottom. The experiment started with a simultaneity judgement task to determine audio-visual temporal precision. Baseline blocks were done to assess unimodal reliabilities and fit linear models between physical position and perceived position for all stimuli. The second and third session started with an association block (Figure adapted from Tong et al., (2020)) followed by two intermixed (white background) and post (grey background) blocks (Figure adapted from Bruns, 2019). Blocks alternated between intermixed and post blocks. Two tones of different sound frequencies (A1 and A2) were paired with visual stimuli of different color (V1 and V2). Importantly each pair (A1V1 and A2V2) was presented with an audio-visual discrepancy in opposite direction inducing a aCVAE in opposite directions. On a trial-by-trial basis the direction of the preceding discrepancy changed allowing to evaluate the transfer of the IVAE. Shifts in audio-visual trials were induced by aVE, aCVAE and aIVAE in common. Throughout the second and third session the absolute disparity was either large or small (the assignment of session order was counterbalanced). **B**: Hypothetical audio-visual shifts induced by aIVAE and aCVAE. Shifts in unimodal trials in intermixed blocks were induced by aCVAE (black) and aIVAE (red) in common, whereas the shifts in post blocks were dominated by the aCVAE. In contrast to previous studies (Bruns & Röder, 2015; Rohlf et al., 2021), we expected that the aIVAE would on average be different from zero (as in (Bosen et al., 2018; Watson et al., 2019)), either due to partial transfer or the association manipulation might induce aIVAEs of different sizes for the different pairs so that the aIVAEs do not sum to zero.

Stimuli

The auditory stimuli were sine tones with four different frequencies (445, 890, 2000, or 4000 Hz) and a white noise burst. Sounds were presented for 78 ms (approximately equal to 7 frames on a 90 Hz display) including 5 ms on- and off-ramps. The spacing of the center frequencies was chosen to assure that at least one critical frequency bandwidth (Zwicker et al., 1957) lies between two subsequent frequencies. The stimulus intensity was randomly varied over a 4-dB range centered at 70 dB(A) to minimize potential differences in the speaker transformation functions.

Visual stimuli were homogeneously colored circles projected onto the virtual plane with 1° visual angle radius. We used five different colors (0°~white, 45°~orange, 135 °~green, 225°~blue, 315°~magenta) defined in the plane spanned by the L-M and S-(L + M) axis of the DKL-Color-Space (Derrington et al., 1984). The choice of colors ensured that none of the colors solely stimulated the s-cone-channel and therefore might bypass the superior colliculus, which has been shown to be important for multisensory spatial orienting (Leo et al., 2008). We used a contrast of 1 and luminance 90° for the white stimulus and a contrast of .5 and luminance of 15° for all other colors.

The colors were calibrated to be approximately isoluminat across participants with a spectrophotometer (i1Publish Pro 3, X-Rite Inc., Michigan, United States). The white stimulus (response noise stimulus) was used to estimate any noise in the pointing responses that was not due to perceptual noise (i.e. non-perceptual noise, NPN, see Tassinari et al., 2006 for a similar procedure).

To indicate to the participants whether to localize the visual or the auditory component of the trial the letter “A” (6°x6° visual angle) surrounded by a black square indicated to localize the auditory component of the trial. A white letter “V” (6°x6° visual angle) indicated to localize the visual component of the trial. A HTC Vive controller was used to collect localization responses of the participants.

Procedure

The study was split into three sessions on separate but not necessarily consecutive days, each session lasting about 3h, see Figure 24 for schematic depiction of the procedure. In the first session we assessed unimodal visual and auditory localization baseline accuracy and reliability. Thereby, we were able to estimate linear distortions of auditory and visual perception at baseline (compare Chapter II and Hong et al., 2021). Additionally, we used a two-interval forced-choice (2-IFC) paradigm to estimate the temporal precision of participants in synchrony perception, that was later used for the model-based analysis. Session 2. and 3. were used to induce the actual experimental manipulations in intermixed blocks that contained unimodal visual and auditory trials as well as bimodal audio-visual trials. We grouped the auditory stimuli in two pairs (445Hz/2000Hz, 890Hz/4000Hz) and one pair was used for each session to avoid carry-over effects between sessions (Bruns & Röder, 2019).

For each session two fixed audio-visual pairs were formed consisting of one of the colored visual stimuli and one of the auditory stimuli. To one of the audio-visual pairs a spatial discrepancy to the left was assigned (-1*absolute disparity) and to the other a spatial discrepancy to the right (absolute disparity). This spatial discrepancy was used to induce the cumulative as well as the immediate ventriloquism aftereffect for both pairs. However, before we induced VE, IVAE and aCVAE one stimulus pair was consistently presented in space and time (common cause association, CCA) and the other was inconsistently presented in space in time, i.e., with large spatial and temporal disparities. This manipulation has led to changes in the a-priori binding tendency in a previous study (Tong et al., 2020). Moreover, we varied the absolute audio-visual discrepancy (9° and 22.5°) between sessions in intermixed blocks. From the CI-model it follows that this manipulation should affect the posterior probability of a common cause.

From the four possible colors two were used for each session. Similarly, we always used one low frequency tone (445Hz and 890Hz) and one high frequency tone (2000Hz and 4000Hz) in each session to guarantee good discriminability between sounds. This factor was counterbalanced with the conditions of interest, i.e., the absolute audio-visual discrepancy and the type of association. All levels of absolute audio-visual disparity and type of association were realized within participant. The signed audio-visual discrepancy and type of association were counterbalanced across participants. Moreover, we counterbalanced the color used for the visual stimulus with the type of association (CCA or DCA) and absolute audio-visual discrepancy (9° or 22.5°). The absolute audio-visual discrepancy was additionally assigned to the number of sessions (2. or 3.) in a counterbalanced manner.

First Session

Headset Calibration

To assure that all visual stimuli are presented in the binocular visual field of the headset, we presented four visual stimuli in the periphery at +/- 15° altitude and +/- 33° azimuth. Importantly, stimuli on the right side were only presented on the left-eye-display and vice versa for left-side-stimuli. Moreover, stimulus positions were fixed in head-centered coordinates i.e., stimuli outside the visual field remained outside, when the head was moved. The color of the stimuli was randomly drawn from orange, green, blue, or magenta and participants were asked to report the color in the order top-right, top-left, bottom-right, bottom-left. Hence, when participants were able to report all colors correctly, it was assumed that all visual stimulus positions (range -31.5° to 31.5°) were in the binocular field of view of the HMD. Otherwise, the HMD position on the head was readjusted and the procedure was repeated. After proper adjustment all participants were able to perform this task. This procedure was repeated after each break, every time a new experimental block started and each time the participant removed the headset or reported a slip.

Temporal Precision Block

In this block we assessed the precision of audio-visual temporal perception in detecting stimulus onset asynchronies (SOA). We used a 2-IFC task with one standard. The standard as well as all comparisons consisted of the white visual stimulus and a white noise burst. Visual and auditory stimulus components were always presented at 0° azimuth and 0° altitude with a duration of 78ms. While the standard was always presented with 0ms SOA, 10 log-spaced SOAs were used for the comparison (+/- 356ms, +/- 289ms, +/- 200ms, +/- 111ms, +/- 22ms), whereby negative values indicate a leading visual stimulus. The interval (first or second) for the standard was chosen randomly and after the second interval, participants were asked whether the synchronous stimulus-pair had occurred in the first or second interval. Participants completed 30 trials per SOA yielding 300 trials overall. Based on the relative number of correct and incorrect trials psychometric functions (cumulative gaussian) were calculated and the temporal precision was calculated as half the variance of the fitted gaussian.

Baseline Block

In this experimental part we evaluated the initial accuracy and reliability of localization responses to unimodal visual and auditory stimuli. Nine different stimuli, i.e., four visual stimuli, four auditory stimuli and the response noise stimulus were presented from six positions (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) 16 times. In total this part consisted of 864 trials in randomized order which were split in four equally sized blocks à 216 trials. Between each block a pause of approximately 5 minutes was introduced.

Each trial started with the presentation of a fixation cross at 0° azimuth surrounded by a circle. A virtual laser pointer providing feedback about the pointing direction was turned on and participants had to point towards the fixation point using the controller and pull a trigger button on the controller. Only if the pointing direction did not deviate more than 1° from the direction of the fixation point the trial started. The circle around the fixation cross disappeared and the laser pointer was turned off. After a random delay (400ms to 600ms) the experimental stimuli were presented for 78ms and 100ms after stimulus offset a letter appeared (A for auditory experimental stimuli and V for visual experimental stimuli). Responses were allowed immediately after the letter appeared and the laser pointer was turned on again. Participants were instructed to respond as accurately but nevertheless as promptly as possible by directing the laser pointer to the location where they had perceived the stimulus and confirming by pulling the trigger. Moreover, we informed participants that all stimuli would be presented at the same height and that they should focus on horizontal pointing accuracy. The procedure diverged for response noise stimuli in so far as the white visual stimulus was displayed persistently until the participant confirmed the response.

Second Session

Association Block

To induce differences in the prior probability of a common cause for the two audio-visual stimulus pairs used in each session we started session 2. and 3. with an association block following the paradigm of Tong et al. (Tong et al., 2020). With regard to the CCA pair, auditory and visual components were always presented together with zero SOA and zero audio-visual disparity at the 2 inner stimulus positions (+/- 4.5°). For each position 96 trials per Block were presented. Regarding the DCA pair, 48 trials were purely auditory, another 48 trials were purely visual, and 96 trials were audio-visual trials per block. Unimodal stimuli as well as the auditory component of the audio-visual stimulus were presented equally often at the 2 inner stimulus positions (+/- 4.5°). The visual component of the audio-visual stimulus was displaced by +/- 13.5°, +/- 18° or +/- 22.5° from the auditory component and the SOA was randomly drawn from the two intervals +/- (300ms, 500ms). In total, participants completed 768 association trials. We intermixed three voluntary pauses of up to 5 minutes that split the 768 trials in 4 parts of equal size. Throughout an association block, participants had to perform a control task that assured that they were attending the stimuli. After 96 randomly chosen regular association trials, participants had to report whether the last two visual stimuli were of the same color:

„Did the last 2 colored dots have the same color?” (Translated from German)

or the last two auditory stimuli had the same pitch:

„Did the last 2 sounds have the same pitch?” (Translated from German).

Both questions were equally likely. The question was presented on a virtual canvas. Two buttons appeared above and below this canvas and each was randomly assigned to the “yes” or “no” response indicated by “yes” or “no” labels. Participants had to point to the button and press the trigger to log their response. The buttons lit up when properly targeted.

Intermixed Block

During this experimental part we induced the VE, aCVAE and IVAE via bimodal stimulation and quantified these effects via localization responses. This was achieved by making use of an adapted version of the paradigm by Bruns & Röder (2015), which relies on the assumption that the aCVAE is frequency-specific and therefore could be induced in different directions for different tones. We refer to Chapter III for a detailed description. Distinct audio-visual discrepancy (either to the left or right) were assigned to the two audio-visual pairs (CCA pair and DCA) and in audio-visual trials visual stimuli were consistently displaced in the corresponding direction across all intermixed blocks of a session. Hence, VEs and aCVAEs in different directions were induced for the CCA pair and the DCA pair. In addition, the two auditory components of the CCA pair and the DCA pair were presented as unimodal trials (analogously to the baseline blocks) intermixed in the bimodal trials. The order of the trials was pseudo randomized in a way that assured an equal number of unimodal trials (4) where the one, two, three or four preceding audio-visual trials had the same unimodal component (previous disparity same) or the different unimodal component (previous disparity different). As each of the two sounds were statically linked with opposing audio-visual discrepancies, this automatically implied that each sound was equally often preceded by the same audio-visual disparity it was linked to and the opposing disparity. This led to eight types of possible sequences of differing length. Assuming, that the IVAE transfers across frequency, the trial-wise variation of the preceding discrepancy should lead to a trial-wise modulation of the IVAE (Bruns & Röder, 2015; Wozny & Shams, 2011b).

Only the 4 inner stimulus positions (+/- 13.5°, +/- 4.5°) were used as auditory stimulus positions in unimodal and bimodal trials. During half of the bimodal trials, vision was task relevant. Participants were instructed to memorize visual and auditory positions during bimodal trials until after 100ms a letter appeared (A for auditory, V for visual) that indicated whether to point to the auditory stimulus (Audition task-relevant) or to the visual stimulus (Vision task-relevant).

We presented each of the two bimodal stimuli 24 times for each auditory position, i.e., 96 times in total. In half of the bimodal trials vision, was task-relevant whereas in the other half audition was task-relevant. Similarly, each unimodal stimulus was presented 24 times per position, i.e., 96 times in total. Overall, each intermixed block contained 384 trials and participants completed 2 intermixed blocks (i.e., 768 trials) per session. Each block was split into two sub-blocks of approximately equal length. The length differed slightly as pauses were set between sequences in order to not split them across two blocks.

Post Block

Each intermixed block was followed by a post block in which the two auditory stimuli from the preceding intermixed block were presented. For these stimuli, the procedure was equivalent to the baseline blocks. In comparison to Bruns et al. (Bruns & Röder, 2015) and Chapter III, the posttest was introduced because the IVAE is known to dissipate rapidly over time (Bosen et al., 2017, 2018), thereby these blocks provide an estimate of the aCVAE that is not affected by the IVAE (see Figure 24). This is important as we did not expect the IVAE to be of equal size for the CCA and the DCA pair. For instance, bimodal CCA trials might lead to larger shifts than bimodal DCA trials due to a larger posterior probability of a common cause on average. It follows that unimodal trials are also on average shifted more in the direction of bimodal CCA trial than bimodal DCA trial solely due to the IVAE. Hence the mean of the unimodal trials in intermixed blocks is affected by the IVAE and the aCVAE (see Figure 24) and thereby not an unbiased measure of the aCVAE (as used in Bruns & Röder, 2015; Rohlf et al., 2020, 2021).

Model-free Analysis

Exclusion Criteria

For all unimodal stimuli of the baseline block, we fitted separate linear models to the localization responses with the actual stimulus positions as predictor. The obtained estimates of slope and intercept served as indicators of localization accuracy with a slope of one and an intercept of zero indicating perfect accuracy. We a-priori defined an absolute intercept above 4.5° or a slope out of the range (0.5, 2.0) as criteria for inaccurate localization. These accuracy criteria were evaluated immediately after participants had finished the unimodal baseline blocks. For each stimulus where the criteria were not fulfilled, we repeated the baseline measurements once in a blocked manner, i.e., one block only included one unimodal stimulus. If the criteria were still not fulfilled for any of the stimuli, the session was aborted, and the participant was not tested any further. Due to the accuracy criteria, in total 7 participants had to be replaced by a newly recruited replacement participant. None of the replacement participants failed the accuracy criteria.

Unimodal Bias

Especially in bimodal trials in intermixed blocks, localization responses might be affected by IVAE, aCVAE and VE in common. Therefore, we will speak of unimodal and bimodal biases when we refer to the raw localization responses, instead of VE, CVAE and IVAE. However, when meaningful decomposition of the biases in components likely reflecting aCVAE, IVAE and VE is possible, we will use these terms.

We formulated one large linear mixed model (LMM) to cover the whole experimental paradigm. First, we calculated trial-wise estimates of sensory biases by fitting linear models separately for each unimodal stimulus to the localization responses of the baseline blocks with veridical stimulus position as predictor. From these linear models we predicted localization responses for the responses in the post and intermixed blocks. Trial-wise residuals (Res) were obtained by subtracting the predicted responses from the actual responses in each trial and multiplication with the sign of the audio-visual disparity. Hence, residuals in the same direction as the audio-visual disparity were positive and negative in the reverse case. The model encompassed block type (post or intermixed), the block number (1 or 2), the stimulus type (bimodal or unimodal), the absolute disparity (larger or small), the previous disparity (same or different) as well as the association type (CCA or DCA) as fixed effects. Bimodal trials were only realized in intermixed blocks, hence stimulus type as well as previous disparity were nested in block type. We will refer to residuals in bimodal trials as Res_{BI} and Res_{UNI} for unimodal trials. This leads to a very large initial LMM:

$$\begin{gathered} Res \sim association\ type * block\ number * block\ type * abs.\ Disp. + \\ association\ type * block\ number\ * block\ type * abs.\ Disp. \\ * previous\ Disparity\ * stimulus\ type + (1 \mid Participant) \end{gathered} \tag{44}$$

Therefore, we performed stepwise model reduction (Hastie & Pregibon, 2017) to drop non-significant fixed effects, the final LMM can be found in Table B. 1. The advantage of one large LMM lies in the fact, that it allows to directly compare the size of unimodal and bimodal biases across blocks and thereby define estimates of aCVAE, IVAE and VE based on contrasts in the LMM. The aCVAE can be evaluated based on the EMMs of the Res_{UNI} in the post blocks. The IVAE can be evaluated as a pairwise contrast (IVAEcontrast) of block type (induc -post) conditioned on unimodal stimuli (i.e., for Res_{UNI}). The trial-wise influence of the IVAE across sound frequencies can be evaluated by a pairwise interaction contrast, were we first calculated the IVAE contrast for the same previous disparity condition and different previous disparity condition separately and then calculated the difference contrast. Moreover, the VE can be evaluated as a pairwise contrast of stimulus type (bimodal – unimodal) conditioned on intermixed blocks, which essentially decomposes the bimodal bias in a component that is also present in bimodal trials and an exclusively bimodal component. Effects of our main experimental manipulations (absolute disparity and association type) could be evaluated in terms of interaction contrasts, by conditioning the VE, IVAE, and CVAE contrast on the levels of absolute disparity or association type and testing the difference between these levels. To keep visualizations and statistical analysis consistent, we generally report EMMs instead of means. Participant-wise data for graphs is approximated by participant-wise linear models including only fixed effects.

Quantification of the Build-up of the aIVAE

To quantify the trial-wise build-up or decay of the aIVAE, a separate LMM was formulated including only auditory trials of the intermixed blocks. An additional factor consecutive Trials (1, 2, 3 or 4) was introduced, indicating how many consecutive bimodal trials contained the same audio-visual disparity as the immediately preceding audio-visual disparity.

$$\begin{aligned} Res \sim\ & \mathit{block\ number} * \mathit{abs.\ Disp.} * \mathit{previous\ Disparity} * \mathit{consecutive\ Trials} \\ & + (1 \mid \mathit{Participant}) \end{aligned} \tag{45}$$

All LMMs were analyzed with type II ANOVAs using Wald chi-square tests (Fox & Weisberg, 2019).

Model-based Analysis

The behavioral manipulations of absolute disparity and association type aimed at delivering a context in which the causal structure of audio-visual stimuli differs noticeably for participants. Two recent studies (compare (Hong et al., 2021) and Study 1 in Chapter III) found seemingly differing results regarding the importance of the causal structure for the aCVAE. Our modelling analysis therefore focusses on the role of the causal structure for aCVAE and IVAE, by making use of the model taxonomy of Chapter II.

Multisensory Integration

Assuming that audio-visual integration is best described by the Causal Inference model (CI-model), a recent version of the Causal Inference model described in Chapter II (see Hong et al., 2021; Odegaard et al., 2015b for similar procedures) was used that allows for linearly biased multisensory perception. More detailed we used the Switching Kalman Filter (Murphy, 1998) formulation of the CI-model in Chapter II, which can elegantly incorporate more than two observations. Here we will mainly focus on the adaptations made to account for the fact that we do not only vary auditory and visual veridical positions but also the temporal offset between visual and auditory stimuli similar to the spatio-temporal Causal Inference model of McGovern et al. (2016).

In each trial the observer makes visual (y_V) and auditory (y_A) noisy measurements of the veridical position and temporal offset between stimuli ($y_{\Delta t}$). Based on these measurements the system calculates estimates for the scenario of a common cause (C=1) or two distinct causes (C=2). The generative model for the measurements is given by:

$$y = \boldsymbol{H}x + \boldsymbol{H}_\beta \beta + z \quad with\, z \sim \; MVN(0, \boldsymbol{R}) \tag{46}$$

Now $y = \begin{pmatrix} y_A \\ y_V \\ y_{\Delta t} \end{pmatrix}$ and $x = \begin{pmatrix} x_A \\ x_V \\ x_{\Delta t} \end{pmatrix}$ are vectors of the measurements and veridical positions, H is a 3x3 Matrix that defines a linear relationship between veridical features (i.e. positions and temporal offset) and measurements. Similarly, $\beta = \begin{pmatrix} \beta_A \\ \beta_V \\ \beta_{\Delta t} \end{pmatrix}$ is a vector of constant biases and H_β is a 3x3 Matrix that defines a linear relationship between constant biases and measurements, hence we will use the identity matrix. R is the 3x3 covariance matrix $\begin{pmatrix} \sigma_A^2 & 0 & 0 \\ 0 & \sigma_V^2 & 0 \\ 0 & 0 & \sigma_{\Delta t} \end{pmatrix}$ of the measurement noise $z = \begin{pmatrix} z_A \\ z_V \\ z_{\Delta t} \end{pmatrix}$. In short, formula (46) implies that sensory measurements reflect the true states of the world in a linear distorted way plus some random noise. The optimal estimates for each scenario are given by:

$$\hat{x}_{k,\boldsymbol{C=c}} = \hat{x}_{k|k-1} - \boldsymbol{K}_{k,C=c}\hat{e}_{k|k-1} \tag{47}$$

For the distinct causal scenarios C {1,2}. The $\boldsymbol{K}_{k,C=c}$ denotes the optimal Kalman gain and $\hat{e}_{k|k-1}$ is an errorterm based on the difference between measurements in a trial k and predicted measurements. The predicted measurements are based on prior distributions for the true states x. If there are two causes the prior is gaussian with diagonal covariance matrix $Q_{k-1,C=2} = \begin{pmatrix} \sigma_{qS}^2 & 0 & 0 \\ 0 & \sigma_{qS}^2 & 0 \\ 0 & 0 & \sigma_{q\Delta t}^2 \end{pmatrix}$, i.e. auditory, visual and temporal expectations are independent. If there is one cause, we derive $\boldsymbol{Q}_{k-1,C=1}$ by setting an intermediate matrix $\begin{pmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_m^2 \to 0 \end{pmatrix}$ $\begin{pmatrix} \sigma_{qS}^2 & 0 & 0 \\ 0 & \sigma_m^2 \to 0 & 0 \\ 0 & 0 & \sigma_{q\Delta t}^2 \end{pmatrix}$ and rotating the first two rows and columns by 45° (Shams & Beierholm, 2011). This means the perceptual system expects almost perfectly correlated auditory and visual positions. In the standard Kalman Filter dynamic transitions over time are considered but here we assume that the mean of expected states stays fixed at zero. The posterior probability in a trial k of each scenario can be calculated by:

$$p(C = c|y_k) = \frac{p(y_k|C = c)p(C=c)}{\Sigma_l p(y_k|C = l)p(C=l)}$$

, whereby

$$p(y_k|C = c)p(C = c) = p(\hat{e}_{k|k-1}|C = c)p(C = c)$$

$$p(\hat{e}_{k|k-1}|C = c)p(C = c) = \varphi(\hat{e}_{k|k-1}, \boldsymbol{S}_{k,C=c})\, p(C = c) \tag{48}$$

with $p(\hat{e}_{k|k-1}|C = c) = \varphi(\hat{e}_{k|k-1}, \boldsymbol{S}_{k,C=c})$ were $\varphi(x, \Sigma)$ is the density function of a MVN with covariance Σ at x. The CI-model estimates of participants percepts are then given as:

$$\hat{x}_k = p(C = 1|y_k)\hat{x}_{k,C=1} + p(C = 2|y_k)\hat{x}_{k,C=2} \tag{49}$$

That means the percepts for both causal scenarios are merged based on their probability.

Multisensory Recalibration

Based on this formulation of the CI-model the model taxonomy in Chapter II dissociates recalibration models based on three factors, the learning mechanism, the errorterm used for learning and whether the errorterm is weighted by the posterior probability of a common cause ($p(C = 1|y_k)$). The intend of this study is to clarify the role of the posterior probability of a common cause. However, as all model factors are to some extent interdependent, we quickly review the formulations of the most likely models based on the results of Chapter II and Hong et al. (2021).

From the studies of Chapter II and Hong et al. (2021), essentially two candidate errorterms remained for the aCVAE, the difference between the merged multisensory percepts for audition and vision:

$$\hat{e}_{MPk} := \hat{x}_{A,k} - \hat{x}_{V,k} \tag{50}$$

And the difference between fully fused estimate and segregated estimate:

$$\boldsymbol{\hat{e}_{VEk}} := \boldsymbol{\hat{x}_{k,C=1}} - \boldsymbol{\hat{x}_{k,C=2}} \tag{51}$$

Moreover, the study of Chapter II revealed that recalibration is best modeled by a recalibration mechanism, were recalibration rates take into account the uncertainty in the errorterms, i.e., low uncertainty implies faster recalibration. Recalibration can thereby again be formulated as Kalman Filtering with the following update rule:

$$\hat{\beta}_k = \hat{\beta}_{k|k-1} - \boldsymbol{K}_{\beta,k} * \hat{e}_{\beta k} \tag{52}$$

Again, we refer to Chapter II for the full derivation of the Kalman recalibration model. The aforementioned models assume a two-step procedure, first multisensory estimates are calculated based on Causal Inference, and then in a second step the bias term is updated, i.e., the likelihoods are corrected for the estimated bias in the sensory modality. Two further candidate models were considered for the IVAE. First, the best fitting model in Chapter II assumed that IVAE and VE are parts of a common process, that means the perceptual system estimates spatial positions and its own biases in a single process (Parallel Recalibration and Integration). This can be modelled by extending the state vector x by β to $x_{PRI} = \begin{pmatrix} x \\ \beta \end{pmatrix}$ and similarly defining the block matrices $\boldsymbol{H}_{PRI} = \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_\beta \end{pmatrix}$, $\boldsymbol{R}_{PRI} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_\beta \end{pmatrix}$, $\boldsymbol{H}_{PRI} = \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_\beta \end{pmatrix}$ and $\boldsymbol{Q}_{PRI} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_\beta \end{pmatrix}$, estimates for $\hat{x}_{PRI}$ are then derived analogous to the filter for multisensory integration. The last remaining candidate model for the IVAE assumed a fixed gain $K_{\beta,k}$ that does not depend on sensory reliabilities, importantly we cannot reliably dissociate in this study whether a fixed or optimal gain is used by participants, as we did not vary the reliabilities, which is why we only included the optimal model.

Causal Inference and Multisensory Recalibration

We now want to formalize the core research question of this study. An optimal observer should only recalibrate when a common cause is likely, to mimic this behavior we suggested the following heuristic adaptation of (52) in Chapter II:

$$\hat{\beta}_k = \hat{\beta}_{k|k-1} - \boldsymbol{K}_{\beta,k} * p(C = 1|y_k) * \hat{e}_{\beta k} \tag{53}$$

Hereby the errorterm is weighted by the posterior probability of a common cause, consequently given a constant gain and errorterm, recalibration is faster when a common cause is likely compared to when it is unlikely. In general, $p(C = 1|y_k)$ should be lower for a large absolute disparity compared to a small absolute disparity.

Our second manipulation aimed at increasing $p(C = 1)$ for the CCA pair and decreasing it for the DCA pair. It is important to stress, that the CI-model only provides optimal estimates when $p(C = 1)$ reflects the veridical statistics in the world. These are of course not static and previous studies showed (Odegaard et al., 2017; Tong et al., 2020) that the perceptual system can adapt to new statistics. We propose the following simple update rule for p-common:

$$p_k(C = 1) = p_{k-1}(C = 1) - a * (p(C = 1|y_k) - p_k(C = 1)) \qquad (54)$$

This formula implies that if the posterior probability of common cause is high, the prior probability will increase over time and vice versa for the opposite case.

Response Model

Participant gave response by pointing towards the perceived stimulus location and previous studies suggest, that the pointing process is affected by multiple sources of noise independent from the perceptual process (Tassinari et al., 2006; Trommershäuser et al., 2003). These sources of noise can be summarized as non-perceptual noise $e_{np} \sim N(0, \sigma_{np})$. Participants only responded to one of the sensory components $m \in M := \{A, V\}$. The resulting model for a response r_k in trial k is as follows:

$$r_k = \hat{x}_{k,M=m} + e_{np} \qquad (55)$$

IVAE on Response Level

The results of Chapter III favor IVAE models on the perceptual level, as for instance modelled by (46). However, Park et al. (2021) discuss the role of memory related drivers of the IVAE which should affect localization on the response level rather than perception itself. Additionally, still an estimated 35% of the participants in Study 1 were better described by a response level model of the IVAE. This hypothesis can be formalized by an alternative response model:

$$\boldsymbol{r_k = \hat{x}_{k,M=m} - \hat{\beta}_{k,iVAE} + e_{np}} \qquad (56)$$

In this case (46) becomes:

$$y = \boldsymbol{H}x + e \quad with\, e \sim MVN(0, \boldsymbol{R}) \qquad (57)$$

This implies that the bias term does not affect the perceptual process but only the response process. Importantly, (56) is based on the already updated bias term $\hat{\beta}_{k,iVAE}$ and not $\hat{\beta}_{k-1,iVAE}$.

Model Parameter

An overview of all competing models and the assignment of free parameter is given in Table B. 2. Free parameter for the integration model were the same as in Study 1 (Chapter III, section *Model Parameter Integration*) extended by $\sigma_{q\Delta t}^{2}$ which however was analytically estimated based on the temporal precision block (see *Temporal Precision Block*). In addition to the parameter used for the corresponding integration model, we introduced six parameters. First, the elements of $\widehat{\boldsymbol{P}}_{\beta}$ (the spatial prior for the auditory and visual biases) were free parameter. In contrast to Study 1, $\widehat{\boldsymbol{P}}_{\beta}$ included prior terms for the CVAE and IVAE, but we also defined the elements for $\widehat{\boldsymbol{P}}_{\beta}$ separately for the audio-visual pairs based on, whether the audio-visual discrepancy was to the left or right. We use $\sigma_{\beta VMD}^{2}$with V in {C, I} indicating whether cumulative or instantaneous priors, M in {A, V} indicating the sensory modality and D in {L, R} as an index for direction (left or right). We set the visual bias prior variances to 0.001 since no visual aftereffects were observed in Study 1 and we did not include a visual posttest. This resulted in four free parameters $\sigma_{ßCAL}, \sigma_{ßCAR}, \sigma_{ßIAL}, \sigma_{ßIAR}$.

To account for potential generalization across frequencies for the IVAE (Bruns & Röder, 2015) and CVAE (Study 1) left τ_{A}^{IVAE} τ_{A}^{CVAE} as free parameter. In addition to the CVAE models, the IVAE models further included the decay factor d_{IVAE} as free parameter.

Models with adaptative $p(C = 1)$ included two further parameters, the learning rate $\lambda_{p(C=1)}$ determined how fast $p(C = 1)$ changed in the association block and a decay factor $d_{p(C=1)}$ that determined how fast after association the internal belief about a common cause returns to its initial state.

Parameter Fitting and Model Comparison

We relied on approximate Bayesian computation (ABC, Beaumont, 2019) for parameter inference and model comparison (see Chapter III) as the likelihood function is not available in closed form. All preprocessing steps of the behavioral data analysis were also performed in advance to model-based analysis, albeit outlier trials were not removed from the data set but only labelled. These trials were still used to model the data sets to preserve sequential effects, but responses in these trials were excluded from calculations of distance measures. Similarly, we did not model responses in the association blocks, as these tasks were only designed to maintain participants' attention and elsewise were unrelated to the underlying perceptual processes. However, the perceptual processing and effects of these trials was modelled.

Table 2
Prior parameter for calculation of pretrained model and participant specific priors.

parameter	*Family*	*Lower Bound*	*Upper Bound*
d_{IVAE}	Uniform	0	1
$d_{p(C=1)}$	Uniform	0	1
$Lambda_{p(C=1)}$	Log-Uniform	0.03	0.75
$p(C = 1)$	Uniform	0	1
$sigma_{ßCAL}$	Log-Uniform	0.001	max $(\hat{\sigma}^2_{mpq})$ *8
$sigma_{ßCAR}$	Log-Uniform	0.001	max $(\hat{\sigma}^2_{mpq})$ *8
$sigma_{ßIAL}$	Log-Uniform	0.001	max $(\hat{\sigma}^2_{mpq})$ *8
$sigma_{ßIAR}$	Log-Uniform	0.001	max $(\hat{\sigma}^2_{mpq})$ *8
t_A^{CVAE}	Uniform	0	1
t_A^{IVAE}	Uniform	0	1

Note: We use $\hat{\sigma}^2_{mpq}$ for the estimated sensory uncertainties with p in {1,2} as an index for pair and q in {CCA, DCA} as an index for association type and m in {A, V} as an index for the sensory modalities.

For the distance measures we split the data into subsets based on the experimental conditions disparity, association type, the type of stimuli (bimodal or unimodal), the task-relevant modality (vision or audition) and previous disparity. For each of these subsets autoregressive integrated moving average models were fitted to the trial-wise *Res* (Hyndman & Khandakar, 2008) to obtain smoothed predictions. By adding these predictions to the trial-wise predictions of the linear baseline models (see *Model-free Analysis*) we obtained smoothed predictions for the raw localization responses. We compared distributions of the empirical and simulated deviations from these smoothed predictions using the Wasserstein-Distance (see Bernton, Jacob, Gerber, & Robert, n.d. for a review on ABC based on Wasserstein Distance) for the approximate Bayesian computations. Importantly we further split the subsets in chunks of 32 trials in induction blocks and 30 trials in pre and post blocks to reduce computational costs.

Parameter inference and model comparison was first done for each participant separately following a three-step procedure. Based on weakly informative priors (see Table 2 for an overview) we calculated participant and model specific priors based on a training sample consisting of 1/4 of the trials of each subset (prior data). We refer to Berger & Pericchi (2004) for a discussion of partial bayes factors. These pretrained priors were derived via ABC sequential Monte Carlo sampling (ABC-SMC, Beaumont et al., 2009) with quantile based threshold (Beaumont et al., 2002). The initial quantile was set to 0.1 and all following to 0.5. We set an acceptance ratio of 0.005 or maximum number of 50 iterations (Batch size = 256000) as stopping criterion to ensure comparable runtimes across models and participants. Each run consisted of 4000 samples. Importantly, all trials were used to simulate datasets but only responses in trials from the prior data were used to calculate distances.

In a second step we run an ABC-SMC sampler on the joint model space (Beaumont, 2019; Marin et al., 2012; Toni & Stumpf, 2010), i.e. we treated the type of model as a discrete parameter (see Chapter III for the exact algorithm). The resulting participant-wise model evidences were submitted to a Bayesian model selection analysis for group studies (Rigoux et al., 2014; Stephan et al., 2009). Model comparisons were then based on protected exceedance probabilities and estimated model frequencies.

For models with an estimated frequency different from zero, we estimated posterior distributions for the parameter based on all trials. Simulated responses for the models were based on random samples from these posterior distributions.

Posterior Simulations

Analogously to Chapter III, we inferred parameters for each participant and each model with an estimated frequency > 0 based on the whole data, but otherwise analogous to the prior data run. Models explicitly specified by the research question (see subsequent paragraph) were additionally included in posterior simulations. The resulting posterior samples were used for posterior simulations and to calculate maximum a posteriori estimates of the model parameter. We took 96000 samples of the approximate posterior per participant and run simulations for all samples. Based on the trial-wise means of the simulated responses posterior simulation plots were created.

Model Comparison Formulation of Research Questions

Our main research questions can now be narrowed down to model comparisons. Firstly, we wanted to investigate whether the aCVAE depends on the causal structure of the stimulus context, i.e., whether models based on (53) or (52) do better account for the aCVAE. Secondly, we wanted to test whether participants learn to update their prior beliefs of a common cause based on sensory evidence by testing models that apply (54) against models that do not incorporate (54). And finally, we wanted to test whether the PRI model will also be the most likely IVAE model in this experimental paradigm compared to models based on (56), i.e., an IVAE on the response level.

Results

Model-free Analysis

In a first step the temporal precision block was analyzed to derive participant-wise estimates for the precision and accuracy of audio-visual simultaneity perception. We fitted psychometric functions (cumulative gaussians) via iteratively reweighted least squares (*Mean AIC*= 127.93, *Min* = 82.23, *Max* = 185.69) to derive estimates of the mean and standard deviation of the temporal likelihood functions. On average, participants perceived audition and vision as synchronous when vision was leading (*Mean PSE* = -28.13 ms, *SEM* = 5.07 ms), replicating previous findings (Fujisaki et al., 2004; Vroomen & Keetels, 2010). Average and participant-wise estimates for $\sigma_{\Delta t}$ and $\mu_{\Delta t}$ are shown in Figure 25. Note that the *PSE* is an estimate for $\mu_{\Delta t}$ (Rohde et al., 2016).

Moreover, we checked, that the average performance (*EMM* = 0.896) in the control task during the association block was well above chance (0.5), *estimate* = 0.396, z = 43.99, p < 0.001 (the estimate is derived from a test against 0.5). These results clearly indicate that participants attended to the visual and auditory stimuli throughout the association phase. A generalized LMM (logit link) with association type and stimulus modality (auditory or visual) revealed a significant main effect of stimulus modality, $\chi^2(1)$ = 29.39, p < 0.001 and a significant interaction of stimulus modality and association type, $\chi^2(1)$ = 34.10, p < 0.001. Overall, it seems that the auditory targets were easier to distinguish, which rendered the visual task for the DCA pair especially difficult as this was the only combination where also stimuli without auditory component were task relevant.

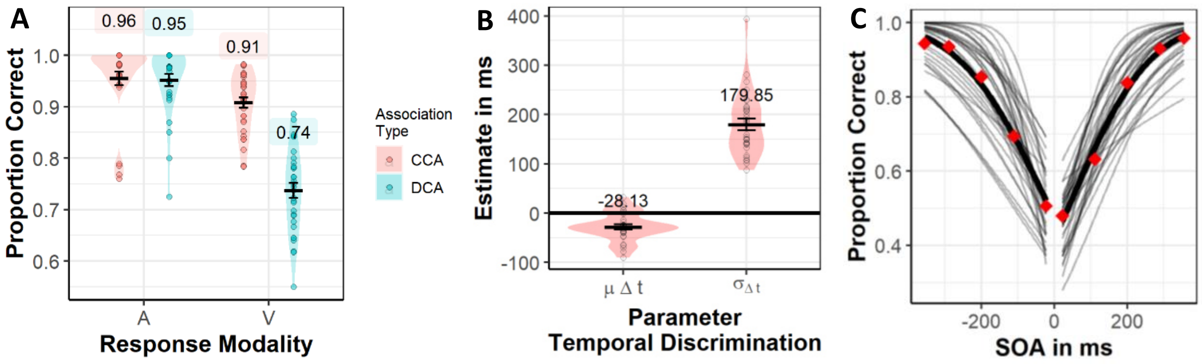


Figure 25. Results of the association and temporal precision block. **A**: Relative frequency of correct responses for control task in association blocks. Colors indicate association type. Results are depicted separately depending on whether the control question referred to visual (V) or auditory stimuli (A). Black bars indicate averages, errorbars denote SEM. Colored dots indicate individual datapoints. **B**: Parameter estimates based on fitted psychometric curves to responses in temporal precision blocks. **C**: Individually fitted psychometric curves as a function of stimulus onset asynchrony (SOA) based on data of the temporal precision block are shown with thin lines and low alpha. The fat black curve indicates a psychometric curve based on all datapoints across participants. Red dots indicate average relative frequencies of correct responses across participants.

Furthermore, we tested whether auditory localization already differed a baseline analyzing again trial-wise residuals $Res_{\,UNI}$ with the fixed factors association type and absolute disparity (note that $Res_{\,UNI}$ are already standardized by the direction of the audio-visual disparity). On average, baseline biases were in the direction of the later used audio-visual disparity (*EMM*) for the DCA pair (Figure 26, both panels, left side) and in the opposite direction for the CCA pair (Figure 26, both panels, right side). Further statistical analysis had to account for these initial biases because the baseline measurement preceded the association blocks, in which the average audio-visual disparity was zero for CCA and DCA pairs. Hence, it might have been that recalibration occurred during the association blocks towards zero.

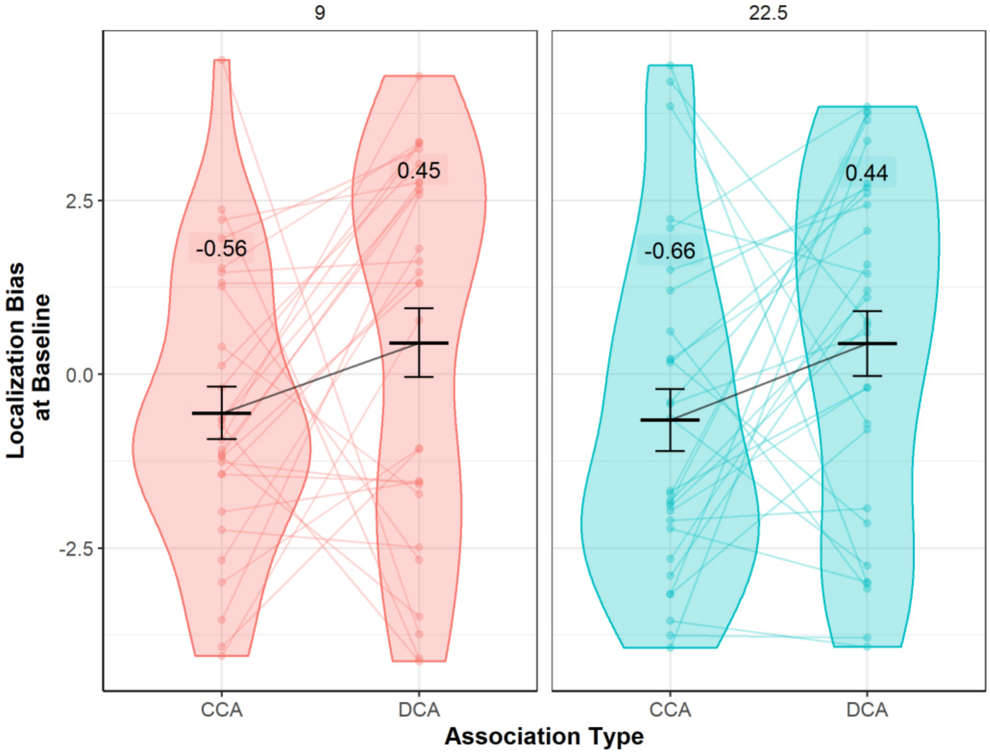


Figure 26. Average localization Biases at Baseline. The panels show mean intercepts of linear models of localization responses regressed on true stimulus position for auditory unimodal trials at baseline grouped by association type (CCA: common cause association, DCA: distinct cause association). Intercepts were standardized by the sign of the used disparity, i.e., positive values indicate a bias at baseline in the direction of the later used audio-visual discrepancy. Individual data is shown in color and low alpha, whereas averages are shown in black. Errorbars denote standard errors of the mean. The left panel shows data for the small (9°) absolute disparity condition while the right panel shows data for the large (22.5°) absolute disparity condition.

The most conservative way to proceed was therefore to assume the worst-case scenario, i.e., average biases of zero before recalibration. For all subsequent behavioral analyses, the predictions to calculate the *Res* were therefore based on zero intercepts. As we explicitly modeled recalibration throughout the association phase, the model-based analysis is not affected by the differences at baseline.

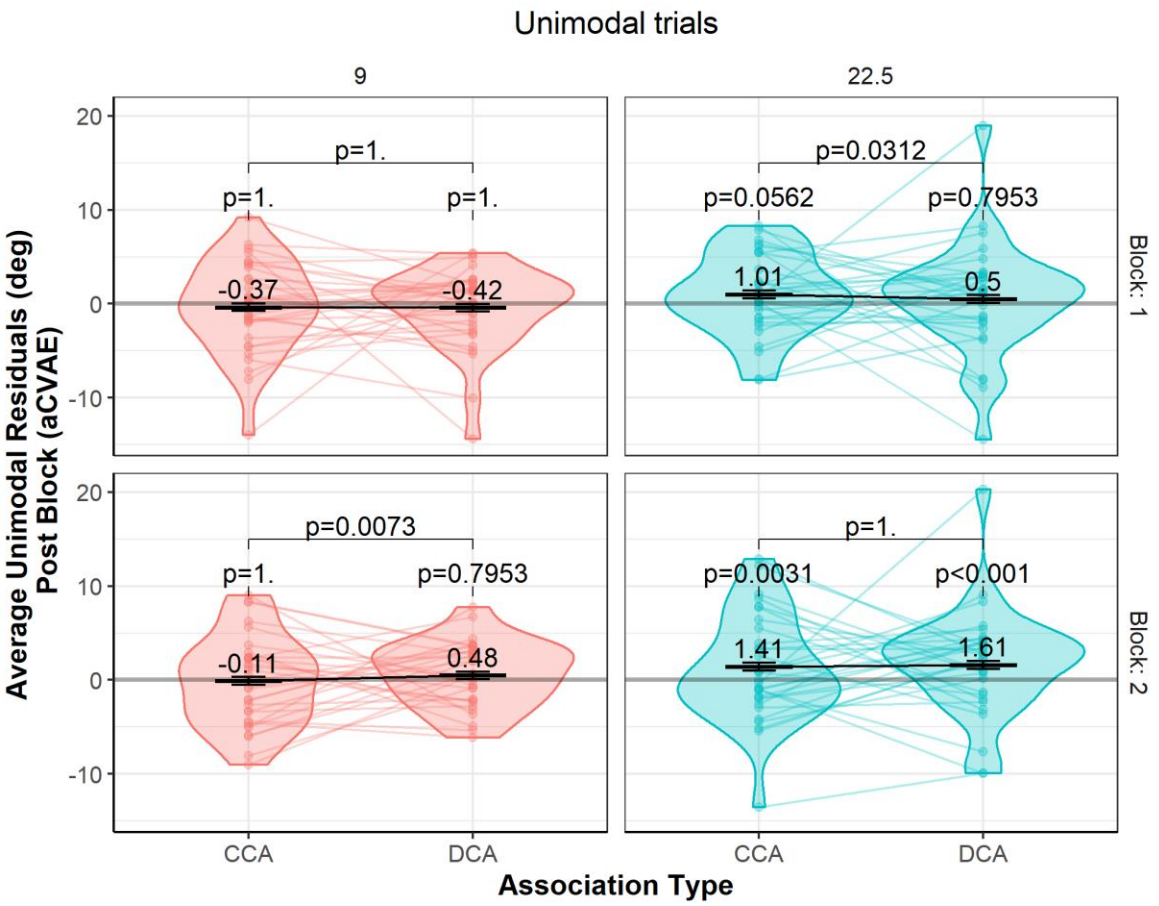


Figure 27. Effects of association type on the unimodal localization in post blocks. The aCVAE is approximated by EMMs of the unimodal residuals and grouped by association type (CCA: common cause association, DCA: distinct cause association). Residuals are standardized in the direction of the audio-visual discrepancy. EMMs are derived from a repeated measures LMM and shown with black lines and dashes. Individual data points are calculated via individual linear models and shown in low alpha. Lines connect data points of unique participants across conditions. Errorbars show the SEs estimated from the LMM. P-Values are calculated via contrasts and holm adjusted. **Left** column: EMMs of unimodal trials after adaptation with a small audio-visual discrepancy (9° left or right). **Right** column: EMMs of unimodal trials after adaptation with a large audio-visual discrepancy (22.5° left or right). **Top** row: EMMs in the first post block of a session. **Bottom** row: EMMs in the second post block of a session.

Post Block and aCVAE

We evaluated the occurrence of the aCVAE based on the *EMMs* of the trial-estimates of the localization biases Res_{UNI} in the post blocks. Substantial aCVAEs were only found for the large absolute disparity, *EMM* = 1.41, *z* = 3.45, *p* = 0.003 for the CCA pair and *EMM* = 1.61, *z* = 3.92, *p* < 0.001 for the DCA pair (Figure 27). A main effect of absolute disparity, $\chi^2(1)$ = 374.20, *p* < 0.001 further implies larger aftereffects when the absolute disparity was large compared to small. Hence in our study the aftereffect increases as function of the audio-visual disparity in line with previous results (Frissen et al., 2012). Moreover, the aCVAE increased from the first block to the second block indicated by a main effect of block number, $\chi^2(1) = 29.84$, $p < 0.001$, indicating that longer adaptation leads to larger aCVAEs (Frissen et al., 2012).

Intermixed Blocks and IVAE

The *EMMs* of the Res_{UNI} in the intermixed blocks are shown in Figure 28. Overall, the *EMMs* are larger in the intermixed blocks compared to the post block, statistically confirmed by a contrast, *estimate* = 0.92 *z* = 11.63, *p*<0.001.

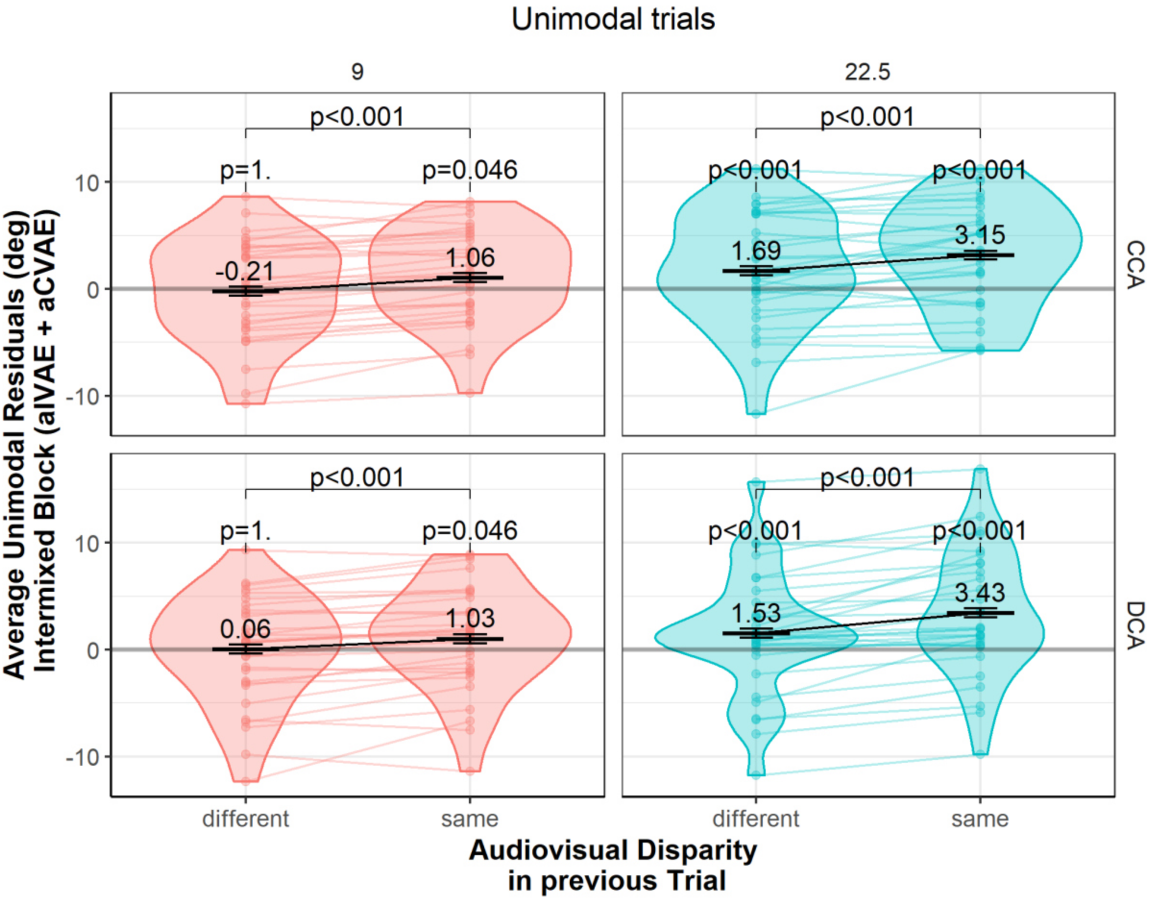


Figure 28. Effects of the previous audio-visual disparity on the unimodal localization in intermixed blocks. The panels show *EMMs* of the unimodal residuals, grouped by previous (prev.) audio-visual disparity. Residuals are standardized in the direction of the audio-visual discrepancy. *EMMs* are derived from a repeated measures LMM and shown with black lines and dashes. Individual data points are calculated via individual linear models and shown in low alpha. Lines connect data points of unique participants across conditions. Errorbars show the SEs estimated from the LMM. P-Values are calculated via contrasts and holm adjusted. Left column: *EMMs* of unimodal trials after adaptation with a small audio-visual discrepancy (9° left or right). Right column: *EMMs* of unimodal trials after adaptation with a large audio-visual discrepancy (22.5° left or right). Top row: *EMMs* of the distinct cause association (DCA) condition. Bottom row: *EMMs* of the common cause association (CCA) condition

The overall larger biases in intermixed blocks compared to post blocks are clear indicators that additionally to the aCVAE in the post block an IVAE occurred during intermixed blocks. Moreover, we performed a contrast to probe the effect of previous disparity. For a unimodal auditory stimulus, the bias was significantly larger when the previous disparity was in the same direction (compared to the opposite) as in the corresponding bimodal stimulus (*estimate* = 1.31°), z = 11.06, p < 0.001. This demonstrates that the IVAE at least partially transferred across frequencies. Moreover, this typical trial-wise modulation is also present for the small audio-visual discrepancy (Figure 28).

Comparable to the aCVAE, the size of the IVAE increased with increasing absolute disparity, *estimate* = 1.95 °, z = 16.33, p < 0.001. Yet, in contrast to the aCVAE, the magnitude of the IVAE did not change from block 1 to 2, neither when the previous disparity was the same, *estimate* = 0.11°, z = 0.66, p = 0.51, nor when it was different, *estimate* = 0.21°, z = 1.25, p = 0.21.

Likewise opposing to the aCVAE, the difference between CCA and DCA did not change from block 1 to 2 in intermixed blocks, neither when the previous disparity was the same, *estimate* = 0.468°, z = 1.4, p = 0.16, nor when it was different, *estimate* = 0.525°, z = 1.567, p = 0.12.

Ventriloquism Effect

Descriptively it becomes apparent that the *EMM*s of the $Res_{\ BI}$ (Figure 29) are noticeably larger than the *EMM*s for the $Res_{\ UNI}$ (Figure 28), indicating the occurrence of the VE during the intermixed blocks. This observation is statistically backed by a main effect of stimulus type, $\chi^2(1)$ = 3841.61, p < 0.001. Overall, trial-wise biases were 6.47° larger in bimodal trials compared to unimodal trials z = 62.08, p < 0.001. Additionally, the previous disparity also affected $Res_{\ BI}$, analogous to $Res_{\ UNI}$, but only when the absolute disparity was large *estimate* = 0.78°, z = 3.94, p < 0.001. We found a significant interaction of stimulus type and previous disparity $\chi^2(1)$ =14.78, p < 0.001, which likely reflects a reduced effect of previous disparity in bimodal trials compared to unimodal trials, *estimate* = -0.80°, z = -3.84, p<0.001. Furthermore, the size of the VE increased with increasing absolute disparity, *estimate* = 4.77°, z = 28.10, p < 0.001.

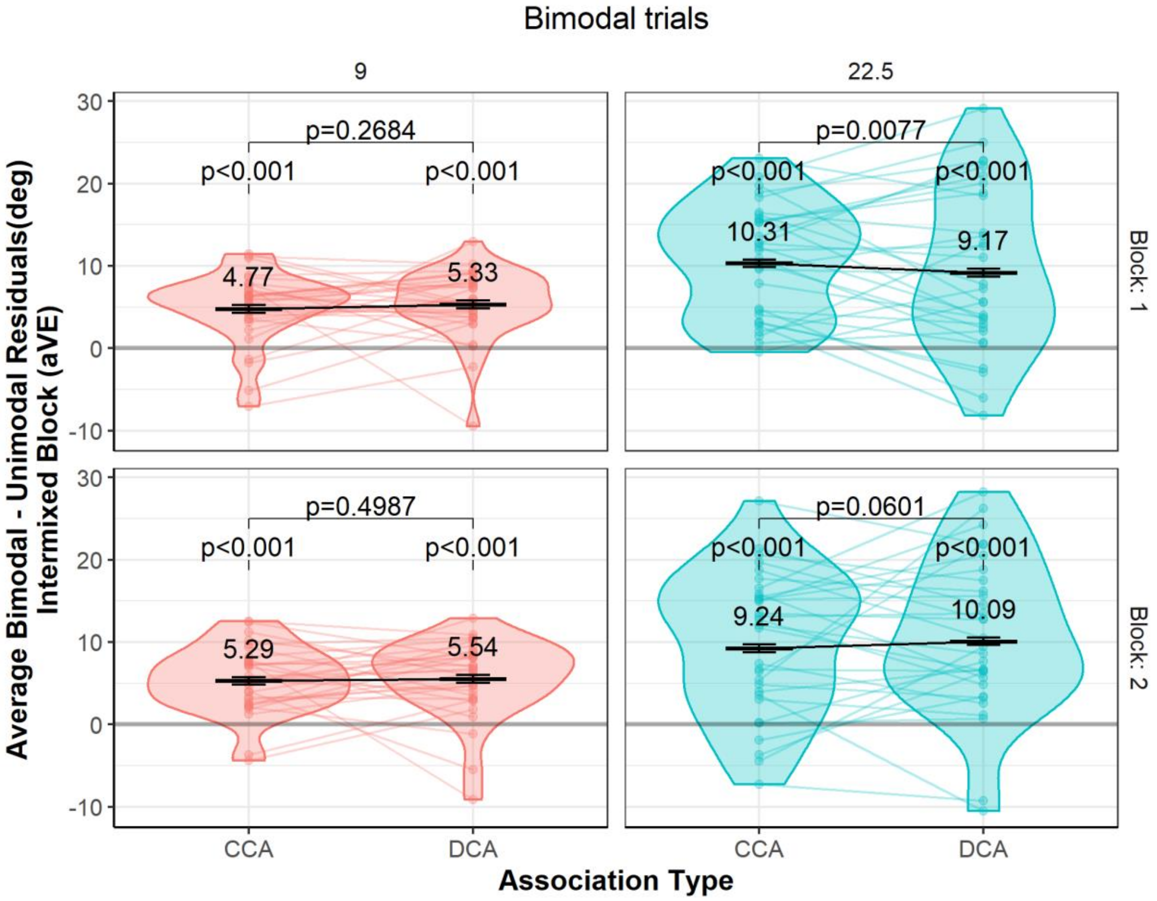


Figure 29. Effects of association type on the bimodal localization (auditory component) in intermixed blocks. The panels show *EMM*s of the bimodal residuals, grouped by association type (CCA: common cause association, DCA: distinct cause association). Residuals are standardized in the direction of the audio-visual discrepancy. *EMM*s are derived from a repeated measures LMM and shown with black lines and dashes. Individual data points are calculated via individual linear models and shown in low alpha. Lines connect data points of unique participants across conditions. Errorbars show the SEs estimated from the LMM. P-Values are calculated via contrasts and holm adjusted. Left column: *EMM*s of unimodal trials after adaptation with a small audio-visual discrepancy (9° left or right). Right column: *EMM*s of unimodal trials after adaptation with a large audio-visual discrepancy (22.5° left or right). Top row: *EMM*s in the first post block of a session. Bottom row: *EMM*s in the second post block of a session.

The Effects of Association Type

No main effect of association was found $\chi^2(1)$ =1.23, p = 0.27. Hence, averaged over all other factors association type did not have a significant effect on the aCVAE. Yet, we found a significant interaction of block number and association type, $\chi^2(1)$ = 21.53, p < 0.001. We found a similar pattern across our measures for aIVAE, aCVAE and aVE that might underly this interaction. The *EMM*s shown in (Figure 27, left column) suggest, that in the first block a larger aCVAE was obtained for the CCA pair compared to the DCA pair but in the second block the order reversed. We will refer to this as association time interaction from now on. This very same pattern holds for the aVE (Figure 29) and unimodal residuals in intermixed blocks (aIVAE + aCVAE) as well. In the design of our study, the type of stimulus (uni vs. bi) is nested in block type, i.e., we did not acquire bimodal trials at baseline. Hence the LMM does not capture an interaction term of block type and stimulus type, which would allow to directly test whether the observed interaction does vary in size across aVE, aIVAE and aCVAE. However, we can compare aIVAE and aCVAE as well as aIVAE and aVE pairwise. A lack of a significant interaction of block number, association type and block type (the factor was removed during stepwise model reduction), implies that the interaction of association type and block number did not significantly differ for post block and intermixed block. Since post blocks are a reliable measure of the aCVAE and no difference between post block and intermixed block was found, the association time interaction in intermixed blocks might be solely induced by the aCVAE. However, a significant interaction of association type, block number, absolute disparity and stimulus type was found, $\chi^2(1) = 7.66, p < 0.0056$. A post-hoc contrast revealed a significantly larger association time interaction for the bimodal residuals compared to the unimodal residuals, *estimate* = 1.27, $z = 2.25, p = 0.025$, when the absolute disparity was large. This indicates that the time association interaction affected the aVE beyond the effects on unimodal localization.

Model-based Analysis

Model Comparison

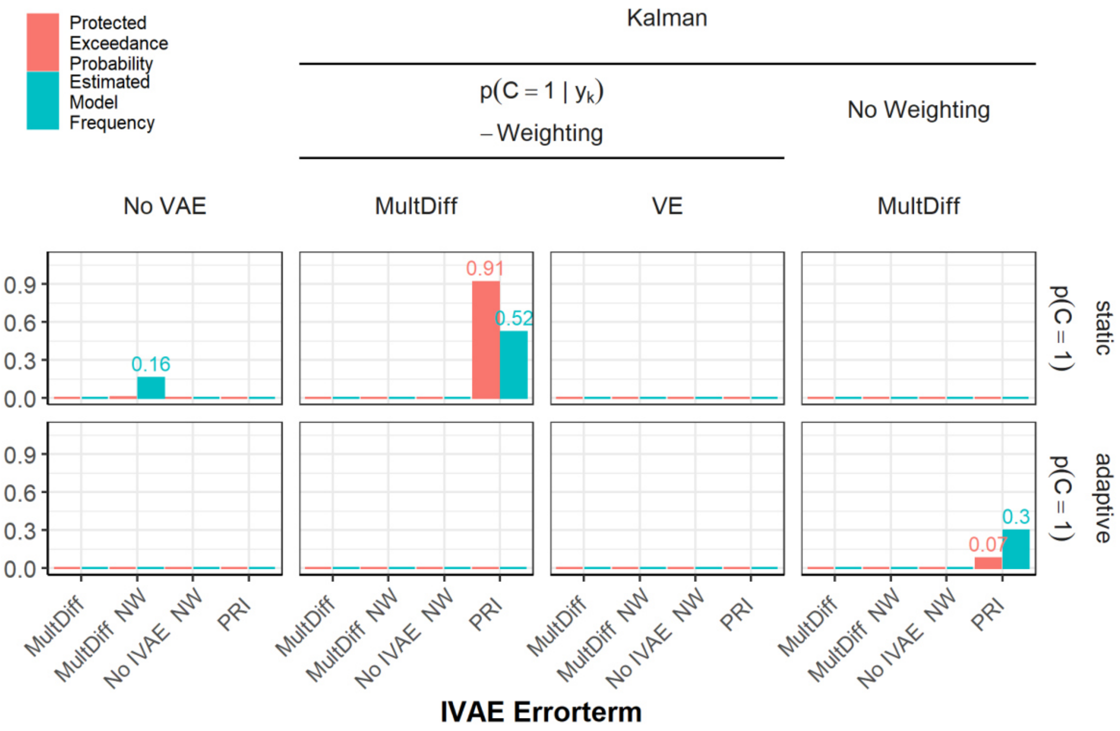


Figure 30. Estimated model frequencies (turquois) and protected exceedance probabilities (red) from random effects Bayesian group model comparison. CVAE-models are grouped by the learning mechanism, type of Weighting and errorterm in separate panels. IVAE-models are grouped by the errorterm and type of Weighting along the x-axis, hereby $p(C = 1|y_k)$ - Weighting as well as Kalman Filtering are assumed unless otherwise indicated (No Weighting abbreviated to NW). All MultDiff-IVAE-models assume the IVAE occurs at response level, whereas in the PRI model the IVAE occurs at perceptual level. Models in the **top** row assume a static $p(C = 1)$, whereas models in the **bottom** row assume an adaptive $p(C = 1)$. Model factors and abbreviations are described in detail in Chapter II.

Estimated model frequency (*eF*) as well as protected exceedance probability (*PEP*) were largest for a CVAE model based on Kalman Filtering, $p(C = 1|y_k)$ -Weighting and the MultDiff errorterm combined with the PRI IVAE model (*eF* = 0.52, *PEP* = 0.91, Figure 30). Moreover, in line with the only marginal behavioral effect of association type, the best fitting model contained a static value of $p(C = 1|y_k)$. The second most frequent model (*eF*=0.30, *PEP* = 0.04) was based on Kalman Filtering, no weighting, the MultDiff errorterm and an adaptive $p(C = 0.07)$. With an *eF* of 0.16 participants did not show a aCVAE and the IVAE was based on the MultDiff errorterm and assumed to occur on the response level, again $p(C = 1)$ was static in this model. For all other models *eF* was zero. Overall, models based on the

MultDiff errorterm for the aCVAE as well as models based on the PRI model of the IVAE yielded an estimated frequency of 0.82.

Posterior Simulations

Recalibration

To descriptively assess whether the inference scheme provided in the Section *Parameter Fitting and Model Comparison* would allow to recover the actual time course of recalibration on a trial-by-trial basis, we evaluated trial-by-trial biases based on posterior simulations and empirical data (Figure 31).

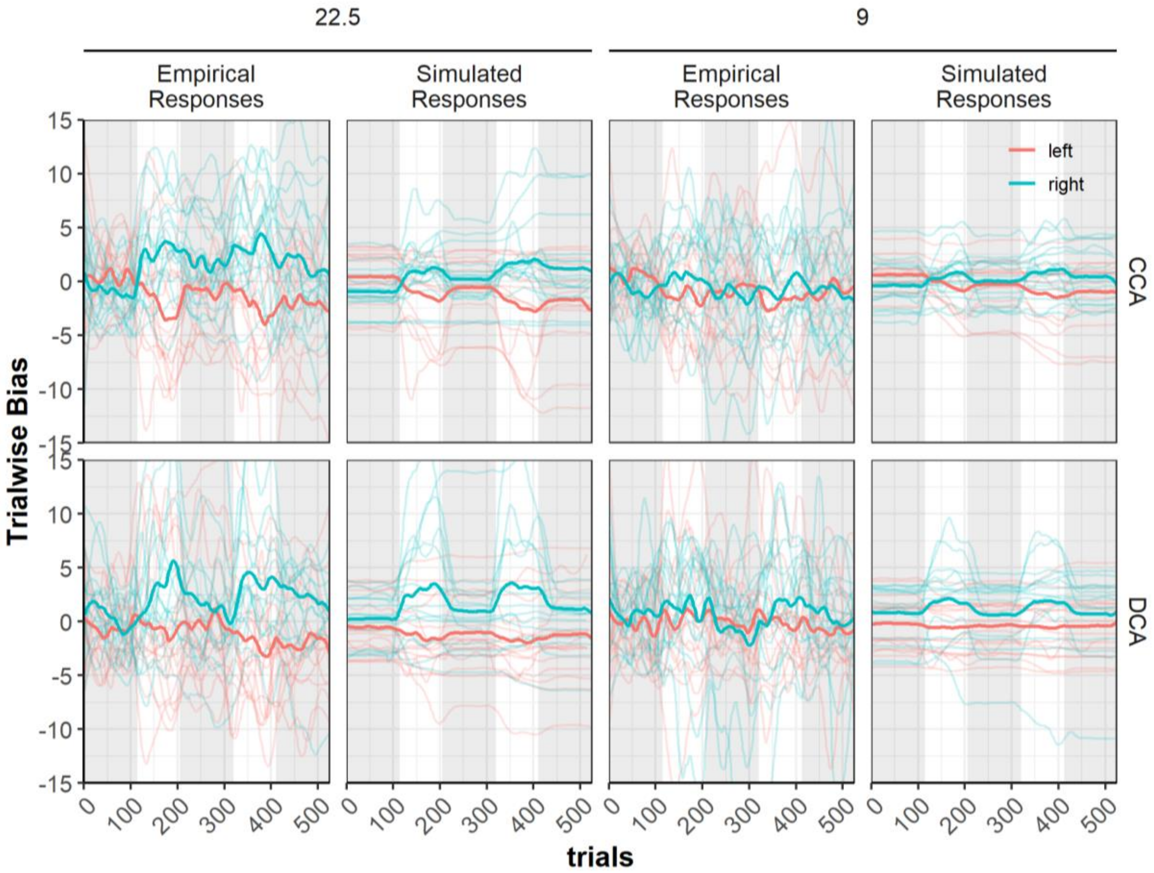


Figure 31. Empirical and simulated trial-wise biases in unimodal trials. Data was simulated for the best fitting CVAE model based on Kalman Filtering, $p(C = 1|y_k)$ -Weighting and the MultDiff errorterm combined with the PRI IVAE model. Trial-wise biases reflect participant-wise LOESS-smoothed (span=0.05) *Res* (see *Methods* for details) of pre, post (both with light grey background) and intermixed blocks (white background) grouped by association type (CCA or DCA). Negative values indicate biases to the right and positive values vice versa. Fat lines indicate LOESS-smoothed (span=0.05) averages over participants whereas thin, low alpha lines indicate individual data. Leftward panels show results for the large audio-visual discrepancy (22.5°) and rightward panels show results for the small audio-visual discrepancy (9°). Within each panel the line color indicates whether audio-visual adaptation was to the left (red) or to the right (turquoise).

Across experimental conditions, the average predictions of trial-wise biases provided a good description of the empirical average trial-wise biases with R^2= 0.78. What was most apparent in the empirical data are prominent bulges in the direction of the audio-visual discrepancy during intermixed blocks, especially in the high absolute disparity condition (Figure 31, most leftward panels).

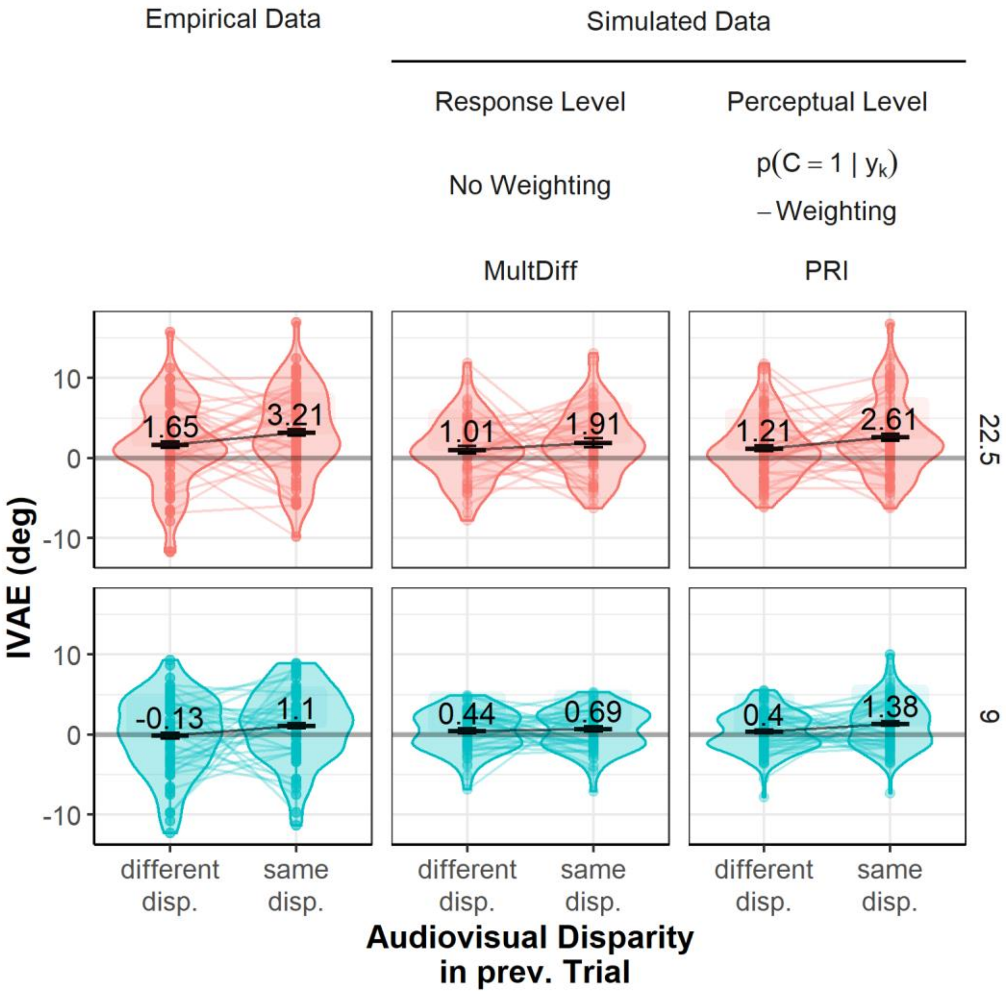


Figure 32. Effect of previous audio-visual disparity on empirical auditory IVAEs and posterior simulations. The two best models are shown that do or do not rely on $p(C=1|y)$ – *Weighting*. The panels show average and participant-wise unimodal *Res* of intermixed blocks separated (row wise) for absolute audio-visual discrepancies. Whitin each panel, results are shown separately for trials preceded by a bimodal trial with either the same or a different disparity. Residuals are standardized in the direction of the audio-visual discrepancy. Average data is shown in opaque black, whereas participant-wise data is depicted by colored low alpha dots. Lines connect corresponding data points across conditions. Errorbars show the SEM. Model factors and abbreviations are described in detail in Chapter II.

These bulges indicate a rapid increase in the beginning of the intermixed blocks and a rapid decay at the beginning of the consecutive post block (Bosen et al., 2018). Importantly, trial-wise biases do not return to baseline, indicating ongoing shifts in the post block, indicative of the aCVAE. The bulges during intermixed blocks imply unimodal shifts in the direction of the audio-visual disparity beyond the aCVAE. The simulated responses demonstrate that this pattern can be recovered well when the IVAE is modelled as not fully transferring across auditory stimuli.

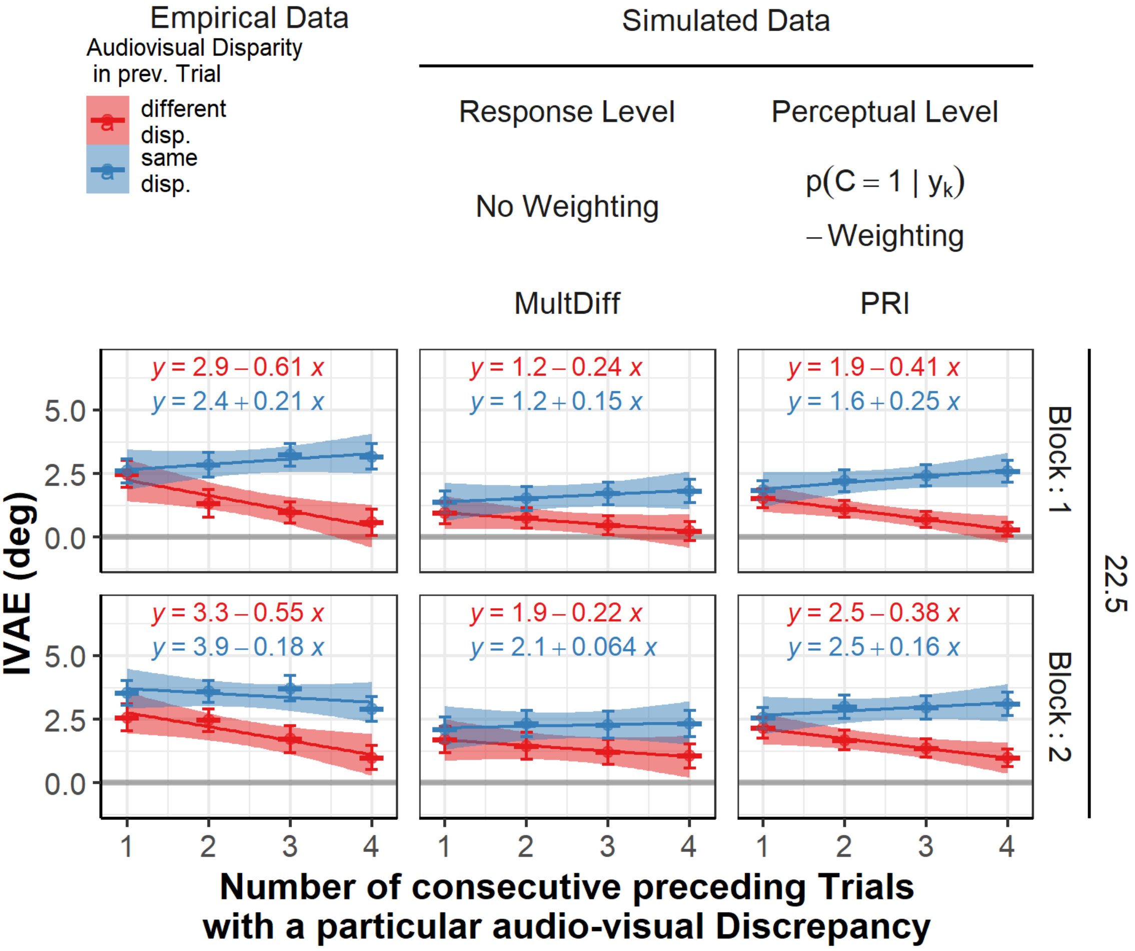


Figure 33. Effect of number of consecutive trials on empirical auditory IVAEs and posterior simulations. The two best models are shown that do or do not rely on $p(C = 1|y)$ – Weighting. The panels show average and participant-wise unimodal *Res* of intermixed blocks (large discrepancy only) separated (row wise) for the block number. Whitin each panel results are shown separately for trials preceded by a bimodal trial with either the same (blue) or a different disparity (red) as in the corresponding bimodal stimulus pair and as a function of number of consecutive trials. *Res* are standardized in the direction of the audio-visual discrepancy used in the corresponding bimodal stimulus pair. Errorbars show the SEM. Model factors and abbreviations are described in detail in Chapter II.

Note that otherwise simultaneous bulges to the left and right would not be possible since leftward and rightward IVAE would cancel each other out or one direction would dominate. Hence, the simulation results imply that the IVAE does not fully transfer across frequencies and therefore builds up over time during bimodal stimulation.

Yet, the IVAE does transfer across frequency to a substantial amount. Figure 32 (upper row) shows that if the audio-visual disparity in the preceding bimodal trial had been different from the audio-visual disparity used in bimodal trials containing the same particular auditory component as the unimodal trial, the unimodal bias in the direction of the audio-visual disparity was highly reduced. Partial transfer for the IVAE model does recover this pattern quite well (Figure 32, bottom row), albeit the overall size of the unimodal bias for the large absolute disparity is slightly underestimated (Figure 32, bottom row, left). With respect to the partial transfer as well as the overall size of the IVAE posterior simulations of the PRI model (Figure 32, middle column) better account for the pattern in the empirical data (Figure 32, left column) compared to the best fitting model on response level (Figure 32, right column). Participant wise comparisons of the simulated IVAE and the empirical IVAE show the same pattern, i.e., the PRI model ($R^2 = 0.80, RMSE = 2.52$) better accounts for the IVAE compared to the best fitting model on response level ($R^2 = 0.74, RMSE = 2.73$).

Additionally, we performed an exploratory analysis to compare how the empirical and simulated IVAEs behave depending on how many consecutive trials of the same or different previous disparity preceded a unimodal trial. Especially, for the large absolute disparity, the PRI model predicts that with increasing numbers of consecutive trials the effects of same or different previous disparities cumulate (Figure 33). A similar effect can be observed in the empirical data but for the same previous disparity only in the first intermixed block (Figure 33). In the second block the slope becomes negative even for the previous disparity same condition (Figure 33, left column, lower row), implying no cumulation of the IVAE in that condition. An exploratory statistical analysis (via an LMM given by (45)) revealed a significant main effect for block number χ^2 (1) = 8.26, p = 0.004, previous disparity χ^2 (1) = 68.97, p < 0.001 and consecutive trials χ^2 (1) = 11.37, p < 0.001. Moreover, a significant interaction of previous disparity and consecutive trials was found χ^2 (1) = 12.73, p < 0.001, indicating that the effects of same and different previous disparity cumulated in opposing directions as hypothesized. However, the interaction between previous disparity, consecutive trials and block number was not significant χ^2 (1) = 1.45, p = 0.23.

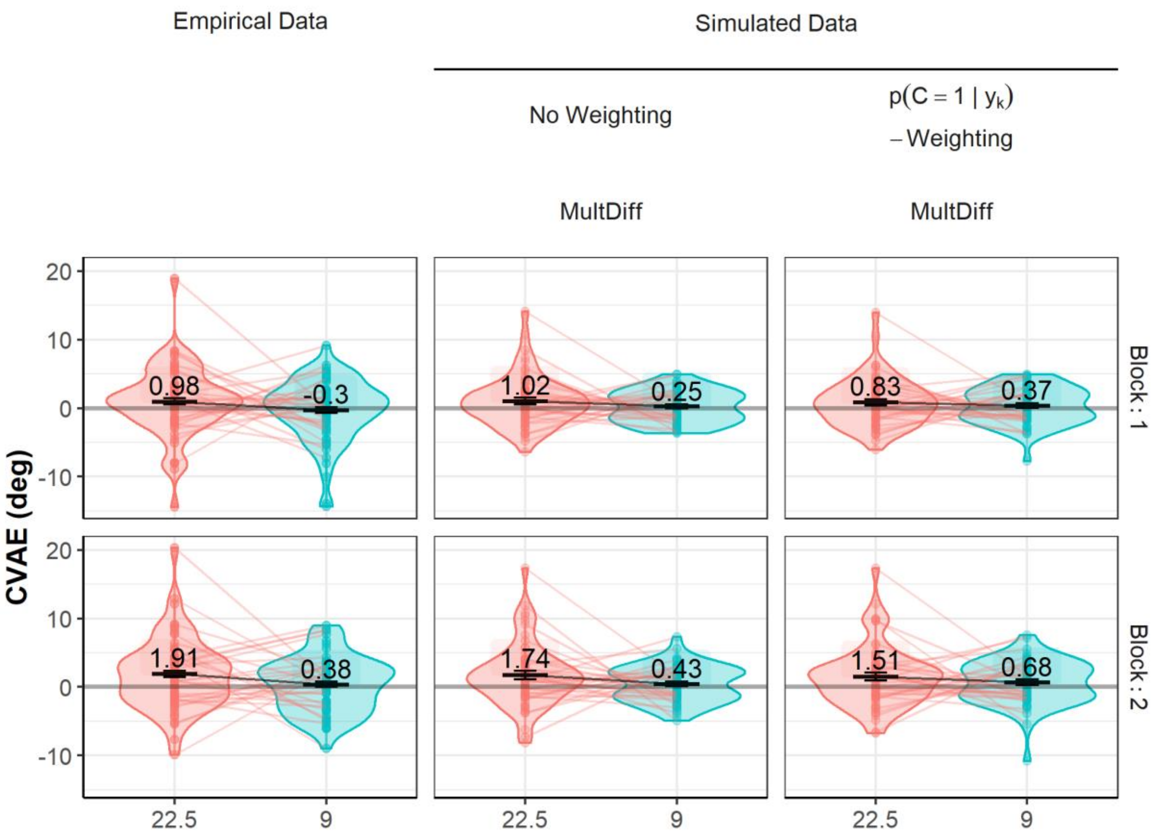


Figure 34. Effect of absolute audio-visual Disparity on empirical aCVAEs and posterior simulations. The two best models are shown that do or do not rely on $p(C = 1|y)$ – Weighting. The panels show average and participant-wise unimodal *Res* of post blocks separated for the block number (row wise). *Res* are standardized in the direction of the audio-visual discrepancy and data for the large discrepancy (22.5°) is shown in red while data of the small discrepancy (9°) is shown in turquoise. Average data is shown in opaque black, whereas participant-wise data is depicted by colored low alpha dots. Lines connect corresponding data points across conditions. Errorbars show the SEM. Model factors and abbreviations are described in detail in Chapter II.

The magnitude of the aCVAE in dependence of the absolute audio-visual disparity is quite well recovered (Figure 34), showing a decrease in the size of the aCVAE with decreasing absolute disparity of similar size in the empirical and simulated data. Moreover, from the first block to the second block the size of the aCVAE increases in both empirical and simulated data, although this increase is slightly underestimated in the simulated data. Although, the best competing model with no Weighting seems to better account for the empirical data (Figure 34, right column) on group level (Figure 34, middle column) compared to the overall best fitting model (Figure 34, right column), participant wise comparison of the modelled aCVAE and the

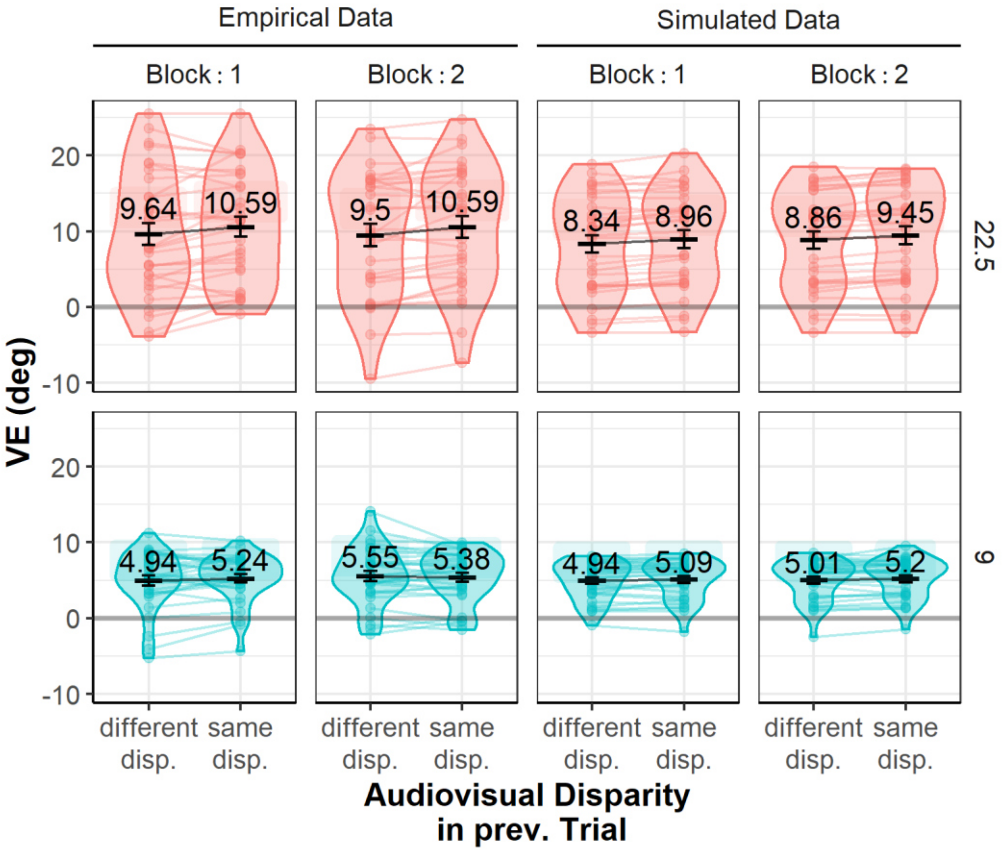


Figure 35. *Effect of previous audio-visual Disparity on empirical auditory VEs and posterior Simulations*. Data was only simulated for the best-fitting model assuming PRI for the IVAE and $p(C = 1|y)$ -Weighting and the MultDiff errorterm for the CVAE. The panels show average and participant-wise bimodal *Res* of intermixed blocks separated for absolute audio-visual discrepancies across rows. Residuals are standardized in the direction of the audio-visual discrepancy. Whitin each panel results are shown separately for trials preceded by a bimodal trial with either the same or a different preceding audio-visual disparity. Average data is shown in opaque black, whereas participant-wise data is depicted by colored low alpha dots. Lines connect corresponding data points across conditions. Errorbars show the SEM.

empirical aCVAE are better for the CVAE model with $p(C = 1|y)$ -Weighting and the MultDiff errorterm ($R^2 = 0.60, RMSE = 3.12$) compared to the CVAE model with no Weighting and the MultDiff errorterm ($R^2 = 0.57, RMSE = 3.24$).

Integration

The general pattern observed for the VE was well recovered by posterior simulations (Figure 35). Most prominently, the VE reduced with decreasing absolute audio-visual disparity, whereby only the size of the VE for large absolute disparities (Figure 35, left) is slightly underestimated. Block-wise comparison of the simulated data show that the best model predicts an increase of VE for the large absolute disparity which is not observed in the empirical data. Similar to unimodal trials an effect of the audio-visual disparity in the preceding trial was observed in the empirical data when the absolute disparity was large (Figure 35, top left), indicating that the IVAE influenced bimodal localization. A similar trend was observed in the simulated data (Figure 35, bottom left), albeit reduced. Participant-wise comparisons of the simulated VE and the empirical VE showed a very good alignment ($R^2 = 0.86,\ RMSE = 2.71$).

Discussion

Main Findings

The present study tried to specify whether immediate and cumulative multisensory recalibration are directly influenced by Causal Inference via $p(C = 1|y)$ -Weighting. To jointly evaluate the effect of Causal Inference on aIVAE, aCVAE and aVE, this study extended the paradigm of Bruns et al. (Bruns & Röder, 2015) by firstly including association blocks following the paradigm of Tong et al. (2020) and secondly varying the absolute size of the audio-visual disparity throughout bimodal stimulation across sessions. Furthermore, baseline- and post-measurements were performed to obtained estimates of aCVAE unaffected by the aIVAE. This allowed to better dissociate the effects of aIVAE, aCVAE and aVE during intermixed blocks as wells as investigating their interplay.

The behavioral results clearly replicated a modulation of aVE (Wallace et al., 2004), aIVAE (Wozny & Shams, 2011a) and aCVAE (Frissen et al., 2012) by absolute disparity. All effects increased with increasing audio-visual disparity. The type of association affected the CVAE and the VE in a varying manner over time but only when the audio-visual disparity was large. Initially aCVAE and VE were increased for the CCA pair, but this effect reversed over time indicating that the effect of association dissipated over time.

The model comparisons partially confirmed the results of Study 1 by demonstrating that the majority of participants were best described by a model that assumes a common process for the VE and the IVAE. Although, the aCVAE was again best described by a model based on the MultDiff errorterm, in contrast to Study 1 most of the participants were best described by a model based $p(C = 1|y)$ -Weighting. A schematic summary of the set of tested models and the model selection is given in Figure 36.

The behavioral Effects of Association

Multisensory Integration

The localization shift during bimodal trials for the CCA pair was about 0.98° larger than for the DCA pair for an absolute audio-visual discrepancy of 22.5° and no difference was observed for the small audio-visual disparity. In contrast to these results Tong et al. (2020) report differences between CCA pair and DCA pair larger than 5° on average for an absolute audio-visual discrepancy of 18°. Since participants completed twice as many association trials compared to Tong et al. (2020) before the first intermixed block started and as the intermixed block contained less bimodal trials (192) than a standard test block (288) of Tong et al. (2020), it is unlikely that a similarly sized effect of association type could have dissipated or not yet have been build up before bimodal localization trials in the first intermixed block. It is more likely that the effect of association type was genuinely smaller in the present study. The association blocks in the present study differed mainly with respect to two aspects from Tong et al. (2020). First, due to the reduced field of view of the VR headset, smaller audio-visual discrepancies were used for the DCA pair (+/- 9° to +/- 22.5° compared to +/- 13.5° up to +/- 40.5° in Tong et al.,2020). Second, smaller SOA were implemented (+/- [300 ms to 500 ms] compared to +/- [750 ms to 1,500 ms]). The latter change was inspired by the suggested mechanism for changing causal priors. The assumption was that the system might store separate priors for the CCA and DCA pair and update them based on the posterior probability of a common cause. In this case, it is not clear how unimodal trials would affect $p(C = 1)$. The SOA used in Tong et al. (2020) essentially led to the perception of distinct visual and auditory events even during bimodal trials for the DCA pair. Smaller SOAs theoretically still lead to small $p(C = 1|y)$ but on average distinct from zero and are therefore better attributable to a multisensory event and thus to $p(C = 1)$ of the particular pair. However, it might be that the perceptual system does not only store and update probabilities for the multisensory but also for the distinct unisensory visual and auditory scenarios, i.e., the probabilities that the auditory or visual component of a pair occurs on its own. In this case, $p(C = 2)$ becomes the product of the two unisensory probabilities for the auditory and visual component of a particular pair. Association trials in which distinct unimodal events are perceived, would then more efficiently update estimated probabilities for the unimodal components and thereby $p(C = 2)$, since there is literally no uncertainty about the causal structure compared to trial in which there is a residual possibility of a common cause. Note that an increase in $p(C = 2)$ always decreases $p(C = 1)$ as both sum to 1. This might generally explain why the large SOAs in Tong et al. (2020) leading to two distinct unimodal events might have affected $p(C = 1)$ and more specifically did so in a more efficient way than the smaller SOAs used in the present study. A model capable of inferring the current stimulus context and learning its probability has been recently proposed for learning sensorimotor repertoires (Heald et al., 2021). Although our explanation for the unexpectedly small effect of association type is speculative in nature, it can be explicitly tested in future experiments, since it predicts that unisensory stimulus presentations alone are sufficient to lower $p(C = 1)$ for audio-visual stimulus pairs containing these unisensory stimuli. Our post-hoc explanation would imply that more general contextual inference as described in Heald et al. (2021) could be a meaningful extension of the Causal Inference framework for multisensory perception.

Multisensory Recalibration

We found an effect of association type on unisensory localization in post and intermixed blocks. But given that this effect did not differ significantly between post and intermixed blocks, the effect in intermixed blocks can be explained by a single alteration of the aCVAE. In other words, the data does not provide evidence for an additional alteration of the IVAE since one would expect a difference between intermixed and post blocks in this case. The association type could have affected the aCVAE via an increased $p(C = 1|y)$ which naturally arises when $p(C = 1)$ is increased. Given that the best fitting aCVAE model was based on $p(C = 1|y)$ – Weighting a higher weight could have accelerated recalibration. However, since our model comparison did not support models where $p(C = 1)$ was updated the computational mechanism of this effect rather remain obscure. Regardless of the computational mechanism, the effect of association type on the aCVAE is yet another example of top-down driven influences on multisensory recalibration. Previously, Eramudugolla,et al. (2011) showed that task-load in central positions can influence audio-visual spatial recalibration in the periphery, presumably by narrowing the spatial specifity of recalibration (Bertelson et al., 2006; Kopco et al., 2009).

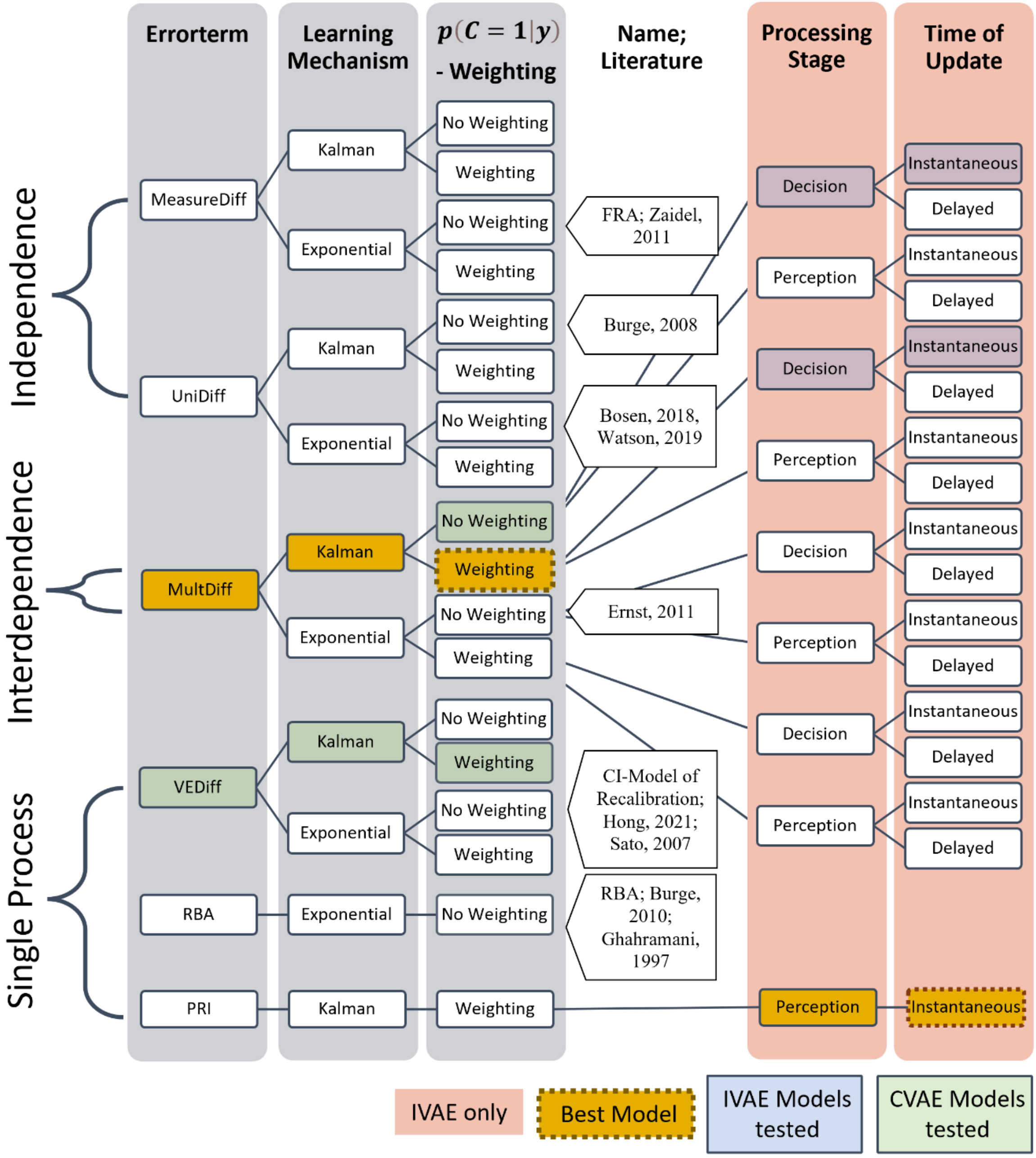


Figure 36. Overview of all tested CVAE and IVAE models. Only a subset of the factors Errorterm, Learning Mechanism and Weighting were realized for the IVAE (blue background) as well as the CVAE (green background). For the IVAE additionally a subset of the levels of Processing Stage and Time of Update were realized, but only for the MultDiff errorterm. In contrast to Study 1, IVAE and CVAE model were jointly fit and compared, hence there is only one overall best model (gold background) comprising of an IVAE and a CVAE model. Model factors and abbreviations are described in detail in Chapter II.

The dissipation of the effect of association type from first to second intermixed and post blocks was expected since bimodal trials during intermixed blocks would, similarly to the association trials, likely update $p(C = 1)$ for the CCA and DCA pairs towards a similar value of $p(C = 1)$. Previous studies (Odegaard et al., 2017; Tong et al., 2020) tried to overcome this wash-out effect by refreshing the association within the test phase. This was not possible in the present study since any refreshing trials especially for the CCA pair could have had overwritten the effects of recalibration. It would be an interesting question for future studies, whether this wash-out effect is simply driven by dissipation over time, i.e., $p(C = 1)$ would drift back to an initial and by association unaffected representation, or explicit counterevidence is necessary to override the newly acquired representation of $p(C = 1)$. In a computational model of the VE (Ursino et al., 2019) the prior $p(C = 1)$ is stored via the strength of crossmodal synapses which implies a stable prior without new sensory evidence.

Computational Principles of Recalibration and the Role of Causal Inference

The Effects of Absolute Disparity

The present study clearly showed that VE, IVAE as well as aCVAE increased with increasing absolute disparity. This increase was well in line with Causal Inference for the VE and the IVAE indicated by the good fit of the posterior simulations (Figure 32 and Figure 35) and model comparison results clearly in favor of a common process of IVAE and VE. In contrast to the results presented in Study 1, none of the participants were better described by a model that does not include an IVAE at all and only an estimated 17% of the participants was better described by an IVAE model based on the MultDiff errorterm and No Weighting. Although the latter model also predicted an increase of the IVAE, it significantly underestimated the IVA on a group level and participant wise predictions of the IVAE were worse than for the single process model. Taken together, these results provide strong evidence in favor of an IVAE based on Causal Inference, which is well in line with the finding of Wozny et al. (2011a), that demonstrated an increased IVAE when participants reported the perception of a common cause in the preceding trial. One might argue that, given the effect of association on the VE, an effect of association on the IVAE should also be observed. Since however, the VE is in general multiple times larger than the IVAE one would expect the modulation to be smaller, which implies that the proposed change in $p(C = 1)$ might not have been large enough to induce a statistically detectable effect on the IVAE.

Our initial hypotheses with respect to the aCVAE were based on a predicted interaction between absolute disparity and association type, and assuming similar effects to the ones observed by Tong et al. (2020) clearly distinct result patterns for models based on $p(C = 1|y_k)$ – Weighting or no Weighting would have been expected. But given that the effect of association type was small essentially all competing models predicted an increase of the aCVAE with increasing absolute disparity at least on a group level. However, participant-wise model comparisons revealed that although the best model based on No-Weighting predicted the group-level aCVAE well, the best $p(C = 1|y_k)$ – Weighting model was favored by the model comparison results as well as participant wise fits of posterior simulations. These results contribute to the clarification of the role of Causal Inference in cumulative recalibration in the light of diverging previous results (compare Study 1 and Hong et al., 2021). In Study 1 most participants were best described by a model assuming no Weighting. Hong et al. (2021) similarly to Study 1, manipulated the visual reliability and observed an increase of the aCVAE with decreasing reliability, which could only be explained by $p(C = 1|y_k)$ – Weighting. The decrease in visual reliability presumably increased $p(C = 1|y_k)$ and thereby accelerated recalibration. If we assume an audio-visual discrepancy that can be reliably perceived, decreasing the visual reliability leads to a wider range of position, where the true stimulus could have been. Hence it is also more likely that it had been closer the auditory stimulus, which is what increases $p(C = 1|y_k)$. The causal structure of the stimuli might have been more salient in the present study then in Study 1. We used highly reliable visual stimuli, therefore variations in the audio-visual disparity led to large changes in the $p(C = 1|y_k)$. This is because changes in the audio-visual disparity are also reliably detected under such conditions. A disparity of several degrees might be still in line with a single cause, but since the sensory evidence is so reliable, a few more degrees of separation are sufficient to make a common cause highly unlikely. In Study 1, the effect of a decrease of visual reliability while keeping the relatively small audio-visual discrepancy (13.5°) constant, might have been more subtle. Given the small discrepancy, $p(C = 1|y_k)$ is already quite high and decreasing the visual reliability would only slightly increase $p(C = 1|y_k)$. Thus, it might be that the perceptual system has some flexibility with respect to whether it takes the causal structure into account or not. Salient differences would highlight the importance of the causal structure in the current context and therefore increase the likelihood of $p(C = 1|y_k)$ – Weighting. The results of Hong et al. (2021) are well in line with this assumption. The reported increase of the aCVAE with lower visual reliability suggests that $p(C = 1|y_k)$ was considerably low when visual reliability was high. Conversely, as visual reliability decreased, the likelihood of a common cause increased, indicating an almost categorical shift from nearly complete segregation to progressively integrated audio-visual spatial percepts.

The Timescales of Recalibration

The adaptation of the approximate Bayesian approach from Study 1 allowed to fit IVAE and aCVAE models based on Causal Inference in parallel for the first time to our knowledge. Note that in Study 1 aCVAE and IVAE were modelled based on the same data but in a stepwise procedure. Previous computational studies that tried to model aCVAE and IVAE in parallel had used simple double exponential learner (Bosen et al., 2018; Watson et al., 2019) which essentially ignored the effects of multisensory integration on recalibration (see Study 1, Chapter III; Hong et al., 2021; Mahani et al., 2017). Importantly, we were able to validate the results of Study 1, by showing that the IVAE was once again best described by assuming a common process for multisensory integration and immediate recalibration. Moreover, in line with the results of Study 1, the aCVAE models with the difference between multisensory percepts as errorterm (MultDiff) best described cumulative recalibration.

With respect to the temporal dynamics of the aCVAE and IVAE, our results are well in line with the general proposition of one slowly emerging but stable and another rapidly emerging but rather transient learning mechanism. The aCVAE continued to increase from the first to the second block indicating that the aCVAE indeed cumulated over time. In contrast, the aIVAE only cumulated over consecutive trials when the previous disparity was in the opposite direction as cumulative recalibration. When the previous disparity was in the same direction as cumulative recalibration the aIVAE ceiled almost within one trial (Figure 33), which points towards a particular importance of the last preceding trial (Bruns & Röder, 2015; Mendonça et al., 2015). The distinct temporal dynamics of the aIVAE compared to the aCVAE depending on the direction of the previous disparity have been reported in a similar manner by Bruns et al. (2015), but still seem somewhat contradictory to the idea that the IVAE builds up over time. An explanation for the lack of increase of the IVAE over consecutive trials might be a hard wired cap for auditory spatial recalibration in general (Wozny & Shams, 2011a). In favor of this interpretation, we observed that the aIVAE increased over consecutive trials throughout the first recalibration block (Figure 33) but not within the second recalibration block where on average the auditory shift is already larger than in the first block. We hypothesize that the aIVAE might reach a cap in direction of cumulative recalibration quite fast, within the second block even after a single exposure. In contrast, the effect of preceding audio-visual disparities in the opposite direction seems to gradually build up over consecutive trials. However, we want to stress that this interpretation is rather exploratory, since the distinct pattern of the aIVAE with respect to the block number was not statistically confirmed, yet previous results showing no significant increase of cumulative recalibration after 180 trials are also well in line with a hardwired cap of recalibration (Frissen et al., 2012).

The Transfer of Instantaneous Recalibration

Our results indicate that the IVAE only partially transferred across the auditory stimuli that differed perceivably in their sound frequency in this study. According to our recalibration models, partial transfer of the aIVAE across auditory stimuli leads to a reduced effect of audio-visual disparities in the opposite direction to cumulative recalibration since the contributing stimulus pairs included tones of different frequencies. This reduced transfer manifests in aIVAEs that on average point in the same direction as the aCVAE, which is exactly what we observed in the behavioral data. Moreover, partial transfer should lead to a slower reduction of aftereffects in the direction of the aCVAE over consecutive trials compared to complete transfer. Importantly, this slower reduction might have prevented the IVAE from flooring within one or two trials in the opposite direction of the aCVAE leading to a negative slope different from zero. Moreover, the reduction of the aIVAE shown here and in a previous study (Bruns & Röder, 2015) demonstrated that the aIVAE does not only passively decay over time (Bosen et al., 2018) but actively takes counterevidence into account. In conclusion, considering the newly observed partial transfer of the aIVAE as well as the assumption of a general cap of auditory spatial recalibration, our results for the aIVAE are well in line with a very fast learning mechanism that is tuned to volatile sources of inaccuracies.

Perceptual vs. Response Level

Previous studies (Bruns, Liebnau, et al., 2011; Wozny & Shams, 2011a; Zierul et al., 2017) have demonstrated that the aCVAE is rather associated with changes of early perceptual representations. Park et al. (Park et al., 2021) found that in elder adults the aIVAE might in contrast also be driven by memory related processes due to decreased sensory precision. Although no such memory related effects were found for healthy adults. Furthermore, Zaidel et al. (2021) demonstrated that visuo-vestibular recalibration triggered by explicit feedback engages neural plasticity in the ventral parietal area, a region that is primarily linked to decision-making processes rather than sensory processing. Both results point towards the possibility of different processes than genuine perception are involved in recalibration. Our computational results however underline the proposition of rather genuinely perceptual processes in healthy adults that alter due to the aIVAE. However, it remains unclear whether the same perceptual representations are altered by the aIVAE and aCVAE. Recent EEG studies point towards partially dissociable neural circuits underlying the aCVAE (Bruns, Liebnau, et al., 2011; Zierul et al., 2017) on the one hand and aIVAE (Park & Kayser, 2019) and VE (Bonath et al., 2007) on the other. Given that the aIVAE emerges at longer latencies (200ms, Park & Kayser, 2019) compared to the aCVAE (Bruns, Liebnau, et al., 2011), it is possible that the aCVAE emerges at early representations and that the associated changes propagate to later representations that also encompass the aIVAE. The build-up of the aIVAE on top of the aCVAE observed during induction blocks would argue for such a propagation as well as the assumption of a common upper bound for the absolute magnitude of auditory perceptual shifts. We observed a direct effect of the aIVAE on the VE which might once again point towards a common underlying spatial representation. However, if the aCVAE is propagated through the auditory spatial processing hierarchy, one would also expect the aCVAE to affect the VE as indicated by our posterior simulations (Figure 35), which we did not find in this study. Note, however, that our manipulation of association might have masked any small changes of integration induced by the aCVAE. The aCVAE was largest for the CCA pair and the large audio-visual discrepancy increasing from block one to block two. At the same time the effect of association, i.e., increasing $p(C = 1)$ dissipated leading to a decreased VE. This effect might have counteracted the shift in bimodal trials induced by the aCVAE. Moreover, Wozny et al (2011a) found alterations in bimodal localization after cumulative recalibration that could be best explained by likelihood shifts attributed to remapping of early auditory representations. The exact number of audio-visual training trials is unfortunately not reported in this study (Wozny & Shams, 2011a). However, from their procedure it follows that participants were exposed to at least 2500 trials with the same audio-visual disparity, suggesting that extensive training might be necessary until the aCVAE affects the aVE. On the other hand, especially in the light of the latter study we perceive our results as evidence in favor of an alteration of bimodal perception by the aIVAE rather than evidence against contributions of the aCVAE to bimodal perception.

Conclusion

The present study confirmed one of the core normative predictions about the relation between multisensory integration and recalibration: Recalibration and integration should both only occur when multisensory signals likely stem from the same source. In combination with the previous study of Chapter III the results suggest that cumulative recalibration flexibly accounts for the causal structure of the scenery especially when differences in causal structures are salient, whereas for immediate recalibration Causal Inference seems to be the preferred default strategy. Moreover, we found a direct effect of the aIVAE on multisensory integration. Both findings highlight the close entanglement of aIVAE and VE. The computational results further validate previous findings of dissociate processes for cumulative and immediate recalibration and thereby support the proposition that the perceptual system encompasses multiple learning mechanisms that are fine-tuned to the temporal volatility changes in the sensory cues.

Chapter IV - Study 2

Causal Inference Differentially Affects Cumulative and Instantaneous Recalibration

Chapter V – Study 3

Feedback Modulates Audio-Visual Spatial Recalibration[*]

[*] This work has been published: Kramer, A., Röder, B., & Bruns, P. (2020). Feedback Modulates Audio-Visual Spatial Recalibration. *Frontiers in Integrative Neuroscience*, *13*. https://doi.org/10.3389/fnint.2019.00074

Introduction

When spatially interacting with our environment, vision and audition communicate in multifaceted ways to guide attention (Driver & Spence, 1998), enhance spatial acuity (Bolognini et al., 2007), and form a coherent representation of our environment. In order to benefit from multiple sensory sources, the signals must be integrated across sensors. Spatial proximity is one of the main cues to decide whether or not two signals belonged to the same event (Holmes & Spence, 2005). In the case of audio-visual spatial perception, assessing spatial proximity is a strikingly complex task, as spatial representations in vision are directly provided by the retina (in eye-centered coordinates), whereas in audition spatial cues emerge from the interaction of the sound waves with the head (Mendonça, 2014) and have to be transformed into a (head-centered) spatial code. It has been argued that the perceptual system uses vision to calibrate auditory spatial perception due to its usually superior spatial resolution and, thereby, resolves misalignments between sensory representations (Bertelson et al., 2006; King, 2009; Knudsen & Knudsen, 1989; Kopco et al., 2009; Radeau & Bertelson, 1974) . Misalignments between sensory representation typically arise during development due to changes in interocular and interaural distance and head size. However, multisensory calibration is not limited to development but rather a lifelong process (Gilbert et al., 2001).

A vivid example of crossmodal recalibration in adults is the cumulative ventriloquism aftereffect (CVAE), in which exposure to audio-visual stimuli with a consistent spatial discrepancy induces a subsequent shift in unisensory auditory localization (Radeau & Bertelson, 1974). The VAE can be induced with various audio-visual exposure durations ranging from a single exposure (Bruns & Röder, 2015; Wozny & Shams, 2011a) over an exposure lasting for several minutes (Bruns, Liebnau, et al., 2011; Lewald, 2002; Recanzone, 1998) to several days (Zwiers et al., 2003). With longer adaptation times, the size of the aftereffect increases (Frissen et al., 2012). The size of the aftereffect is usually only a fraction of the original audio-visual discrepancy (10% - 50%) (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009). More drastic interventions such as the use of prisms over days (Zwiers et al., 2003) to weeks (Bergan et al., 2005) while continuously interacting with the environment have been shown to result in a stronger and more complete realignment of audition with the new visual world.

In case of the VAE, the mere existence of an audio-visual discrepancy implies that at least one of the sensory estimates must be inaccurate. However, without external feedback, the perceptual system cannot infer which sensory estimate was inaccurate and, thus, which sensory representation should be recalibrated (Zaidel et al., 2013). While the CVAE as a form of recalibration manifests in subsequent unisensory shifts, auditory localization is also biased towards vision during audio-visual stimulation, referred to as the ventriloquism effect (VE). Studies investigating such immediate effects as examples of multisensory integration have found that a unified multisensory percept is formed as a weighted average based on the precision of the individual cues, which is considered optimal since such a combination rule maximizes the precision of the multisensory percept (Alais & Burr, 2004; Ernst & Banks, 2002). It has been demonstrated that auditory localization accuracy is positively correlated with precision along the horizontal plane (Garcia et al., 2017). If accuracy is correlated with precision and precision is directly accessible to the perceptual system (Ernst & Luca, 2011), some authors have argued that it would be beneficial if recalibration was based on the reliability of the individual cues, too (Burge et al., 2010; Ghahramani et al., 1997; Makin et al., 2013; van Beers et al., 2002). However, precision does not necessarily imply accuracy (Ernst & Luca, 2011). Thus, several authors have argued that the perceptual system forms prior beliefs about the accuracy of individual senses which are independent of precision (Block & Bastian, 2011; Ernst & Luca, 2011). Recalibration is then assumed to be based on the prior beliefs about accuracy rather than on current reliability. Accordingly, it has been proposed that sensory estimates are adapted according to a fixed ratio (fixed-ratio adaptation) which is relatively stable over time and independent of short-term variations in sensory precision (Zaidel et al., 2013). Crossmodal recalibration consistent with a fixed-ratio adaptation was indeed observed in visual-vestibular motion perception (Zaidel et al., 2011).

Regardless of whether recalibration is reliability-based or follows a fixed-ratio, it would lack external validation in a purely sensory context in which accuracy can only be inferred either from the same cues that are subject to recalibration, which would be circular, or from prior beliefs that can turn out to be wrong when the environment changes. Several authors have argued that this circularity can only be overcome by the use of external feedback which provides independent information about the state of the world (Di Luca et al., 2009; Zaidel et al., 2013). While it is known that unisensory and sensorimotor perceptual learning is susceptible to external feedback (Adams et al., 2010), to our knowledge only one study has investigated whether crossmodal recalibration is modulated by external feedback (Zaidel et al., 2013).

Zaidel et al. (2013) demonstrated that, unlike recalibration without external feedback (unsupervised recalibration), crossmodal recalibration depended on cue reliability when external feedback about the sensory accuracy was provided which was based on the spatial location of one of the two sensory cues (supervised recalibration). In a visual-vestibular motion CVAE paradigm, Zaidel et al. (2013) manipulated visual reliability such that it was either set higher or lower than vestibular reliability. Feedback was either given based on motion implied by visual motion stimuli or based on vestibular motion stimuli which were presented simultaneously. Whereas unsupervised recalibration was independent of cue reliability (Zaidel et al., 2011), supervised recalibration was found to be based on the discrepancy between the multisensory (i.e., integrated) percept and the location indicated by feedback. As the multisensory percept in visual-vestibular motion perception is highly dependent on cue reliability (Fetsch et al., 2009; Gu et al., 2008) supervised recalibration therefore also depended on cue reliability. Zaidel et al. (2013) argued that both mechanisms together result in accurate, precise and consistent multisensory and unisensory representations of space. The idea is that unsupervised recalibration aligns sensory modalities, thereby providing a consistent representation of space, and supervised learning realigns this internally consistent representation with the external world.

However, in order to accept these ideas as a general rule, it has to be demonstrated that they hold for other combinations of sensory modalities such as for audio-visual stimulation. In fact, empirical results have suggested that audio-visual spatial recalibration in the CVAE might be unaffected by top-down processes. For example, the CVAE did not differ between audio-visual trials which included matching voices and faces or percussion sounds and a video of hands playing bongo, compared to trials in which the visual stimulus was simply a synchronously modulated diffuse light (Radeau & Bertelson, 1977, 1978). Furthermore, although attentional load was found to influence the spatial pattern of the CVAE, the overall size of the CVAE remained unaffected (Eramudugolla et al., 2011). These results were taken as evidence for the idea that the CVAE is largely independent of top-down effects such as attention. In accordance with this proposal are findings that the CVAE occurs even when participants are asked to ignore visual stimuli or become aware of the audio-visual discrepancy (Bertelson, 1999). However, it is not known whether the CVAE is modulated by external feedback regarding the spatial accuracy of either the auditory or visual cue. In fact, such feedback would be a crucial prerequisite to guarantee external accuracy of perception, that is, a correct relation between sensory representations and the external world.

In order to test whether crossmodal recalibration is affected by external spatial feedback, we extended the classical CVAE paradigm (Radeau & Bertelson, 1974; Recanzone, 1998) by introducing feedback similar to that employed by Zaidel et al. (2013). During an audio-visual block, participants had to localize audio-visual stimuli with a fixed spatial discrepancy. In contrast to previous studies, feedback about the localization error was provided. Each participant completed four sessions and in half of the sessions feedback in audio-visual blocks was calculated based on the discrepancy between the participant's response and the true visual position, and in the other half of the sessions feedback was based on the discrepancy between the participant's response and the true auditory position.

As there are a few reports of visual aftereffects in the ventriloquism paradigm (Lewald, 2002; Radeau & Bertelson, 1976) which could potentially be increased by feedback that is based on the auditory stimulus position, we tested both auditory and visual unimodal localization before and after the audio-visual block to assess both auditory and visual aftereffects. Based on the assumption that feedback would update the perceptual system's beliefs about the accuracy of the involved sensory cues, we hypothesized that the CVAE would decrease for the sensory modality that feedback was based on. The opposite effect was expected for the other modality for which feedback did not indicate the true stimulus location. Moreover, as accuracy was found to be correlated with precision in audition (Garcia et al., 2017) and precision modulated effects of feedback in visual-vestibular recalibration (Zaidel et al., 2013), we additionally tried to manipulate the reliability of the visual stimulus. In accordance with Zaidel et al. (2013), we hypothesized that recalibration in the presence of feedback is based on relative cue reliabilities. Hence, the CVAE would be increased for the less reliable sensory modality.

Methods

Participants

In order to counterbalance all control conditions (see *Procedure* for details), we were restricted to multiples of 24 for our sample size. We aimed for a sample size of 24 participants, which has 80% power (at an α level of .05) to detect a medium-sized effect (d_z = 0.52) for a directional difference between two within-subject conditions (corresponding to our main hypothesis that the CVAE is reduced when feedback is based on the auditory position rather than on the visual position). The power analysis was conducted in G*Power 3.1 (Faul et al., 2009).

A total of 37 healthy adult volunteers were recruited through an online subject pool of the University of Hamburg, because 13 datasets had to be removed from the initial sample due to technical issues which led to a wrong presentation of auditory stimulus locations. All affected datasets were replaced such that complete datasets from 24 participants were acquired. At the analysis stage, six additional datasets had to be excluded from the 24 participants which completed all sessions. One participant reported visual field restrictions in one hemifield after completion of the experiment and had to be removed from the sample. Moreover, five participants had to be removed due to untypically inaccurate responses or poor performance in catch trials (see *Data Analysis* for details).

The remaining 18 participants (4 male, 14 female) were from 19 to 39 years of age (mean: 24.4 years) and reported normal hearing and normal or corrected-to-normal vision. Participants received course credits as compensation. Additionally, participants received monetary rewards (mean = 25.56€, possible min. = 0€, possible max. = 46.80€, empirical min. = 17.55€, empirical max. = 39.60€) as part of the experiment. Written informed consent was obtained from all participants prior to taking part. The study was performed in accordance with the ethical standards laid down in the 2013 Declaration of Helsinki. The procedure was approved by the ethics commission of the Faculty of Psychology and Human Movement of the University of Hamburg.

Apparatus

Experiments were conducted in a sound-attenuated and darkened room. Participants were seated in the center of a semicircular frame (90 cm radius) on which six speakers were mounted at ear level. Hence, all auditory stimuli were presented at the same height. Speaker locations ranged horizontally from 22.5° left from straight-ahead (0°) to 22.5° right from straight-ahead in steps of 9° (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). Participants positioned their head on a chin rest to fix the head position across trials. An acoustically transparent curtain covered the speakers. A schematic illustration of the apparatus is shown in Figure 37. Visual stimulation was provided via four laser pointers which projected a light point onto the curtain for 200 ms. Two laser beams were diffused resulting in circular, red, light blobs with approximately Gaussian luminance amplitude envelopes. The sizes (horizontal and vertical) of the visual stimuli (VS), defined by the standard deviation of the luminance distribution, were 12.84° for the low reliable VS and 2.83° for the high reliable VS. The position of a VS was defined as the center of its luminance distribution. The center of the luminance distribution in the vertical dimension was always at the same height as the speakers. A third and fourth laser pointer were not diffused and purple and green in color. The laser pointers were mounted on a step motor with an angular resolution of 0.9° and a horizontal range of 180°. Auditory stimuli were narrow-band filtered (1/2 octave) pink noise bursts with four different center frequencies (250, 500, 1000, or 2000 Hz) and were presented for 200 ms including 5 ms on- and off-ramps. The stimulus intensity was randomly varied over a 4-dB range centered at 70 dB(A) to minimize potential differences in the speaker transformation functions. Participants localized stimuli with a custom-build pointing stick which recorded azimuthal position with 1° resolution.

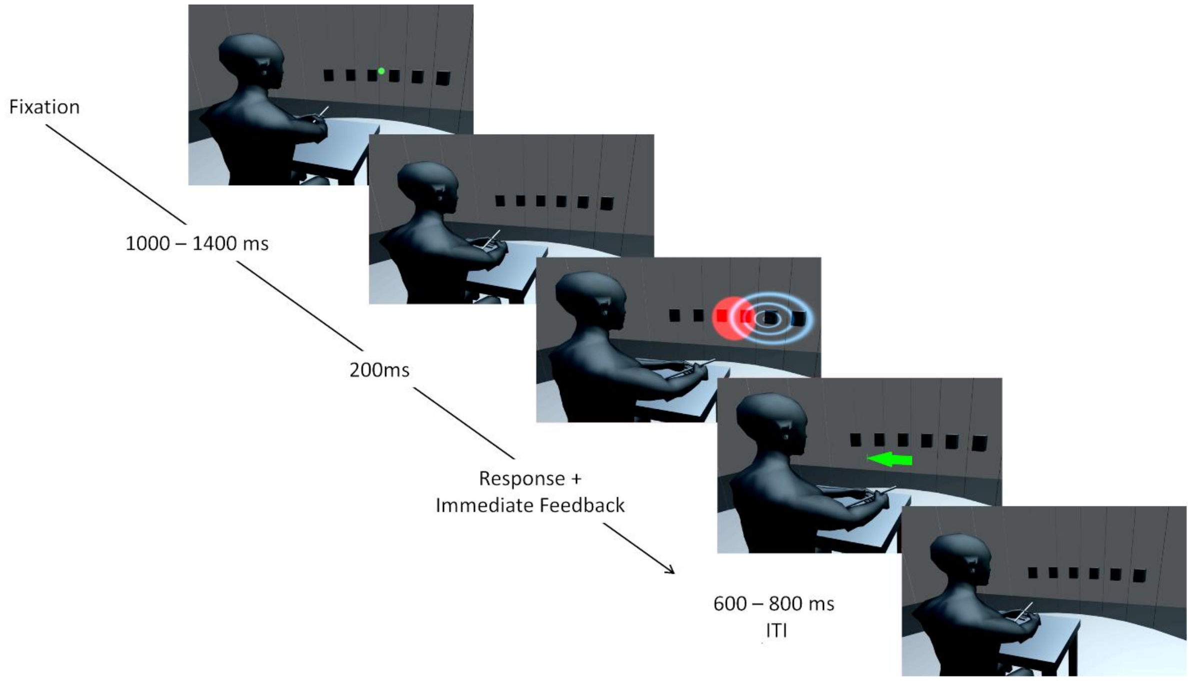


Figure 37. Illustration of the setup and an audio-visual trial. Six speaker positions from -22.5° to 22.5° in steps of 9° are represented by black boxes. The curtain covering the speakers is only transparent for illustration purposes and was visually opaque and only acoustically transparent. A chin rest used to fixate the head is not displayed. At first, a green laser dot appeared as fixation point and participants could start the trial by pointing to the fixation dot and pressing a button. The trial started when the pointing error was below ±10°. During a second interval, a step motor adjusted a second laser used for stimulus presentation. Auditory (indicated by blue waves) and visual (red light cone) stimuli were presented for 200 ms in synchrony. Participants could respond immediately by pointing towards the perceived direction and pressing a button on the pointer. Corrective feedback followed instantaneously in form of a centrally presented arrow. The color of the arrow (green for reward, red for no reward) and a unique sound indicated whether a reward was obtained. After a varying interval (600-800 ms) the green laser dot reappeared, and the participant could start the next trial. Avatar image adapted from "Low Poly Character" by TehJoran, 2011 (https://www.blendswap.com/blend/3408) licensed under CC BY.

To deliver feedback, an LED-panel (APA 102, Shiji Lighting, Shenzhen, China) measuring 32 cm in width and 8 cm in height with a pixel width of 0.5 cm and a spacing of 0.5 cm (2.54 ppi) was attached to the semi-circular frame between ±10.2° azimuth and 2 cm below the lower edge of the speakers. An Arduino Leonardo (Arduino SRL, Strambino, Italy) was used to interface between the experimental computer and the LED-panel.

Procedure

The study was split into four sessions which were conducted on separate (but not necessarily consecutive) days (see Figure 38 A). Each session started with a unimodal pretest to measure baseline localization accuracy and precision for visual stimuli and auditory stimuli presented in isolation. Afterwards, an audio-visual adaptation block (see below) was conducted to induce auditory and potentially visual CVAEs. The adaptation block was followed by unimodal test blocks to assess the magnitude of the aftereffects. To ensure that aftereffects did not decay over unimodal test blocks, each test block was preceded by a short re-adaptation block. The general procedure of a session is illustrated in Figure 38 B.

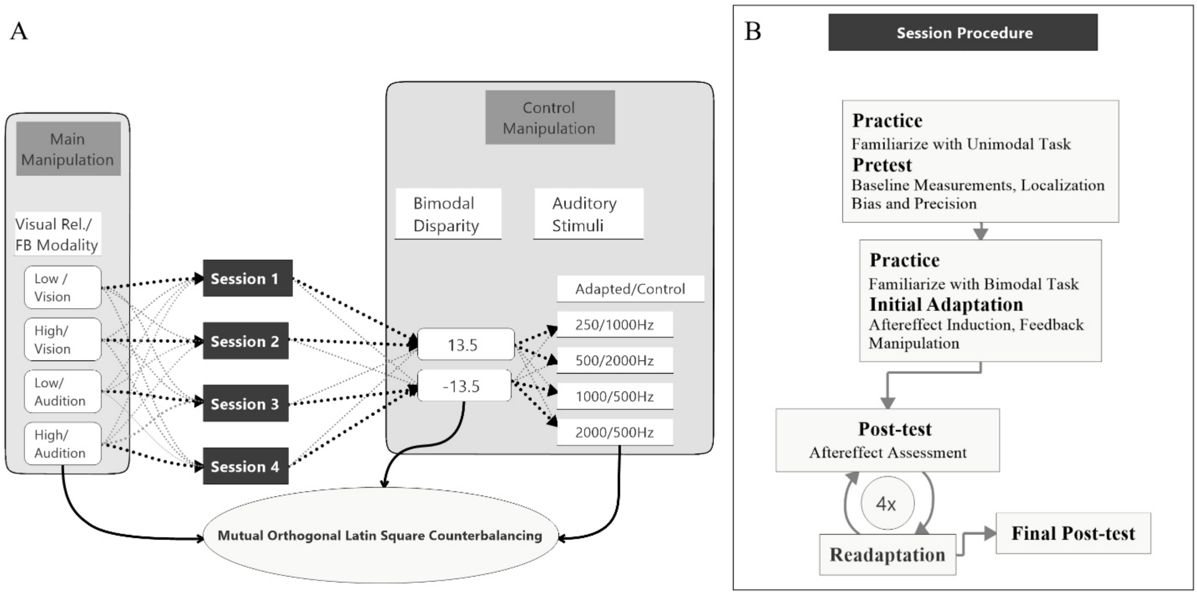


Figure 38. Study design and session procedure. **A**: The flow diagram shows the counterbalancing procedure. An exemplary procedure for one participant is depicted with bold black pointed lines. All possible assignments between the main conditions, session number, bimodal disparity and auditory stimulus pair are depicted with light grey pointed lines. Assignments of main conditions to session number, bimodal disparity and auditory stimulus pairs were mutually counterbalanced by orthogonal Latin squares. **B**: The flow diagram visualizes the procedure of a single session. All four sessions were performed following the same procedure.

Two factors were varied between sessions, the reliability of the visual stimulus (manipulated by the size of the circular light cone) and the feedback modality. During adaptation blocks participants were asked to localize the auditory stimulus and feedback about the magnitude and direction of their localization errors was provided. Error feedback was consistently calculated either based on the position (i.e., center of the luminance distribution) of the visual stimuli (vision feedback modality) or based on the position of the auditory stimuli (audition feedback modality) within each session. All participants completed all combinations of visual reliability (high vs. low) and feedback modality (vision vs. audition) across sessions. The auditory stimuli were grouped into four pairs (250 Hz/1000 Hz, 500 Hz/2000 Hz, 1000 Hz/250 Hz, 2000 Hz/500 Hz) with non-overlapping frequency spectra. The first stimulus of each pair was the adapted auditory stimulus (AS) and was used during both unimodal blocks and audio-visual adaptation blocks. The second stimulus was only used during the unimodal blocks and served as a control stimulus (CS). Thereby, the CS allowed to test for a sound-frequency transfer of the aftereffect. Each session was conducted with a unique pair of auditory stimuli to avoid carry-over effects between session (Bruns & Röder, 2019).

Moreover, to avoid those participants became aware of the audio-visual discrepancy during adaption blocks and, thus, might apply explicit response strategies, in half of the sessions the visual stimuli were consistently displaced to the left and in the other half to the right of the sound source. To avoid effects of session order, auditory stimulus assignment or visual discrepancy direction on the feedback modality and reliability conditions, these factors were counterbalanced across participants using a mutual orthogonal Latin square design (Julian et al., 1996). For factors with four levels (discrepancy was dummy coded by taking each discrepancy twice) three mutual orthogonal 4x4 Latin squares exist, so that there were six possible ways of assigning Latin squares to the three factors (session order, auditory stimulus assignment, visual discrepancy direction). As four participants are necessary to realize one Latin square, in total 24 participants were necessary for a balanced design that realizes all combinations of Latin squares. However, factors relevant for the data analysis (visual reliability and feedback modality) were measured within-subject and, thus, were counterbalanced irrespective of participant exclusion (see *Procedure* for details).

Unimodal Blocks

Unimodal pre- and posttests were identical, except that the posttest was split into several blocks. The two auditory stimuli (AS, CS) were presented from all six speakers (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). One visual stimulus (VS) was presented from the same six positions as the auditory stimuli. Either the low reliable VS or the high reliable VS was consistently used across the whole session according to the counterbalancing procedure. The VS was described to participants as a diffuse light cloud, and they were instructed to localize the center of this light cloud. For each position and stimulus type (AS, CS, VS) 10 trials were presented, yielding 180 trials in total. For the pretests, all 180 trials were presented in a random order. For the posttests, the 180 trials were split into five blocks of 36 trials each. Two trials per position and stimulus type were presented in each block of the posttest. Each trial started with the presentation of a green fixation laser point at 0° azimuth. Participants were required to direct the pointing stick towards the fixation point and started the trial by a button press. The trial only started when the pointing direction deviated less than ±10° from 0°. This procedure assured a constant starting position for all pointing movements. After a random delay between 400 and 600 ms the presentation of the VS was prepared: the step motor carrying the laser pointer was first moved to a random position between -50° and 50° and then moved to the target position. This was done to avoid that the duration of the sound evoked by the moving step motor provided a cue for the VS position. After another delay of 600 to 800 ms, the VS was presented. During AS and CS trials only a random delay between 1000 and 1400 ms was used after fixation, followed by the presentation of the stimuli. Responses were allowed immediately after stimulus onset. Participants were instructed to respond fast and accurately, but to prioritize accuracy over response speed. Moreover, participants were informed that all stimuli (during unimodal and audio-visual blocks) would be displayed at the same height and that they should focus on localizing stimuli accurately in the horizontal plane. No feedback or reward was provided during pre- and posttest trials. Between trials a random delay between 600 to 800 ms was introduced.

Audio-visual Blocks

In order to induce the VAE, the AS and the VS were synchronously presented for 200 ms with a spatial displacement of the VS of either 13.5° to the left or 13.5° to the right of the sound location. The spatial discrepancy was constant during a session. In the initial audio-visual adaptation block, stimuli were presented 20 times at each of six positions (sound at -22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). The four audio-visual re-adaptation blocks (prior to each of the following unimodal posttest blocks) only contained 10 trials per position and were conducted to counteract a potential decay of the aftereffect (for similar procedures, see Bruns et al., 2011; Zierul et al., 2017). Overall, each session included 360 audio-visual adaptation trials and 360 unimodal test trials (720 trials in total). Participants were instructed to localize the sound (i.e., to ignore the visual location) in audio-visual trials. Immediately after the response, feedback about the azimuthal localization error the was given. The localization error was either calculated as the deviation of the azimuthal pointing direction from the true azimuthal location of the AS or as the deviation of the azimuthal pointing direction from the true azimuthal location of the VS. The modality used for calculating the localization error was held constant within a session. Feedback consisted of a centrally presented arrow with the origin at 0° and heading in the direction participants had to correct their localization response to in order to reduce the error. The length of the arrow equaled the magnitude of the localization error in cm rounded to the next integer, with an upper bound of 16 cm (10.2°) and a lower bound of 4 cm (2.55°). Errors below 4 cm (2.55°) were indicated with a filled circle with a radius of 3 cm (1.9°). Furthermore, participants received a monetary reward (0.03€) when the error fell below an individual threshold which was set to the participant's 30th percentile of the absolute localization error in the auditory trials of the pretest. A reward was indicated by a unique sound (400 ms custom rebuild of the Super Mario coin sound effect) and a green feedback arrow or circle. A localization error above the individual threshold was indicated by another unique sound (300 ms tone that changed pitch from 100 Hz to 60 Hz after 150 ms) accompanied by a red feedback arrow. The whole sequence of an audio-visual trial is depicted in Figure 37. After each block participants were informed about the amount of reward they had collected during the block. The total amount of reward was disbursed at the end of the session.

In order to assure that participants attended to both visual and auditory stimuli, deviant trials were presented intermixed between regular trials with a probability of 0.1. In deviant trials, participants were instructed to localize a laser point as fast and accurately as possible that differed in color (purple) and was not accompanied by a sound. The laser point was presented until a response was given. When the reaction time fell below the 50th percentile of the reaction time in visual trials of the pretest and localization error was less than 5°, a reward (0.03€) was earned in these trials. The same visual and auditory feedback was used as for regular trials, except that always circular shapes were used.

Data Analysis

Data were acquired for 24 participants in order to counterbalance control conditions (session order, stimulus assignment and audio-visual disparity). However, overall six participants had to be excluded from further analyses. One participant reported partial vision in one hemifield after the study was completed. Another two participants failed to respond properly to audio-visual deviant trials. The deviant trials required participants to respond fast and accurately (see *Procedure* for details) to receive a reward. Hence, not attending to the visual stimuli or closing the eyes during audio-visual blocks would lead to a low number of rewards in deviant trials. These two participants consistently received rewards in less than 2% of the deviant trials across all sessions, whereas on average participants received rewards in 55% (min. = 15%, max. = 82%) of the deviant trials. Hence, we excluded their data from further analyses. For each of the remaining participants we fitted linear models between true azimuthal stimulus positions and azimuthal localization responses for each session and each stimulus (a slope of one and an intercept of zero indicate perfect localization). Three participants with either a slope or an intercept that differed three standard deviations from the mean of all participants were excluded as this indicated an extremely inaccurate localization behavior. All further data analyses were based on the data of the remaining 18 participants.

Importantly, all factors relevant for further data analyses (i.e., Feedback Modality and Visual Reliability) were still fully counterbalanced after exclusion of the participants. The reduction of the sample size only affected the counterbalancing of session order, assignment of sound pairs to sessions and assignment of audio-visual discrepancy directions to sessions. The final numbers of participants for each combination of these factors are summarized in Table C. 1- 4.

To test whether participants changed their localization behavior in audio-visual adaptation trials according to the error feedback, we took the mean localization error in the first ten adaptation trials of the initial adaptation block and compared this score with the mean localization error of the last ten adaptation trials in the last re-adaptation block. We performed two separate *t* tests for the conditions of feedback modality (audition or vision) comparing the mean of the first ten trials to the mean of the last ten trials.

Measurements for accuracy and reliability were derived from unimodal blocks and based on a common model of measurement error (Grubbs, 1973). Each trial is interpreted as a measurement y_{ik} for the true stimulus position $x_k$where i is an index over the trial numbers and k over stimulus positions. The measurement model is then formalized as

$$\boldsymbol{y_{ik} = x_k + a_k + e_{ik}}, \tag{58}$$

where a_k is a constant bias for the kth stimulus position and e_{ik} are independent mean zero random errors. As an estimator for accuracy, we calculated the constant error $\hat{a}_k$ by averaging localization responses of all trials for each combination of stimulus position, condition and participant. For a given stimulus position this is a robust estimator of the bias term a_k and thus accuracy. We will further refer to $\hat{a} := M(\hat{a}_k)$ as *constant bias*, which is an overall measure for the tendency to systematically mislocalize in one direction across all locations. Reliability is defined as the inverse of the variance of e_{ik}. Due to the direct relation between variance and reliability we assessed the *variable error*, a robust estimator of the variance (Brown & Forsythe, 1974), as a measure for reliability. The variable error is defined as the mean absolute deviation of the localization response from the mean localization response for a given stimulus position, that is, if $\hat{y}_{ik}$ are the participant's responses the variable error is defined as $M(|\hat{y}_{ik} - \hat{a}_k|)$. A high variable error indicates a low reliability and vice versa.

First, we tested whether we were successful in manipulating the reliability of the visual stimuli (high or low) and controlled that auditory reliabilities did not differ prior to adaptation. Therefore, variable errors calculated from all pretest trials were submitted to a repeated measures MANOVA (O'Brien & Kaiser, 1985) with factors Feedback Modality (audition or vision), Stimulus Type (AS, CS, VS), Stimulus Position (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) and Visual Reliability (low or high). This approach is not affected by violations of the sphericity assumption and allows for post-hoc interaction contrasts, which were conducted to further analyze significant MANOVA effects.

The CVAE was measured as change in the constant bias between pre- and posttest blocks. For this purpose, data from the five posttest blocks were pooled. More specifically, the difference of posttest constant bias ($\hat{a}_{post}$) and pretest constant bias ($\hat{a}_{pre}$) multiplied with the sign of the audio-visual discrepancy (Diff_{AV}) was taken as a measure for the CVAE, thus CVAE = $(\hat{a}_{post} - \hat{a}_{pre}) * sign(\text{Diff}_{AV})$ (for a similar procedure see Bruns & Röder, 2019). This procedure assured that aftereffects in the direction of the VS always had a positive sign irrespective of whether the VS was displaced to the left (-13.5°) or to the right (13.5°). The resulting values were submitted to a repeated measures MANOVA (O'Brien & Kaiser, 1985) with Feedback Modality (audition or vision), Stimulus Position (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) and Stimulus Type (AS, CS, VS) as within-subject factors.

Audio-visual Blocks

In order to induce the CVAE, the AS and the VS were synchronously presented for 200 ms with a spatial displacement of the VS of either 13.5° to the left or 13.5° to the right of the sound location. The spatial discrepancy was constant during a session. In the initial audio-visual adaptation block, stimuli were presented 20 times at each of six positions (sound at -22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°). The four audio-visual re-adaptation blocks (prior to each of the following unimodal posttest blocks) only contained 10 trials per position and were conducted to counteract a potential decay of the aftereffect (for similar procedures, see Bruns et al., 2011; Zierul et al., 2017). Overall, each session included 360 audio-visual adaptation trials and 360 unimodal test trials (720 trials in total). Participants were instructed to localize the sound (i.e., to ignore the visual location) in audio-visual trials. Immediately after the response, feedback about the azimuthal localization error the was given. The localization error was either calculated as the deviation of the azimuthal pointing direction from the true azimuthal location of the AS or as the deviation of the azimuthal pointing direction from the true azimuthal location of the VS. The modality used for calculating the localization error was held constant within a session. Feedback consisted of a centrally presented arrow with the origin at 0° and heading in the direction participants had to correct their localization response to in order to reduce the error. The length of the arrow equaled the magnitude of the localization error in cm rounded to the next integer, with an upper bound of 16 cm (10.2°) and a lower bound of 4 cm (2.55°). Errors below 4 cm (2.55°) were indicated with a filled circle with a radius of 3 cm (1.9°). Furthermore, participants received a monetary reward (0.03€) when the error fell below an individual threshold which was set to the participant's30th percentile of the absolute localization error in the auditory trials of the pretest. A reward was indicated by a unique sound (400 ms custom rebuild of the Super Mario coin sound effect) and a green feedback arrow or circle. A localization error above the individual threshold was indicated by another unique sound (300 ms tone that changed pitch from 100 Hz to 60 Hz after 150 ms) accompanied by a red feedback arrow. The whole sequence of an audio-visual trial is depicted in Figure 37. After each block participants were informed about the amount of reward they had collected during the block. The total amount of reward was disbursed at the end of the session.

In order to assure that participants attended to both visual and auditory stimuli, deviant trials were presented intermixed between regular trials with a probability of 0.1. In deviant trials, participants were instructed to localize a laser point as fast and accurately as possible that differed in color (purple) and was not accompanied by a sound. The laser point was presented until a response was given. When the reaction time fell below the 50th percentile of the reaction time in visual trials of the pretest and localization error was less than 5°, a reward (0.03€) was earned in these trials. The same visual and auditory feedback was used as for regular trials, except that always circular shapes were used.

Results

Unimodal Precision

Unimodal pretests were performed in order to assess localization biases and reliabilities for all stimulus types and positions. We evaluated whether we succeeded in manipulating the visual reliability and whether auditory reliability significantly differed across conditions at baseline. Therefore, variable errors at pretest (see *Data Analysis* for a definition) were submitted to a repeated measures MANOVA (O'Brien & Kaiser, 1985) with factors Feedback Modality (audition or vision), Stimulus Type (AS, CS, VS), Stimulus Position (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) and Visual Reliability (low vs. high). Only a main effect of Stimulus Type was found, $F(1,17) = 35.22$, $p < 0.001$, showing that visual reliability was higher than auditory reliability independent of the reliability manipulation (see Figure 39). Since no main effect of visual reliability was found (see Table *3* for full results), this factor was not further considered in the following analyses.

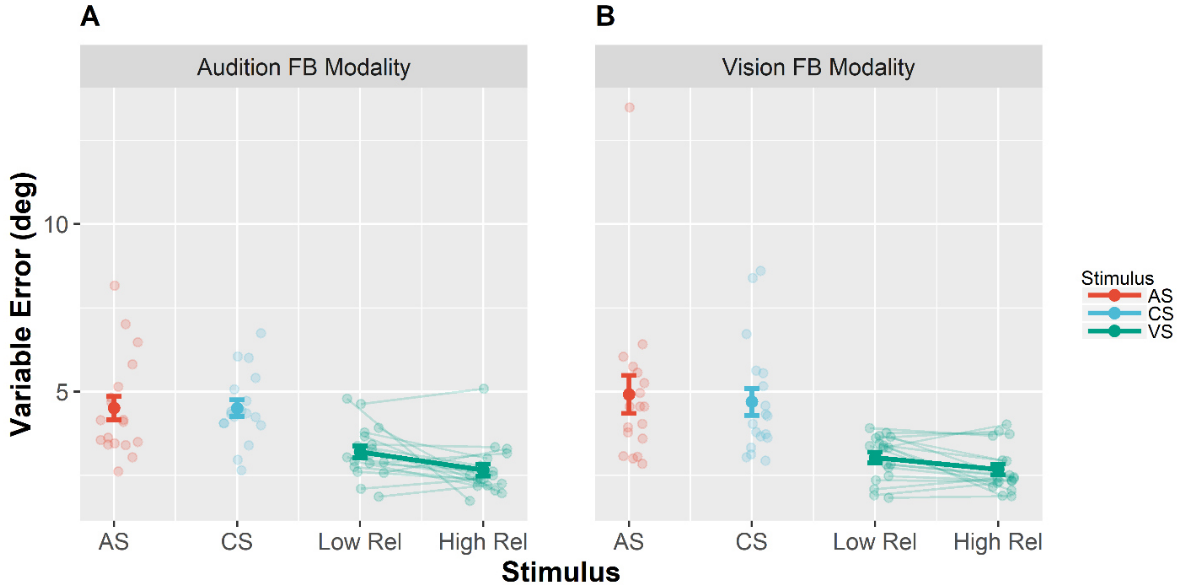


Figure 39. Mean variable errors in the pretest. Variable errors were defined as absolute trial-wise deviation from the mean localization response, averaged across stimulus positions and participants. **A**: Results when audition was the feedback modality. **B**: Results for vision as the feedback modality. Each panel shows the variable error separately for the different stimuli (adapted sound AS, control sound CS, and visual stimulus VS). Moreover, results for the VS are shown separately for the VS with low reliability (Low Rel) and high reliability (High Rel). Individual data are shown with light-colored points and lines, whereas sample averages are indicated by dark-colored points and bold lines. Paired data points (i.e., individual data from a single participant) are connected via lines. Error bars represent standard error of the mean. Mean values are depicted on top of each bar.

Additionally, we performed pairwise contrasts to assess whether the variable error changed from pre- to posttest separately for all stimulus types (AS, CS, VS). Results are summarized in Table 4. Importantly, the variable error did not decrease for auditory stimuli (AS and CS), but it decreased for the VS, both when audition was the feedback modality, $F(1,17) = 16.75, p < .001$, and when vision was the feedback modality, $F(1,17) = 6.43$, $p = .021$.

Moreover, a contrast was performed to test whether in the posttest blocks the variable error differed for the AS between the conditions audition feedback modality (*M* = 4.4°, *SD* = 1.3°) and vision feedback modality (*M* = 4.9°, *SD* = 1.9°). No significant difference was found, $F(1,17) = 2.50, p = .132$.

Table 3
Repeated measures MANOVA on variable errors in the pretest.

Effect	Num Df	Den Df	**Pillai test statistic**	**Approx. *F***	***p***
(Intercept)	1	17	0.93	249.66	<0.001
Feedback Modality	1	17	0.04	0.11	0.43
Visual Reliability	1	17	0.01	0.03	0.74
Stimulus Type	1	16	0.81	27.49	<0.001
Feedback Modality: Visual Reliability	1	17	0.06	0.54	0.29
Feedback Modality: Stimulus Type	1	16	0.10	1.47	0.43
Reliability: Stimulus Type	1	16	0.21	1.92	0.14
Feedback Modality: Reliability: Stimulus Type	1	16	0.04	0.39	0.70

Table 4
Pairwise contrasts for auditory variable errors between pre- and posttest.

Contrast	**Stimulus**	**FB-Modality**	**Mean Variable Error at Pretest**	**Mean Diff-erence**	**Pillai test statistic**	**Approx. *F***	**Num Df**	**Den Df**	***p***
Post-pre	AS	Audition	4.51	-0.14	0.011	0.20	1	17	0.663
Post-pre	AS	Vision	4.92	-0.02	< 0.001	<0.01	1	17	0.942
Post-pre	CS	Audition	4.51	0.16	0.028	0.51	1	17	0.487
Post-pre	CS	Vision	4.70	0.47	0.20	4.33	1	17	0.053
Post-pre	VS	Audition	2.93	-0.49	0.50	16.75	1	17	< 0.001
Post-pre	VS	Vision	2.86	-0.25	0.27	6.43	1	17	0.021

Note all p-values are uncorrected.

Audio-visual Blocks

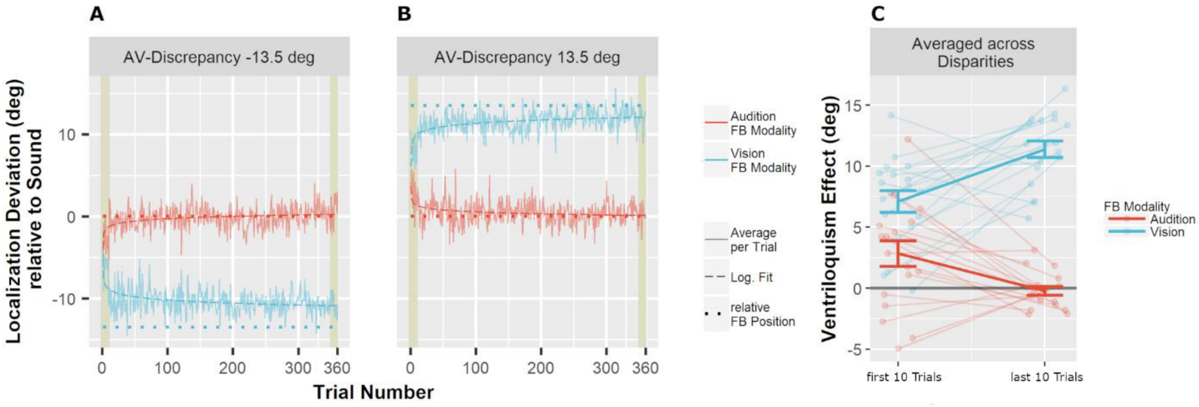


Figure 40. Mean localization deviations in audio-visual adaptation blocks. **A** and **B**: Averages across participants and stimulus positions for each adaptation trial are displayed depending on whether audition (red) or vision (blue) was the feedback modality. Mean deviations were derived by averaging across all participants for one specific trial. The trial number reflects the order of the trials during audio-visual blocks. The position of the sound was used as reference (relative position of 0°). Sessions including an audio-visual discrepancy to the left (-13.5°) are depicted in panel A, and sessions with a discrepancy to the right (13.5°) are depicted in panel B. The actual data (solid line) were logarithmically interpolated (dashed line) to visualize the trend across trials. The relative position that was used to calculate error feedback is indicated by the dotted lines (rel. FB Position). In all conditions, participants adjusted their localization behavior in the direction implied by the error feedback. Participants started with an offset towards the visual position which reflects the well-known ventriloquism effect. The first and last ten trials are highlighted by khaki rectangles. These trials were averaged per participant for statistical analyses. **C**: Localization deviations averaged across the first ten and the last ten audio-visual adaptation trials. Individual data are shown with light-colored points and lines whereas sample averages are indicated by dark-colored bold lines. Paired data points (i.e., individual data from a single participant) are connected via lines. Error bars represent the standard error of the mean. The effect of feedback was very prominent already within the first ten trials (see panels A and B). As a consequence, localization responses already differed at baseline (i.e., over the first ten trials) depending on whether audition or vision was the FB modality (see panel C). Nevertheless, a comparison of the first ten trials and the last ten trials demonstrated a clear effect of FB modality (see text for details).

To test whether feedback altered auditory localization in bimodal trials during adaptation, we calculated the difference of the auditory localization response from the true auditory position. The VE was apparent in a shift of auditory localization towards the accompanying VS (Figure 40). Crucially, when feedback was given based on to the true auditory position, the VE decreased over the course of the adaptation trials. In contrast, feedback based on the visual position increased the VE. To statistically test the change of the VE size over the course of the audio-visual adaptation trials, we calculated the means of the first ten trials and the means of the last ten trials in the audio-visual blocks, multiplied with the sign of the audio-visual discrepancy (thus, a shift of auditory localization towards the VS was always positive). These values were compared with Bonferroni-Holm corrected paired-sample *t* tests. Feedback based on to the auditory position significantly decreased the VE from the first ten trials of the audio-visual block (*M* = 2.8°, *SD* = 4.5°) to the last ten trials of the audio-visual block (*M* = -0.2°, *SD* = 1.5°), *t*(17) = 4.27, *p* < .001. When feedback was given based on the visual position, the bias significantly increased from the first ten trials of the audio-visual block (*M* = 7.1°, *SD* = 3.7°) to the last ten trials of the audio-visual block (*M* = 11.4°, *SD* = 2.9°), *t*(17) = 5.10, *p* < .001.

Table 5
Average reward per session received in audio-visual blocks.

Reliability	**FB-Modality**	**Absolute Mean**	**SD**	**Min**	**Max**	**Rel. Reward**
Visual Rel. low	Audition	6.27	1.86	3.21	10.11	0.58
Visual Rel. high	Audition	6.31	1.92	2.49	10.08	0.58
Visual Rel. low	Vision	6.30	2.34	2.34	10.53	0.58
Visual Rel. high	Vision	6.66	2.39	1.29	10.17	0.62

During audio-visual blocks participants received a monetary reward when the error fell below an individual threshold (see *Procedure* for details). A summary of the received rewards is given Table 5. A repeated measures MANOVA with factors Feedback Modality (audition or vision) and Visual Reliability (low or high) did neither reveal any significant main effects nor a significant interaction of Feedback Modality and Visual Reliability (see Table 6).

Ventriloquism Aftereffect

We next examined whether the magnitude of the CVAE depended on whether feedback was given based on the visual or based on the auditory position (see Figure 41). In contrast to the standard ventriloquism aftereffect for the auditory modality (aCVAE), we will refer to visual aftereffects as *visual Ventriloquism Aftereffect* (vCVAE). A reliable aCVAE was observed for auditory stimuli when vision was the feedback modality. By contrast, no aCVAE was observed for auditory stimuli when audition was the feedback modality. In none of the two conditions a vCVAE significantly different from zero was found. However, mean visual localization responses when vision was the feedback modality compared to when audition was the feedback modality differed significantly.

Table 6
Repeated measures MANOVA on reward in audio-visual blocks.

Effect	Num Df	Den Df	**Pillai test statistic**	**Approx. *F***	***p***
(Intercept)	1	17	0.94	279.26	<0.001
Feedback Modality	1	17	0.02	0.29	0.60
Visual Reliability	1	17	0.02	0.37	0.55
Feedback Modality: Visual Reliability	1	17	0.01	0.14	0.72

A detailed depiction of mean auditory and visual localization behavior can be found in the Supplementary Material (see Figure C. 1 and Figure C. 2). A repeated measures MANOVA (2x3x6) with factors Feedback Modality (audition or vision), Stimulus Type (AS, CS, VS) and Stimulus Position (-22.5°, -13.5°, -4.5°, 4.5°, 13.5°, 22.5°) revealed a significant interaction of Feedback Modality and Stimulus Type, $F(2,16) = 7.14$, $p = .006$. Furthermore, a significant main effect of Stimulus Type was found, $F(1,17) = 11.07$, $p = .001$, as well as a significant interaction between Feedback Modality and Stimulus Position, $F(5,13) = 4.84$, $p = .010$.

Subsequent pairwise contrasts between the two levels of feedback modality separately calculated for the three levels of Stimulus Type (CS, AS, VS) revealed that the CVAE significantly differed for the AS, $F(1,17) = 12.7$, $p < .001$, and the VS, $F(1,17) = 7.91$, $p = .024$, such that the aCVAE for the AS increased when vision was the feedback modality and the vCVAE increased when audition was the feedback modality. No effect of feedback modality was found for the CS, $F(1,17) = 1.36$, $p = .259$. We additionally performed Bonferroni-Holm corrected post-hoc *t* tests to test whether aftereffects were different from zero for each stimulus type and feedback modality. When vision was the feedback modality, significant aftereffects were found for the AS (*M* = 3.2°, *SD* = 2.4°), *t*(17) = 7.05, *p* < .001, and the CS (*M* = 2.1°, *SD* = 1.4°), *t*(17) = 6.21, p < .001, but not for the VS (*M* = -0.6°, *SD* = 1.1°), *t*(17) = -2.52, *p* = .088. No significant aftereffects were found when audition was the feedback modality (see Table 7 for all results).

Table 7
One-Sample post-hoc t tests comparing aCVAE and vCVAE against zero.

Stimulus	**FB-Modality**	**Mean**	**SD**	**t**	**Df**	***p***
AS	Audition	0.53	2.36	0.95	17	0.355
AS	Vision	3.17	1.90	7.05	17	<0.001
CS	Audition	1.16	2.62	1.89	17	0.230
CS	Vision	2.11	1.44	6.21	17	<0.001
VS	Audition	0.65	1.63	1.68	17	0.230
VS	Vision	-0.62	1.05	-2.52	17	0.088

Note all *p* values are Bonferroni-Holm corrected. The aCVAE for the AS and CS as well as the vCVAE for the VS were tested against zero depending on whether feedback was based on the position of the auditory or visual stimuli during audio-visual blocks.

In addition, we performed post-hoc contrasts (Bonferroni-Holm corrected) separately for each pair of stimuli (CS, AS, VS) when vision was the feedback modality, to test whether the aCVAE differed between stimuli. The aCVAE for the AS was larger than the aCVAE for the CS, *F*(1,17) = 12.89, *p* = .009, and larger than the vCVAE for the VS, *F*(1,17) = 46.09, *p* < .001. The aCVAE for the CS was larger than the vCVAE for the VS, *F*(1,17) = 32.84, *p* < .001.

In order to test whether the influence of the feedback modality was greater for the AS than for the CS, we performed an interaction contrast comparing the difference of the aCVAE between the conditions vision feedback modality and audition feedback modality for AS (*M* = 2.6°, *SD* = 3.4°) and CS (*M* = 1.0°, *SD* = 3.4°). The difference between aCVAEs was larger for the AS, *F*(1,17) = 6.65, *p* = .020. These results suggest that the effect of feedback modality generalized to the CS only partially.

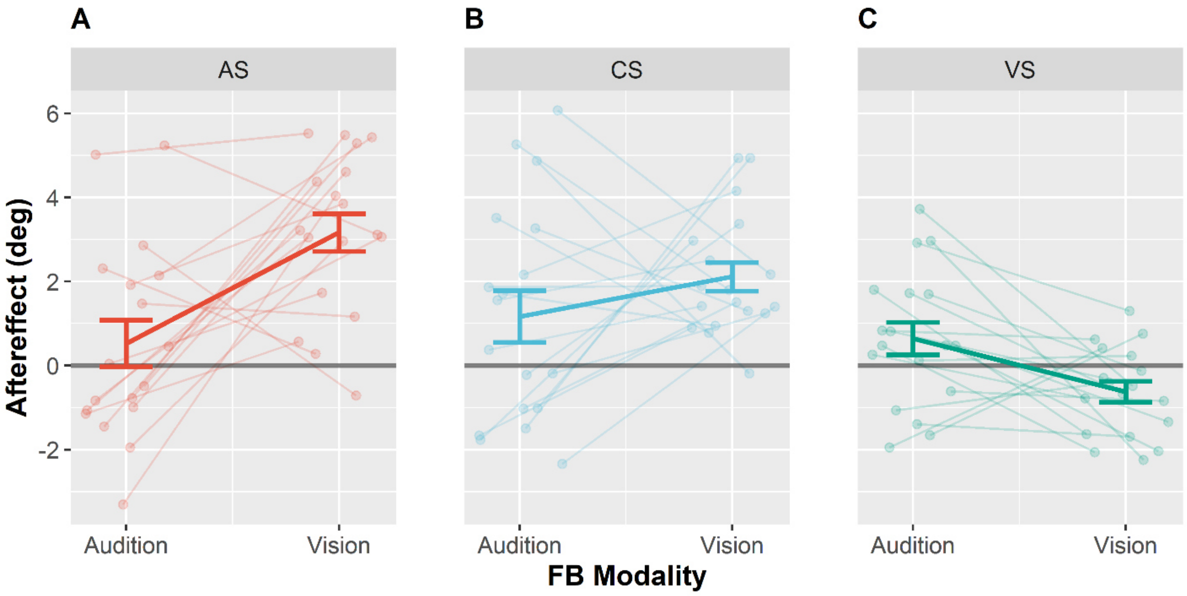


Figure 41. Ventriloquism aftereffects. Aftereffects (CVAEs) were collapsed over leftward and rightward audio-visual disparities for the adapted sound AS (Panel **A**), the control sound CS (Panel **B**), and the visual stimulus VS (Panel **C**). Each panel shows aftereffects separately for the conditions Audition FB modality and Vision FB Modality. Individual data are shown with light-colored points and lines whereas sample averages are indicated by dark-colored bold lines. Paired data points (i.e., individual data from a single participant) are connected via lines. Values were calculated as differences between pre- and posttest localization error multiplied with the sign of the audio-visual discrepancy. Thus, shifts in the direction of the competing stimulus during adaptation are positive. Error bars represent the standard error of the mean.

Discussion

The present study investigated whether crossmodal recalibration, as operationalized with the CVAE, and multisensory integration, as operationalized with the VE, are top-down modulated by feedback. We adapted the standard CVAE paradigm by adding feedback during audio-visual adaptation. By giving feedback either based on the position of the auditory stimuli or based on the position of the visual stimuli, we were able to assess whether feedback modulates the magnitude of the VE and the CVAE. During adaptation, we found that the VE was reduced if feedback was based on the position of the auditory stimulus. A significant CVAE for auditory stimuli was only found when vision was the feedback modality, but not when audition was the feedback modality. Finally, we observed a generalization of the aCVAE to an untrained sound with a different frequency spectrum.

Ventriloquism Effect

The analysis of audio-visual trials during adaptation revealed a clear modulation of the VE by feedback. In the ongoing debate of whether the VE is a rather automatic perceptual process (Bertelson, Pavani, et al., 2000; Bertelson & Aschersleben, 1998; Radeau, 1985) or at least to some degree susceptible to top-down processes (Bruns et al., 2014; Mario Maiworm et al., 2012), our results provide further evidence for the latter assumption. The results show similarities to the study of Bruns et al. (2014) in which it was demonstrated that reward can reduce the VE. In their VE paradigm participants received a monetary reward for precise and accurate auditory localization. Any visual bias induced by the VE was, thus, in conflict to the motivational goal of maximizing the reward. Importantly, the amount of reward depended on the hemifield in which the auditory stimulus was presented. When audio-visual stimuli were presented in the hemifield associated with a high reward, the VE was reduced compared to when the audio-visual stimuli were presented in the hemifield associated with a low reward. Noteworthy, feedback in our study did not only comprise information about the localization error but also a monetary reward when the localization error fell below a threshold. Thus, our findings extend the results of Bruns et al. (2014) by showing that additional corrective feedback can not only reduce but even extinguish the VE when feedback is based on the auditory stimulus position. By contrast, feedback and reward increased the VE when they were based on the visual stimulus position.

One explanation for the modulation of the VE might be that feedback and reward enhanced auditory processing when audition was the feedback modality. It has been shown that feedback can facilitate visual perceptual learning (Herzog & Fahle, 1997) and that reward can facilitate unisensory discrimination performance (Pleger et al., 2008, 2009). Similarly, feedback in our study might have led to an increase in auditory localization reliability. Given that the size of the VE depends on the relative reliabilities of vision and audition (Alais & Burr, 2004; Ernst & Banks, 2002) this would have resulted in a decreased VE. If this was the case, feedback would have modulated multisensory integration via changed bottom-up processing rather than top-down influences. However, we did not find any differences in unisensory auditory localization reliability (indicated by the variable error) between unimodal trials in the pretest and posttest blocks. Moreover, we did not find differences in localization reliability depending on which modality was feedback-relevant either. In fact, only visual reliability increased from pre- to posttest, regardless of whether audition or vision was feedback-relevant. Thus, changes in reliability-based bottom-up processing should have resulted in an increased

VE regardless of which sensory modality was feedback-relevant. Hence, it is unlikely that the decrease or increase of the VE was simply due to altered auditory reliabilities and thus altered bottom-up processing.

In accordance with the present findings, recent studies showing a top-down modulation of the VE did not find changes in unisensory processing. Therefore, the authors (Bruns et al., 2014; Mario Maiworm et al., 2012) argued that it might be the process of crossmodal binding itself that is altered by top-down processing. Binding refers here to the problem of inferring whether two signals have a common or distinct source. For both scenarios different strategies are optimal: If the signals emerged from a common cause, a reliability-weighted average is the optimal estimate (cue integration; see Alais & Burr, 2004; Ernst & Banks, 2002). Otherwise, perceptual estimates should be derived separately from unisensory cues (cue segregation). In fact, the brain seems to form estimates for both scenarios at different stages of the cortical hierarchy (Rohe & Noppeney, 2015). In a further processing step, the probability of a common or distinct cause is estimated and a final multisensory percept is formed as a weighted average of the estimates derived by cue segregation and integration (Beierholm et al., 2010; Körding et al., 2007). Each estimate is weighted by the probability of the underlying model (Körding et al., 2007). This approach has proven to describe the VE well in a range of studies (Beierholm et al., 2010; Rohe & Noppeney, 2015; Wozny et al., 2010) and is referred to as *Causal Inference* (Körding et al., 2007).

In fact, decreasing the binding tendency and relying on unisensory estimates would have been a beneficial strategy in our paradigm. The shift in localization behavior during bimodal trials towards the feedback-relevant sensory modality indicates that participants picked up the relation between sensory modality and feedback. Thus, the feedback-relevant modality might have been identified as task-relevant. It is known that task relevance modulates auditory and visual weights in multisensory integration independently from bottom-up factors such as reliability (Rohe & Noppeney, 2016). This up- or down-weighing might be mediated by attentional shifts towards one modality (Mozolic et al., 2007; Padmala & Pessoa, 2011) or reallocation of cognitive control resources (Pessoa, 2009) to the feedback-relevant modality.

Although the VE seems to be independent from spatial attention, several examples exist in multisensory integration where attentional shifts to a specific modality (rather than to a specific location) lead to decreased integration of task-irrelevant stimuli presented in another modality (Johnson & Zatorre, 2005; see Keil & Senkowski, 2018 for a review). Recent studies have demonstrated that audio-visual integration occurs at different stages of the cortical hierarchy in parallel (Calvert & Thesen, 2004; Rohe & Noppeney, 2015) and that these different stages are associated with distinct computational principles (Rohe & Noppeney, 2015, 2016). It has been argued that multisensory integration associated with late processing stages might be prone to top-down modulation whereas integration associated with early stages might be more or less automatic (Koelewijn et al., 2010). Following this argument, feedback might have modulated late stages of the cortical hierarchy which are linked to audio-visual percepts based on Causal Inference (Aller & Noppeney, 2019; Rohe & Noppeney, 2015).

The importance of top-down processing seems to increase when tasks include motivational incentives, monetary reward (Bruns et al., 2014; Rosenthal et al., 2009), emotional valence (Mario Maiworm et al., 2012) or avoiding harm (Shapiro et al., 1984). For instance, the sound-induced flash illusion was only susceptible to feedback when feedback was accompanied by a reward (Rosenthal et al., 2009). Similarly, explicit knowledge of a spatial discrepancy between audition and vision did not alter the VE (Bertelson & Aschersleben, 1998). However, here we show that corrective feedback paired with a monetary reward clearly increased or decreased the VE depending on whether audition or vision was feedback relevant.

Ventriloquism Aftereffect

In order to maintain accuracy, the perceptual system must infer which sensory modality is inaccurate and to what extent. Ideally, each sensory modality should be recalibrated according to the magnitude of its inaccuracy. In the standard CVAE paradigm audition is calibrated towards vision which can provide internal consistency (Kopco et al., 2009; Pages & Groh, 2013; Radeau & Bertelson, 1974; Zaidel et al., 2011). However, when audition is accurate, and vision is biased, recalibrating audition towards vision introduces inaccuracies in the perceptual system.

As predicted by the assumption that the maintenance of accurate sensory modalities is the primary objective of crossmodal recalibration (Block & Bastian, 2011; Di Luca et al., 2009; Zaidel et al., 2013), we found that feedback based on audition can suppress the aCVAE. Hence, the perceptual system did not recalibrate auditory spatial perception when feedback implied that audition was already accurate. By contrast, when vision was feedback-relevant a substantial aCVAE of 23.5% of the size of the audio-visual discrepancy (13.5°) was found. We did not provide direct sensory feedback (as often used in sensory-motor adaptation paradigms) about the true stimulus position which would have allowed the perceptual system to infer sensory prediction errors in a bottom-up manner (Izawa & Shadmehr, 2011). Instead, a centrally presented arrow indicated magnitude and direction of the localization error, requiring participants to consciously infer the semantic meaning of the feedback. Hence, feedback must have modulated crossmodal recalibration in a top-down manner.

In contrast to our assumption that external accuracy drives recalibration, one could argue that the aCVAE in our study followed the principles of reliability-based adaptation (Burge et al., 2010; Ghahramani et al., 1997; Makin et al., 2013; van Beers et al., 2002). Feedback might have facilitated unisensory auditory processing, as has been shown in unimodal experiments (Pleger et al., 2008, 2009), and, thereby, increased auditory reliability. Thus, according to this assumption audition would be weighted more in the recalibration process, leading to less recalibration. Analogously to our results for the VE, it is unlikely that changes in reliability could explain the results as we did not find an increase in auditory localization reliability between pretest and posttest and reliability in AS trials did not differ depending on which sensory modality was feedback-relevant.

Zaidel et al. (2013) proposed that external feedback invokes a second recalibration process which is superimposed on unsupervised crossmodal recalibration without external feedback and relies on cue reliabilities. Hence, both processes occur in parallel when feedback is present. According to Zaidel et al. (2013), feedback based on the less reliable sensory modality leads to increased supervised recalibration to an extent that outreaches the effect of unsupervised recalibration. Importantly, supervised and unsupervised recalibration result in shifts in opposite directions for the cue that feedback is based on. This results in an overall recalibration of the less reliable sensory modality away from the reliable sensory modality (negative aftereffect). In contrast to Zaidel et al. (2013), we did not find any significant negative aftereffects although audition was clearly less reliable than vision (Figure 39).

Interestingly, Pages & Groh (2013) argued that the aCVAE without external feedback might be a form of supervised learning itself, whereby vision functions as the supervisor for audition. In line with this assumption, they demonstrated that a aCVAE only occurred when the visual stimuli were presented long enough for participants to perform saccades towards them. When visual stimuli were extinguished before participants could accomplish saccades, no aCVAE occurred. Our results support the assumption that external feedback in audio-visual spatial recalibration needs to provide information about the magnitude and direction of the localization error in order to be effective.

We did not observe a recalibration of vision (a vCVAE) in our study, neither when audition was feedback-relevant nor when vision was feedback-relevant. There are only a few reports of vCVAEs (Lewald, 2002; Radeau & Bertelson, 1976), and even prism adaptation for several weeks usually does not result in visual aftereffects (Welch, 1978). Hence it is questionable whether it is possible to induce visual aftereffects through audio-visual adaptation at all (Lewald, 2002; Welch, 1978; Zaidel et al., 2011). Ernst & di Luca (2011) have argued that in order to stay accurate, the perceptual system has to infer to which extent a sensory discrepancy can be attributed to individual inaccuracies of the contributing sensory modalities. As there is no direct information in the sensory cues allowing to assess accuracy, a way to resolve this assignment problem is to form prior beliefs about the probability of a sensory cue to be biased (bias prior). Sensory recalibration then only depends on the ratio of the bias priors. The lack of visual aftereffects could be explained by a remarkably small bias prior for vision. Our results indicate that it might not be possible to update this bias prior on the time scale and by the type of external feedback that was used in the present study (fixed prior, van Wassenhove, 2013). It has been argued that vision, as the most reliable spatial sense, serves as a reference to calibrate the other senses (Bertelson et al., 2006; Knudsen & Knudsen, 1989; Kopco et al., 2009; Radeau & Bertelson, 1974). If the visual system serves as a reference for other sensory modalities, a fixed prior is beneficial to avoid unstable visual sensory estimates in an ever-changing multisensory environment.

To efficiently recalibrate, the perceptual system must infer whether the discrepancy between two sensory cues is due to sensory inaccuracies or whether the cues simply reflect distinct sources. Ideally, recalibration should only occur when a discrepancy can be attributed to sensory inaccuracies (Mahani et al., 2017). We argue that during bimodal trials the VE might have decreased when feedback was based on audition relative to when feedback was based on vison due to a decreased binding tendency which manifests in a reduced prior probability of a common cause (Körding et al., 2007). Hence the increased probability of distinct causes in bimodal trials might have also reduced recalibration. A recent fMRI study (Zierul et al., 2017) showed that the aCVAE is associated with activity changes in the planum temporale, a region which has also been associated with the VE (Bonath et al., 2007), suggesting that neural circuitries involved in the VE and aCVAE are overlapping (see also Park & Kayser, 2019). Thus, Causal Inference processes might affect the aCVAE via the same neural circuitry as the VE (Rohe & Noppeney, 2015).

In contrast to previous studies (Bruns & Röder, 2015; Lewald, 2002; Recanzone, 1998) we found a significant transfer of the aCVAE to an untrained auditory stimulus (see Figure 41). However, there is an ongoing debate whether the aCVAE is sound frequency-specific (Bruns

& Röder, 2015; Lewald, 2002; Recanzone, 1998) or generalizes across sound frequencies (Frissen et al., 2003, 2005), and generalization might depend on the sensory context in which audio-visual adaptation takes place (Bruns & Röder, 2019). Although a significant aCVAE emerged for the CS, our results indicate that feedback had a specific effect on the auditory stimulus used during adaptation (AS) as the difference of the aCVAE between the conditions vision feedback modality and audition feedback modality was significantly reduced for the auditory control stimulus (CS) which was only presented during pre- and posttest.

In summary, the suppression of the aCVAE by feedback based on audition challenges the assumption that the aCVAE is an automatic process which is independent from top-down influences (Epstein, 1975; Passamonti et al., 2009; Radeau & Bertelson, 1978). Although the aCVAE readily occurs when top-down processing can be excluded (Passamonti et al., 2009), our findings demonstrate that the perceptual system can flexibly integrate external feedback into the process of crossmodal recalibration, highlighting the importance of external accuracy as a driving factor for crossmodal recalibration.

Chapter V – Study 3
Feedback Modulates Audio-Visual Spatial Recalibration*

Chapter VI - General Discussion

Summary

Accurate, precise, and adaptive representations of our environment are paramount for successful interactions throughout the lifespan. To make use of redundant and complementary information from different senses, perceptual systems must solve manifold computational problems. Sensory information is often noisy, inaccurate or ambiguous, leading to discrepancies between sensory inputs when none are present in the scenery, or vice versa. Therefore, sensory discrepancies do not only reflect actual discrepancies in the scenery, but also emerge because of noise or distinct sources of inaccuracies. This thesis investigated the computational strategies of the perceptual system for solving this complex credit assignment problem. In line with previous literature, it was hypothesized that multisensory integration solves the Causal Inference problem and resolves multisensory conflict that is likely due to noise. Yet, the computational principles of recalibration at multiple timescales were less understood, especially with regard to the interdependence of integration and recalibration. Multisensory integration was operationalized with the well-studied VE, whereas recalibration was operationalized by the cumulative (CVAE) and immediate (IVAE) aftereffect of the VE.

In the first study, we varied the reliability of the visual stimulus. The results showed a visual VE when the visual reliability was low, and no visual VE when visual reliability was high. In general, visual, and auditory localization responses in bimodal trials were well explained by the Causal Inference model assuming optimal weighting. No visual aftereffects (vCVAE and vIVAE) were found. The aCVAE was reduced when the visual reliability was low, and best described by a model where recalibration is based on the residual discrepancy between multisensory auditory and visual percepts. Moreover, the learning mechanism seemed to approximate Kalman Filtering, i.e., more sensory uncertainty led to slower recalibration. For most participants, the aCVAE was insensitive to CI. While the aIVAE was also reduced when the visual reliability was low, the behavioral results were best in line with a model that assumed a common process for the aIVAE and aVE, and Kalman Filtering as learning mechanism. Moreover, the aIVAE did not only affect unimodal responses, but also bimodal responses, indicating that immediate recalibration might indeed serve to provide more accurate inputs for multisensory integration. Since we did find a visual VE, we will again use VE when we refer to vVE and aVE in common.

In the second study, we varied the audio-visual discrepancy and the type of association (common cause association or distinct cause association; CCA). aVE, aIVAE and aCVAE increased with increasing audio-visual disparity. Auditory VE and aCVAE were increased for the CCA pair, but this effect dissipated over time. Model comparisons confirmed several findings of Study 1. More specifically, the aCVAE was, again, best described by a model that is based on the residual difference between multisensory auditory and visual percepts (MultDiff errorterm). Further, the behavioral results were best in line with a model that assumes a common process for the aIVAE and aVE. Importantly, the proportion of participants for which the aCVAE followed the principles of Causal Inference was largely increased in comparison to Study 1, suggesting the role of Causal Inference is to some extent context-specific for the aCVAE.

In the third study, reward feedback was given throughout audio-visual recalibration, either consistent with the position of the visual stimulus, or consistent with the position of the auditory stimulus. Moreover, the reliability of the visual stimulus changed across sessions. The effects of the reliability manipulation were negligible. However, during recalibration, the aVE was reduced if feedback was based on the position of the auditory stimulus. A significant aCVAE for the auditory stimuli was only found when feedback was consistent with vision, but not when feedback was consistent with audition. Again, no vCVAE was found.

The model taxonomy in Chapter II provided a strong formalism to investigate the computational interdependence of integration and recalibration, both cumulative as well as immediate. Over two independent studies, the approach confirmed a common process for integration and immediate recalibration. Moreover, the model dissociates cumulative recalibration from immediate recalibration and integration based on the computational principles. The results of Study 1 and Study 2 are in line with previous results indicating two recalibration mechanisms at different timescales (Bosen et al., 2018; Bruns & Röder, 2015; Watson et al., 2019). Importantly, the computational models provide mechanistic insights into how the perceptual system solves the credit assignment problem.

Dissociating Noise and Multiple Sources of Inaccuracy

The PRI model of integration and immediate recalibration extends the Causal Inference model by jointly estimating biases in the sensory input with the position of external sources. The fast build up and decay render the aIVAE suitable to account for volatile changes in the accuracy of the auditory system (Bosen et al., 2018; Noppeney, 2021; Watson et al., 2019). The underlying problem then becomes to dissociate volatility in accuracy from stochasticity of sensory cues (i.e. noise) in the sensory system (Piray & Daw, 2021). In the Bayesian framework, the volatility is captured by the bias prior. As the true bias changes over time, the system becomes more and more uncertain about the true bias, and the prior for the bias becomes wider. The Kalman Filter allows us to model this process explicitly. When a new observation is made, the uncertainty of the bias and the uncertainty with respect to the true position are weighted against each other, and both the bias and the positional estimate are updated in proportion to their uncertainty. Importantly, the bias term is only updated under the assumption of a common cause. As the final perceptual estimates are based on averaging over the common and distinct cause scenario, the bias is only partially updated.

Although aIVAE and aVE are part of a common process in the PRI model, their distinctive tuning to bias volatility and cue stochasticity has several implications. The perceived uncertainty about sensory accuracies, i.e., the width of the bias prior, is an individual parameter that depends on the history of each observer and is independent from other important parameters of CI like the sensory reliabilities. Hence, the size of the aIVAE and aVE do not necessarily have to correlate across observers.

Moreover, while sensory reliabilities can in principle be estimated based on the sensory input itself (Ma & Jazayeri, 2014), some authors suggest that it is also learned over time (Beierholm et al., 2020; Sato & Körding, 2014). Volatility, in turn, cannot be inferred from a single observation, and thus must be learned over time. In principle, increased bias volatility leads to increased autocorrelation between consecutive trials, whereas a decrease in reliability decreases autocorrelation and increases covariance over time, making it possible to dissociate them based on experience (Piray & Daw, 2021). Importantly, increased bias volatility should increase the learning rate for the aIVAE, whereas a decrease in reliability should decrease the learning rate for the aIVAE. Indeed, this pattern has been observed in sensorimotor recalibration (Burge et al., 2008).

Assuming reliability and volatility might be learned based on trial-wise variations of the auditory input , underestimating bias volatility can lead to underestimation of reliability (Piray & Daw, 2021). This could explain the visual overweighting that has been observed in several studies of audio-visual integration (Arnold et al., 2019; Battaglia et al., 2003; Meijer et al., 2019). Furthermore, standard paradigms investigating the VE usually use a wide range of audio-visual disparities, therefore increasing auditory bias volatility. This, in turn, can lead to larger shifts of auditory responses in direction of the visual stimulus than predicted by the standard CI, due to the aIVAE. This implies that future modelling studies, even if primarily interested in the VE, should also account for the aIVAE.

Moreover, learning reliability and volatility in conjunction is generally a significantly harder problem then learning one of them when the other is known (Piray & Daw, 2021). A recent study (Rohlf et al., 2020) in audio-visual spatial perception showed that younger children (under 6 years) showed integration, but no rapid recalibration. These results could be interpreted in terms of a stepwise learning process. Interestingly, a recent study found that rapid audio-visual temporal recalibration only occurred after sensory reliabilities reached an adult-like state (Han et al., 2022). Thus, it might be concluded that the increased reliability, and furthermore, an accurate representation of that reliability, facilitated learning bias volatility as prerequisite for rapid recalibration. Negen et al. (2019) showed that feedback about the true auditory or visual position facilitated audio-visual cue combination, arguing that feedback allowed participants to learn about sensory reliabilities and accuracies. Therefore, this study similarly highlights the importance of accurate knowledge about the reliability and accuracy of sensory cues. Hence, rather than learning the reliability and the volatility of the auditory system jointly, the perceptual system might first learn its own reliability sufficiently accurately allowing for integration, followed by a second step of volatility learning allowing for rapid recalibration.

In contrast to the aIVAE, the aCVAE seems to reflect a computationally distinct and subsequent processing stage from integration (Study 1, Study 2). Nevertheless, according to the model comparisons in Studies 1 and 2, the aCVAE depends on the outputs of integration, since its errorterm is the residual difference between visual and auditory multisensory estimates. Analogous to the aIVAE, the bias prior reflects the current estimate of sensory bias, as well as the uncertainty of this estimate. The narrow prior is tuned to rather slow changes in the sensory accuracies, and mirrors processes with a lower degree of volatility compared to aIVAE. This is well in agreement with the proposed hypothesis that cumulative recalibration should be tuned to accuracies of the contributing senses (Di Luca et al., 2009; Zaidel et al., 2011). From a computational perspective, the higher degree of independence of the aCVAE from the VE in comparison to the aIVAE seems very reasonable. As laid out earlier, estimating volatile biases without accounting for noise leads to overestimation of the bias, and conversely, estimating noise without accounting for volatile biases leads to overestimation of sensory uncertainty. This ambiguity makes it reasonable to account for both sources of uncertainty jointly via VE and aIVAE. However, for more stable sources of inaccuracy, this ambiguity is highly reduced - since information is accumulated over long periods of time, the noise can be average out. Hence, the aCVAE can handle more stable sources of inaccuracies quite independently from the VE. This suggests the perceptual system is fine tuned to the computational challenges underlying perceptual inference by solving independent problems with independent processes.

Although it has often been argued that recalibration should be beneficial for subsequent multisensory integration, effects of audio-visual spatial recalibration on subsequent bimodal localization have been reported so far only for the aCVAE (Wozny & Shams, 2011a). In Studies

1 and 2, we report initial evidence for an effect of the aIVAE on bimodal localization. Thus, instantaneous and cumulative recalibration might indeed improve multisensory integration.

Interdependence of Cumulative Recalibration and Integration

Recalibration based on the difference of multisensory percepts (i.e., MultDiff errorterm) would predict that full integration should abolish recalibration. This implies that the shifts induced by the VE, aIVAE and aCVAE together do not exceed the actual audio-visual discrepancy. Moreover, large VEs and aIVAEs should decrease the aCVAE (Welch & Warren, 1980). As a natural consequence, cumulative recalibration based on the difference between multisensory percepts should be incomplete, which is well in line with the observed pattern in several studies (Bruns & Röder, 2019; Frissen et al., 2012; Lewald, 2002). Moreover, the results of Study 2 are in line with this hypothesis. Almost complete integration but no aCVAE was observed when the audio-visual discrepancy was small, possibly due to a very small errorterm. In turn, the VE only compensated for a smaller fraction of the large audio-visual discrepancy, yielding large errorterms for the aCVAE. In this condition, reliable aCVAEs were observed. Importantly, uncertainty about the sensory accuracy can vary largely from observer to observer. Hence negative correlations across participants between VE and aIVAE on the one hand, and VE and aCVAE on the other hand, are not a necessity. For instance, smaller errorterms can still produce a large aCVAE when the learning rate is high. However, one study using extensive training (2500 training trials, Recanzone, 1998) showed almost complete recalibration (7.08° with 8° discrepancy). Under such circumstances, models based on the MultDiff errorterm would predict that at least the speed of the build-up of the aCVAE should be negatively correlated with the initial degree of integration.

Causal Inference in Recalibration

While it is now considered well-evidenced that multisensory integration is sensitive to the possible causal structures of the scenery (see Noppeney, 2021 for a review), this was not obvious for instantaneous and cumulative recalibration. Wozny et al. (2011a) reported that the aIVAE was larger when the tone and visual stimulus were perceived as fused in the preceding bimodal trial. The model comparison results of Study 1 and 2 provide further support for a Causal Inference based aIVAE. In fact, both studies suggest that the aIVAE is directly embedded in the Causal Inference process. Moreover, across both studies at a behavioral level, the pattern of the aIVAE followed the pattern of the aVE. Direct support for a CI-based aCVAE, however, was only found in Study 2, where most participants were better described by a model assuming a modulation of recalibration by the posterior probability of common cause.

These results might imply a certain degree of flexibility with respect to the impact of Causal Inference on cumulative recalibration. In Study 1, we used a moderate audio-visual discrepancy of 13.5°, and the manipulation of the visual reliability might have had only moderate influence on the posterior probability of a common cause. By contrast, the influence of the distinct audio-visual discrepancy (9° vs. 22.5°) combined with the distinct association might have led to distinctively perceived causal structures between experimental conditions in Study 2. Hence, causal structure was more behaviorally relevant in Study 2, leading to a higher proportion of participants recalibrating in a manner that is sensitive to the causal structure. Hong et al. (2021) reported an increase of the aCVAE with increasing reliability, which indicates that the posterior probability of a common cause must have been very low when the visual reliability was high. With decreasing visual reliability, the authors argued that a common cause became more likely in their study. This implies an almost categorical difference from full segregation to increasingly integrated audio-visual spatial percepts. On the one hand, clear differences in the perceived causal structure led to more distinct predictions from CI-based models and models that are not CI-based. Hence, the design of Study 2 and of the study of Hong et al. (2021) might simply be more suitable to dissociate these models. On the other hand, it might be that the perceptual system is somewhat flexible and rather takes the causal structure into account when changes are salient. In line with the latter hypothesis, Mahani et al. (2017) found that participants initially integrated audio-visual cues, but through the experiment, switched to a strategy were they selected the most reliable cue. Hereby no recalibration was observed. Model-based analysis revealed that a common cause was unlikely in this condition. However, when auditory, visual, and tactile cues were used in a single session, recalibration occurred, and model-based analysis revealed that this might have been due to an overall increased likelihood of a common cause in this condition.

Malleability of Priors

So far, two studies have shown that the prior of a common cause is malleable in audio-visual spatial perception (Odegaard et al., 2017; Tong et al., 2020). The association paradigm in Study 2, adapted from Tong et al (2020), implicitly relied on statistical learning of multisensory associations (Quintero et al., 2022), assuming that participants accumulated evidence for a common or distinct cause for an audio-visual stimulus pair over time. Although we found a small effect of this association on bimodal and unimodal localization behavior, it dissipated quickly over time. This effect did not provide enough evidence to justify more complex models, which assume a learning process for the prior probability of a common cause. It remains an open question whether cumulative and immediate recalibration are susceptible to changes in the prior probability of a common cause induced by associative learning.

The aim of Study 3 was to test whether audio-visual recalibration and integration are sensitive to explicit feedback about sensory accuracies. It was assumed that spatial feedback consistent with the auditory position would decrease uncertainty about the accuracy of the auditory system. In line with this assumption, we found a decreased aCVAE when feedback was consistent with auditory stimuli compared to visual stimuli. In Kalman Filter models of recalibration, decreased uncertainty would imply narrower bias priors, which in turn lead to slower learning rates since new incoming observations are weighted down. A similar pattern was found for multisensory integration. The aVE effectively diminished when feedback was consistent with audition and increased almost to the size of the audio-visual disparity, when feedback was consistent with vision. This finding opens an alternative interpretation for the results of Study 3: feedback consistent with audition might have altered the binding tendency, i.e., the prior probability of a common cause, in a top-down manner. More specifically, when feedback was consistent with the auditory stimulus, it provided two pieces of information. On the one hand, it indicated where the auditory stimulus had been. On the other hand, when visual information was reliable, reward feedback also provided information about the true size of the audio-visual discrepancy. Hence, it allowed participants to learn that auditory and visual stimuli are spatially dissociable. In fact, decreasing binding tendency and only relying on the auditory stimulus would have maximized reward in this study. A recent study showed that attending to visual and tactile stimuli throughout bimodal stimulation increased integration as well as recalibration, whereas attending to only one sensory modality reduced integration and recalibration (Badde et al., 2020). Importantly, these effects were mediated by changes of the binding tendency.

In conclusion, reward feedback altered integration and recalibration in a top-down driven manner. While the results for the aVE are well in line with a change in the prior probability of the common cause, the results for the aCVAE are also in line with a change of the bias prior. An increased certainty about sensory biases should be directly linked to less volatility, as argued beforehand. Thus, a more direct test of whether bias priors are malleable or not, might directly manipulate the volatility of the bias in auditory spatial perception (see Burge et al., 2008; Piray & Daw, 2021 for potential paradigms). Yet, both mechanisms, i.e. down-Weighting the prior probability of a common cause as well as narrowing the bias prior, are well in line with the general hypothesis that the maintenance of accurate sensory modalities is the primary objective of multisensory recalibration (Block & Bastian, 2011; Di Luca et al., 2009; Zaidel et al., 2013), since both abolish audio-visual recalibration when the direction is not compatible with external feedback.

General Principles of Multisensory Recalibration

Sensitivity to Sensory Reliabilities

Similar to Hong et al. (2021), we found that reliability affects recalibration via its influence on multisensory integration. Furthermore, reliability affects the learning rates in recalibration. At first glance, these results might seem to contradict the findings for visuo-vestibular self-motion perception (Zaidel et al., 2011), where no effects of visual reliability were found. This may further raise the question of whether the proposed principles for audio-visual integration as well as immediate and cumulative recalibration, generalize to other stimulus dimensions and combinations of senses. However, even under the assumption that the model proposed here generalizes to other multisensory combinations, there are several reasons why the effects of reliability might vary across studies. First, if recalibration is CI-based, as the results of Studies 2, 3 as well as Hong et al. (2021) suggest, the effect of visual reliability should not be monotonic but bell shaped. With decreasing visual reliability, aftereffects first increase because the $p(C = 1|y_k)$ increases. However, at the same time the size of the errorterm and the learning rate decrease. Initially the former effect dominates. Intuitively this is plausible since the $p(C = 1|y_k)$ is always zero for perfectly reliable stimuli. With decreasing reliability, the $p(C = 1|y_k)$ initially rapidly increases since larger uncertainties make audio-visual discrepancies more likely. At a certain point, the decrease in the errorterm dominates the increase in $p(C = 1|y_k)$, and aftereffects start to decrease. It follows that stimulus conditions and study design must be carefully chosen to be able to detect reliability dependence. Since the effect of reliability should be bell shaped, it is generally a good practice to test 3 levels of reliability as done in Hong et al. (2021), especially if the study intends to show that there is no effect of reliability. Moreover, the reliability levels must be chosen so that they induce large differences in the multisensory percepts, since the emerging effects on the aftereffects will only be a fraction from the effects on multisensory integration (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009).

Finally, the buildup of the aftereffect throughout recalibration must be tracked (see Bosen et al., 2018; Watson et al., 2019 for potential paradigms) to be able to detect differences in learning rates between reliability conditions. This is especially important given that aftereffects usually reach a ceiling within a session (Bruns & Röder, 2019; Frissen et al., 2012), which means that differences in the learning rate in general cannot solely be assessed based on localization shifts in post-tests.

These model-based methodological restrictions can serve as a guide to explain seemingly opposing results regarding the role of sensory reliability (Burge et al., 2010; but Rohlf et al., 2021; Zaidel et al., 2011) in cumulative recalibration. Although the lack of reliability-dependence in visuo-vestibular recalibration might be in line with CI-based recalibration, the methodological considerations above clearly imply that studies can be designed where reliability dependence should occur. Moreover, several other forms of multisensory recalibration, for instance audio-tactile (Bruns, Spence, et al., 2011) or visuo-tactile (Samad & Shams, 2018) recalibration, have not been investigated with regard to reliability dependence. Hence, in order to accept or reject reliability dependence as a general principle of multisensory recalibration, more research accounting for the above-mentioned methodological concerns is necessary.

Prior Beliefs about Accuracy

Throughout the present thesis, we argued that recalibration should be fine-tuned to the sources of inaccuracies in the sensory systems, and that these sources define how accurate a sense is over time. In summary, the results of this thesis strongly argue in favor of this hypothesis. We did not observe visual cumulative or immediate visual aftereffects, either when vision was massively blurred (Study 1) or when explicit reward consistent with audition was given (Study 3). This pattern could be best explained by assuming bias priors differing in their variance for vision and audition, similar to the modeling results of Hong et al. (2021). The generally narrower priors for visual biases and wider priors for auditory biases might reflect the differences in visual and auditory accuracy. Sensory modality specific recalibration rates were also found for visual and vestibular self-motion perception (Zaidel et al., 2011) as well as visuo-tactile recalibration (Badde et al., 2020). Hence, the perceptual system's belief about its own sensory accuracies might be a major driver of multisensory spatial recalibration.

Yet, it is still unclear how these prior beliefs are formed, and how functional and structural constraints limit the malleability of these beliefs. The general assumption in the Bayesian framework (Berniker et al., 2010; Knill, 2007a; Sato & Körding, 2014) is that priors are shaped via experience. As argued before, it should then be possible to directly manipulate the perceived volatility of the bias in spatial perception (Burge et al., 2008; Piray & Daw, 2021), and thereby alter the magnitude of recalibration.

Priors: Experience-Based Knowledge or System Constraints?

It might be that the underlying differences in neural organization provide hardwired limits to what can be learned or not. For instance, no visual aftereffects were observed in audio-visual spatial recalibration of healthy adults (Bruns et al., 2022; Study 1, Chapter III; Study 3, Chapter IV) whereas visual aftereffects are observed in visuo-vestibular recalibration (Zaidel et al., 2011) and audio-visual temporal perception (Di Luca et al., 2009). The lack of purely visuo-spatial aftereffects highlights the semantic double role of priors in the Bayesian framework, because this behavior would be accounted for by an infinitively narrow prior. Hereby, the prior does however reflect structural and functional constraints in contrast to perfect prior knowledge about the visual sensory accuracy. Prism adaptation might serve as another example where prior knowledge about sensory accuracies is at least partially ignored. In several studies, participants were aware of the effect of the prism glasses, indicated by strategic behavioral adaptations (Redding & Wallace, 2002). Prior knowledge should then point towards highly inaccurate vision. Yet, vision is often not recalibrated at all, or to a lesser degree than auditory mappings (Canon, 1970). Moreover, there is no evidence that visual aftereffects are the result of a remapping of retinotopic spatial maps. Rather, it seems that the relative estimated position of the eye to the head is recalibrated (Crawshaw & Craske, 1974; Redding & Wallace, 1997, 2002).

Nevertheless, considering structural and functional constraints across different multisensory combinations might allow for extraction of general principles. For instance, auditory spatial perception must be inferred from binaural cues, and is encoded in a few, spatially broadly tuned channels. By contrast, for tactile spatial perception, spatiotopically somatosensory information must be combined with proprioceptive and visual cues (Azañón & Soto-Faraco, 2008). It would be a strong argument for general principles of the interplay of multisensory integration and recalibration if, despite these different neurophysiological underpinnings, visuo-tactile and audio-tactile recalibration follow the principles of Causal Inference and Kalman Filtering. The modelling framework from Chapter II provides a proper tool to test for and investigate these principles. Moreover, comparative studies would allow us to test whether bias priors generalize, i.e., whether the perceptual system uses the same auditory bias prior for audio-visual and audio-tactile recalibration. In fact, studies investigating visuo-tactile, audio-tactile and audio-visual spatial recalibration within individuals might allow for disentangling connectivity constraints on recalibration from prior uncertainty-based effects.

Another approach to dissociate the effects of structural and functional constraints from those of rapid experience-based learning on priors is to investigate typical and atypical developmental trajectories. Recent studies with individuals, who had been born blind due to dense bilateral cataracts and who regained sight later in life (Bruns et al., 2022; Senna et al., 2022) showed that the ability to integrate audio-visual spatial information, as well as to recalibrate auditory space, was unimpaired (Bruns et al., 2022). However, the same individuals also showed auditory recalibration of visual space. This likely implies that atypical input due to early visual deprivation led to additional functional connections, allowing for recalibration of visual space.

A question for future studies could be whether the emergence of these potential additional connections is a byproduct of overarching reorganization or can be interpreted as a form of uncertainty-based learning. Hereby, the lack of early visual input might induce a belief of higher uncertainty with respect to visual accuracy. This, however, does not necessarily mean that the accuracy is truly degraded. Accordingly, one would then predict that with increasing visual experience after surgery, auditory recalibration of visual space diminishes. Alternatively, these atypically acquired recalibration capabilities might be preserved over time (Keuroghlian & Knudsen, 2007). A comparative approach between typical and atypical development would then allow us to address the question of whether recalibration based on atypically acquired (or preserved) connectivity follows the same principles as typically present recalibration. If typical bias priors in multisensory recalibration do contain an experience-based component, does the same experience-based prior adjustment based on multisensory percepts and Kalman Filtering occur for visual bias priors in cataract-reversal individuals? If so, this might imply general computational principles for the interplay of integration and recalibration that might emerge whenever required connectivity is present. Similar domain-general learning mechanisms have, for instance, been proposed for visual statistical learning (Kirkham et al., 2002).

Multiple Timescales of Learning

An increasing body of literature provides evidence for dissociated learning mechanisms in audio-visual spatial perception, operating at different timescales (Bosen et al., 2018; Bruns & Röder, 2015; Watson et al., 2019) and emerging with distinct developmental trajectories (Rohlf et al., 2020, 2021). Study 1 and 2 contributed to the evidence by demonstrating that the aIVAE is likely embedded in multisensory integration, whereas the aCVAE is based on an interdependent but distinct process. Based on the literature that demonstrates a dissociation between cumulative and immediate learning across several combinations of senses in a multitude of behavioral tasks, it is reasonable to assume that the emergence of multiple learning mechanisms across different timescales is an overarching principle of multisensory perception, and plasticity in general. Cumulative and immediate learning has been found in audio-visual temporal perception (Van Der Burg et al., 2015), visual-vestibular heading perception (Shalom-Sperber et al., 2022; Zaidel et al., 2011), sensorimotor adaptation (Inoue et al., 2015) and various forms of perceptual adaptation (Bao & Engel, 2012; Dhruv et al., 2011; Mesik et al., 2013).

This generality nurtures speculations of whether there are common underlying neural principles leading to similar outcomes, in domains which are quite diverse apart from their temporal trajectories. Along the potential timescale of learning, what is referred to as cumulative and instantaneous recalibration throughout this thesis are certainly forms of rapid learning (Lewald, 2002; Recanzone, 1998) compared to long-term exposure to displaced vision over several days or weeks (Bergan et al., 2005; Linkenhoker & Knudsen, 2002); and might resemble the outcome of an ontogenetic learning process (Rohlf et al., 2020). Röder et al. (2021) provided a hypothesis of typical and atypical emergence of plasticity that fits the proposed dissociation of priors in terms of necessary connectivity on the one hand, and short-term experience-based plasticity on the other hand. They argue that strong synapses in neurons, though fewer in number than weak synapses, connect neurons with similar response properties and exert a dominant influence (Cossell et al., 2015). These strong synapses are believed to form the underlying structure or *scaffold* (Röder et al., 2021) that ensures representational stability and memory (Rose et al., 2016). During development, this specific subset of neural circuitry may be formed and stabilized. In adulthood, neural circuits are thought to adjust in response to unexpected inputs, predominantly in a top-down drive manner (Kral et al., 2005). If typical input is missing throughout sensitive periods, the organization of the scaffold might emerge in an atypical manner (Röder et al., 2021). Rose et al. (2016) proposed that adult plasticity might alter weak synaptic connections rather than the scaffold. Properties of the aIVAE and aCVAE are well in line with the involvement of these weak synapses. The aIVAE quickly dissipates over time (Bosen et al., 2018; Watson et al., 2019), and participants' response behavior shifts back to the pre-learned state, indicating that indeed the initial state of the system must have been preserved. Moreover, the aCVAE can build up over several consecutive days, although participants leave the lab in between sessions, and likely readjust their hearing in a normal hearing environment (Bruns & Röder, 2019). This points towards a dissociation of a relatively stable representation, that is quickly reestablished under typical input, and additionally a newly formed representation that might be specific to a particular hearing context (Bruns & Röder, 2019). Further, in line with a preserved stable representation of auditory space, several studies showed that altered spatial auditory cues (Carlile et al., 2014;

Carlile & Blackman, 2014; Hofman et al., 1998; Mendonça et al., 2013) can lead to recalibration that does not produce aftereffects when normal hearing conditions are restored.

One neurophysiological reason why rapid recalibration, i.e., aIVAE and aCVAE, might emerge later than integration could be that the *scaffold* has to be set up first. The *scaffold* might provide the neural circuitry necessary for proper mappings between sensory modalities. These mappings, in turn, enable the system to calculate errorterms between sensory modalities, and propagate them to initiate adultlike recalibration (Rohlf et al., 2020). Moreover, simultaneous emergence of rapid recalibration and integration could interfere with the slow stabilization of the scaffold. For instance, it has been proposed that unsupervised sensorimotor learning contributes to the formation of auditory space (Aytekin et al., 2008; Rauschecker, 1995). If multiple learning mechanisms varying in their learning rates and processing stages evolve simultaneously, biases in early spatial representations might be compensated for by rapid recalibration of later stages, eliminating potential error signals from the interaction with the environment.

A not-yet stabilized scaffold does relate to the computational framework provided here as an additional layer of uncertainty. The models described in Chapter II and in the literature in general (see Table 1 for an overview) assume perfect computations with accurate prior knowledge but noisy inputs, whereas more realistic models should rather assume noisy computations and biased prior knowledge. It might be hypothesized that stabilization during sensitive periods might go hand in hand with less noisy computations and more accurate priors.

This hypothesis perfectly agrees with a theoretical neural modelling approach of the VE and the aCVAE (Cuppini et al., 2017). Throughout the course of learning multisensory integration, the receptive fields of neurons become progressively narrower, reflecting the spatial reliability of external stimuli. Moreover, the spatial density of receptive fields adjusts to reflect unisensory priors. Finally, crossmodal synapses between visual and auditory unisensory layers adapt to represent a prior of a common cause (Ursino et al., 2019). Thereby, the network supports an increasingly accurate and precise architecture for multisensory integration. Cuppini et al. (2017) assumed that rapid recalibration, i.e. the aIVAE and aCVAE, might be realized via the alteration of lateral synapses in the unisensory auditory layer. Adopting the concept of the formation of *scaffold* in a phase of increased plasticity (Röder et al., 2021) to this neural network model would imply that only a subpopulation of weak lateral synapses would remain highly plastic after the *scaffold* is set, whereas most of the strong synaptic connections would decrease in their plasticity, which is in fact typically observed in the auditory system (Kral, 2013). It could further be hypothesized that aIVAE and aCVAE only affect weak synaptic connections. The relative contributions of weak, plastic synapses and strong, less plastic synapses might then explain the commonly found incomplete recalibration (Bertelson et al., 2006; Frissen et al., 2012; Kopco et al., 2009) since the scaffold of strong synapses would be relatively unaffected by aIVAE and aCVAE. Hence the stable representation of the scaffold might implement the hypothesized hardwired cap for the aIVAE and aCVAE. Moreover, this hypothesis would explain how the initial state of auditory spatial representations is stored, which seems to be the case, since auditory shifts induced by the aIVAE, for instance, quickly dissipate. Several studies show instant recovery when normal hearing conditions are restored after recalibration (Carlile et al., 2014; Carlile & Blackman, 2014; Hofman et al., 1998; Mendonça et al., 2013).

In line with initially less precise and accurate computations is the finding that young children have a reduced prior probability of a common cause in audio-visual spatial perception. Although Causal Inference describes their performance best, model fits were worse for young children compared to older children and adults (Rohlf et al., 2020). Given that the system might be uncertain about the computations involved for Causal Inference, it might prefer to assume distinct causes, leading to more robust predictions. A falsely assumed common cause certainly leads to more fundamental errors when the two sources split at some point, as compared to the opposite assumption. By contrast, sensory dominance as observed in visual-haptic integration in children (Burr & Gori, 2012) might rather emerge when there are pronounced differences in uncertainties with respect to the mappings between sensory cues and perceptual dimensions. Burr & Gori (2012) argue that with respect to object size, touch directly encodes object size, whereas vision must infer size by a complex calculation based on retinal size and object distance, making it reasonable for the system to assume that touch is more accurate although less precise. In such a scenario, when an object is sensed by vision and touch in a temporal and spatially highly coherent manner, a common cause is very likely, and the cue conflict with respect to object size must be resolved. Due to the high uncertainty about visual accuracy, touch might then dominate vision. Such a dependency of multisensory integration on not only relative reliabilities, but also relative accuracies, is exactly what the single process model of integration and recalibration would predict. Interestingly, a recent study showed (Nava et al., 2020) that visual-haptic integration in children can become optimal for children aged 4–5 years after gamified training. The authors argued that one reason could have been that the weights for each sensory modality were calculated more accurately after training. That could mean that the involved computations became less noisy to a point were optimal calculations became beneficial, in comparison to simple heuristics.

A summarizing hypothesis would be that differences in the developmental trajectory of multisensory integration and recalibration across different sensory combinations arise due to differences in sensory uncertainties, accuracies, and, moreover, different levels of uncertainty in the computations themselves. In sum, these factors might make different suboptimal heuristics (i.e., sensory dominance, or a bias to segregation) more suitable for different sensory combinations, until uncertainties are sufficiently reduced to implement closer-to-optimal inference schemes.

Conclusion

The perceptual system does not only learn about the statistics of relevant features of the external world like space and time, but also about the dynamics of these features, providing distinct mechanisms to account for distinct sources of uncertainty. Whereas audio-visual integration implements Causal Inference, recalibration combines prior knowledge about sensory inaccuracies, and sensory evidence about inaccuracies. Recalibration mechanisms are fine-tuned to multiple levels of volatility and can incorporate top-down information about sensory accuracy. When joint computations are beneficial, as for the aIVAE and VE, computations are performed jointly. With decreasing computational entanglement of perceptual inference, processes become more independent, as for instance VE and aCVAE.

A further going hypothesis raised by the present thesis would be that whenever the necessary neural circuitry for multisensory interactions is available, multisensory integration and recalibration emerge based on similar computational principles across sensory modalities. However, the resulting developmental trajectories as well as adult mechanisms might greatly depend on the precision and accuracy of the sensory input and on the level of uncertainty in the actual computations that a particular sensory system can provide. Importantly, by extending the modelling framework proposed in this thesis to account for noisy and uncertain computations, for example by assuming adaptive Kalman Filters (Akhlaghi et al., 2017; Rao, 1999), where not only the states but also the system underlies uncertainty, this hypothesis can be tested empirically, potentially providing a unifying framework for seemingly contradictory results in the literature

References

Acerbi, L., Dokka, K., Angelaki, D. E., & Ma, W. J. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Computational Biology*, *14*(7), e1006110. https://doi.org/10.1371/journal.pcbi.1006110

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058. https://doi.org/10.1038/nn1312

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2010). Efficient visual recalibration from either visual or haptic feedback: The importance of being wrong. *Journal of Neuroscience*, *30*(44), 14745–14749. https://doi.org/10.1523/JNEUROSCI.2749-10.2010

Akhlaghi, S., Zhou, N., & Huang, Z. (2017). Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation. *2017 IEEE Power & Energy Society General Meeting*, 1–5. https://doi.org/10.1109/PESGM.2017.8273755

Alais, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/S0960-9822(04)00043-0

Alais, D., & Burr, D. (2004). Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*. https://doi.org/10.1016/S0960-9822(04)00043-0

Aller, M., Mihalik, A., & Noppeney, U. (2022). Audiovisual adaptation is expressed in spatial and decisional codes. *Nature Communications*, *13*(1), 1–17. https://doi.org/10.1038/s41467-022-31549-0

Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biology*, *17*(4), e3000210. https://doi.org/10.1371/journal.pbio.3000210

Andersen, R. A. (1997). Neural mechanisms of visual motion perception in primates. *Neuron*, *18*(6), 865–872. https://doi.org/10.1016/S0896-6273(00)80326-8

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal Filtering*. Prentice Hall.

Arnold, D. H., Petrie, K., Murray, C., & Johnston, A. (2019). Suboptimal human multisensory cue combination. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-018-37888-7

Arnott, S. R., & Alain, C. (2011). The auditory dorsal pathway: Orienting vision. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2162–2173. https://doi.org/10.1016/j.neubiorev.2011.04.005

Aytekin, M., Moss, C. F., & Simon, J. Z. (2008). A sensorimotor approach to sound localization. *Neural Computation*, *20*(3), 603–635. https://doi.org/10.1162/neco.2007.12-05-094

Azañón, E., & Soto-Faraco, S. (2008). Changing Reference Frames during the Encoding of Tactile Events. *Current Biology*, *18*(14), 1044–1049. https://doi.org/10.1016/j.cub.2008.06.045

Badde, S., Navarro, K. T., & Landy, M. S. (2020). Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. *Cognition*, *197*(December 2019), 104170. https://doi.org/10.1016/j.cognition.2019.104170

Bao, M., & Engel, S. A. (2012). Distinct mechanism for long-term contrast adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(15), 5898–5903. https://doi.org/10.1073/pnas.1113503109

Barraza, J. F., & Grzywacz, N. M. (2008). Speed adaptation as Kalman filtering. *Vision Research*, *48*(23–24), 2485–2491. https://doi.org/10.1016/j.visres.2008.08.011

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391–1397. https://doi.org/10.1364/JOSAA.20.001391

Bauer, R. W., Matuzsa, J. L., Blackmer, R. F., & Glucksberg, S. (1966). Noise Localization after Unilateral Attenuation. *The Journal of the Acoustical Society of America*, *40*(2), 441–444. https://doi.org/10.1121/1.1910093

Beaumont, M. A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and Its Application*, *6*, 379–403. https://doi.org/10.1146/annurev-statistics-030718-105212

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., & Robert, C. P. (2008). Adaptive approximate Bayesian computation. *Biometrika*, *96*(4), 983–990. https://doi.org/10.1093/biomet/asp052

Beaumont, M. A., Cornuet, J. M., Marin, J. M., & Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, *96*(4), 983–990. https://doi.org/10.1093/biomet/asp052

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

Beierholm, U., Quartz, S. R., & Shams, L. (2010). The ventriloquist illusion as an optimal percept. *Journal of Vision*, *5*(8), 647–647. https://doi.org/10.1167/5.8.647

Beierholm, U., Rohe, T., Ferrari, A., Stegle, O., & Noppeney, U. (2020). Using the past to estimate sensory uncertainty. *ELife*, *9*, 1–22. https://doi.org/10.7554/ELIFE.54172

Bergan, J. F., Ro, P., Ro, D., & Knudsen, E. I. (2005). Hunting increases adaptive auditory map plasticity in adult barn owls. *Journal of Neuroscience*, *25*(42), 9816–9820. https://doi.org/10.1523/JNEUROSCI.2533-05.2005

Berger, C. C., & Ehrsson, H. H. (2016). Auditory motion elicits a visual motion aftereffect. *Frontiers in Neuroscience*, *10*(DEC), 1–6. https://doi.org/10.3389/fnins.2016.00559

Berger, T. O., & Pericchi, L. R. (2004). Training samples in objective Bayesian model selection. *Annals of Statistics*, *32*(3), 841–869. https://doi.org/10.1214/009053604000000229

Berniker, M., Voss, M., & Körding, K. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE*, *5*(9), 1–9. https://doi.org/10.1371/journal.pone.0012686

Bernton, E., Jacob, P. E., Gerber, M., & Robert, C. P. (2019). Approximate Bayesian Computation with the Wasserstein Distance. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *81*(2), 235–269. https://doi.org/10.1111/rssb.12312

Bertelson, P. (1999). Chapter 14 Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müssler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (pp. 347–362). Elsevier. https://doi.org/10.1016/S0166-4115(99)80034-X

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489. https://doi.org/10.3758/BF03208826

Bertelson, P., Frissen, I., Vroomen, J., & De Gelder, B. (2006). The aftereffects of ventriloquism: Patterns of spatial generalization. *Perception and Psychophysics*, *68*(3), 428–436. https://doi.org/10.3758/BF03193687

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*(12), 1634–1642. https://doi.org/10.1016/S0028-3932(00)00067-1

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332. https://doi.org/10.3758/BF03205552

Blauert, J. (1996). *Spatial Hearing*. The MIT Press. https://doi.org/10.7551/mitpress/6391.001.0001

Block, H. J., & Bastian, A. J. (2011). Sensory weighting and realignment: independent compensatory processes. *Journal of Neurophysiology*, *106*(1), 59–70. https://doi.org/10.1152/jn.00641.2010

Bolognini, N., Leo, F., Passamonti, C., Stein, B. E., & Làdavas, E. (2007). Multisensory-mediated auditory localization. *Perception*, *36*(10), 1477–1485. https://doi.org/10.1068/p5846

Bonath, B., Noesselt, T., Krauel, K., Tyll, S., Tempelmann, C., & Hillyard, S. A. (2014). Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *NeuroImage*, *98*, 425–434. https://doi.org/10.1016/j.neuroimage.2014.04.077

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, *17*(19), 1697–1703. https://doi.org/10.1016/j.cub.2007.08.050

Bond, K. M., & Taylor, J. A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *Journal of Neurophysiology*, *113*(10), 3836–3849. https://doi.org/10.1152/jn.00009.2015

Bosen, A. K., Fleming, J. T., Allen, P. D., O‘Neill, W. E., & Paige, G. D. (2017). Accumulation and decay of visual capture and the ventriloquism aftereffect caused by brief audio-visual disparities. *Experimental Brain Research*, *235*(2), 585–595. https://doi.org/10.1007/s00221-016-4820-4

Bosen, A. K., Fleming, J. T., Allen, P. D., O’Neill, W. E., & Paige, G. D. (2018). Multiple time scales of the ventriloquism aftereffect. *PLoS ONE*, *13*(8), e0200930. https://doi.org/10.1371/journal.pone.0200930

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brown, M. B., & Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, *69*(346), 364. https://doi.org/10.2307/2285659

Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *Journal of Neuroscience*, *16*(14), 4420–4437. https://doi.org/10.1523/jneurosci.16-14-04420.1996

Bruns, P. (2019). The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update. *Frontiers in Integrative Neuroscience*, *13*(September), 1–8. https://doi.org/10.3389/fnint.2019.00051

Bruns, P., Dinse, H. R., & Röder, B. (2020). Differential effects of the temporal and spatial distribution of audiovisual stimuli on cross-modal spatial recalibration. *European Journal of Neuroscience*, *52*(7), 3763–3775. https://doi.org/10.1111/ejn.14779

Bruns, P., Li, L., Guerreiro, M. J. S., Shareef, I., Rajendran, S. S., Pitchaimuthu, K., Kekunnaya, R., & Röder, B. (2022). Audiovisual spatial recalibration but not integration is shaped by early sensory experience. *IScience*, *25*(6). https://doi.org/10.1016/j.isci.2022.104439

Bruns, P., Liebnau, R., & Röder, B. (2011). Cross-Modal Training Induces Changes in Spatial Representations Early in the Auditory Processing Pathway. *Psychological Science*, *22*(9), 1120–1126. https://doi.org/10.1177/0956797611416254

Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception, & Psychophysics*, *76*(6), 1815–1827. https://doi.org/10.3758/s13414-014-0699-y

Bruns, P., & Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Scientific Reports*, *5*, 12739. https://doi.org/10.1038/srep12739

Bruns, P., & Röder, B. (2019). Repeated but not incremental training enhances cross-modal recalibration. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(4), 435–440. https://doi.org/10.1037/xhp0000642

Bruns, P., Spence, C., & Röder, B. (2011). Tactile recalibration of auditory spatial representations. *Experimental Brain Research*, *209*(3), 333–344. https://doi.org/10.1007/s00221-011-2543-0

Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, *8*(4), 20. https://doi.org/10.1167/8.4.20

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual–Haptic Adaptation Is Determined by Relative Reliability. *The Journal of Neuroscience*, *30*(22), 7714–7721. https://doi.org/10.1523/JNEUROSCI.6427-09.2010

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

Burr, D., & Gori, M. (2012). Multisensory Integration Develops Late in Humans. In *The Neural Bases of Multisensory Processes*. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/pubmed/22593886

Butler, R. A. (1986). The bandwidth effect on monaural and binaural localization. *Hearing Research*, *21*(1), 67–73. https://doi.org/10.1016/0378-5955(86)90047-X

Callan, A., Callan, D., & Ando, H. (2015). An fMRI study of the ventriloquism effect. *Cerebral Cortex*, *25*(11), 4248–4258. https://doi.org/10.1093/cercor/bhu306

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1–3), 191–205. https://doi.org/10.1016/J.JPHYSPARIS.2004.03.018

Campbell, H., & Gustafson, P. (2022). Bayes factors and posterior estimation: Two sides of the very same coin. *ArXiv Preprint*, 1–21. http://arxiv.org/abs/2204.06054

Canon, L. K. (1970). *Intermodality Inconsistency of Input and Directed Attention As Determinants of the Nature of Adaptation*. *84*(1), 141–147.

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal Inference in the Multisensory Brain. *Neuron*, *102*(5), 1076-1087.e8. https://doi.org/10.1016/J.NEURON.2019.03.043

Carlile, S., Balachandar, K., & Kelly, H. (2014). Accommodating to new ears: The effects of sensory and sensory-motor feedback. *The Journal of the Acoustical Society of America*, *135*(4), 2002–2011. https://doi.org/10.1121/1.4868369

Carlile, S., & Blackman, T. (2014). Relearning Auditory Spectral Cues for Locations Inside and Outside the Visual Field. *Journal of the Association for Research in Otolaryngology*, *15*(2), 249–263. https://doi.org/10.1007/s10162-013-0429-5

Carlile, S., Delaney, S., & Corderoy, A. (1999). The localisation of spectrally restricted sounds by human listeners. *Hearing Research*, *128*(1–2), 175–189. https://doi.org/10.1016/S0378-5955(98)00205-6

Carlile, S., Fox, A., Orchard-Mills, E., Leung, J., & Alais, D. (2016). Six Degrees of Auditory Spatial Separation. *JARO - Journal of the Association for Research in Otolaryngology*, *17*(3), 209–221. https://doi.org/10.1007/s10162-016-0560-1

Carlile, S., Hyams, S., & Delaney, S. (2001). Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *The Journal of the Acoustical Society of America*, *110*(1), 416–424. https://doi.org/10.1121/1.1375843

Carlile, S., Martin, R., & McAnally, K. (2005). Spectral Information in Sound Localization. In *International Review of Neurobiology* (Vol. 70, pp. 399–434). https://doi.org/10.1016/S0074-7742(05)70012-X

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

Chen, Y.-C., & Spence, C. (2017). Assessing the Role of the 'Unity Assumption' on Multisensory Integration: A Review. *Frontiers in Psychology*, *8*(MAR), 445. https://doi.org/10.3389/fpsyg.2017.00445

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, *518*(7539), 399–403. https://doi.org/10.1038/nature14182

Crawshaw, M., & Craske, B. (1974). No retinal component in prism adaptation. *Acta Psychologica*, *38*(6), 421–423. https://doi.org/10.1016/0001-6918(74)90001-8

Cuppini, C., Shams, L., Magosso, E., & Ursino, M. (2017). A biologically inspired neurocomputational model for audiovisual integration and causal inference. *European Journal of Neuroscience*, *46*(9), 2481–2498. https://doi.org/10.1111/ejn.13725

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space. *Neuron*, *66*(6), 937–948. https://doi.org/10.1016/j.neuron.2010.05.018

de Winkel, K. N., Katliar, M., & Bülthoff, H. H. (2015). Forced Fusion in Multisensory Heading Estimation. *PLoS ONE*, *10*(5), e0127104. https://doi.org/10.1371/journal.pone.0127104

de Winkel, K. N., Katliar, M., & Bülthoff, H. H. (2017). Causal Inference in Multisensory Heading Estimation. *PLoS ONE*, *12*(1), e0169676. https://doi.org/10.1371/journal.pone.0169676

Dean, I., Harper, N. S., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, *8*(12), 1684–1689. https://doi.org/10.1038/nn1541

DeBello, W. M., Feldman, D. E., & Knudsen, E. I. (2001). Adaptive axonal remodeling in the midbrain auditory space map. *Journal of Neuroscience*, *21*(9), 3161–3174. https://doi.org/10.1523/jneurosci.21-09-03161.2001

Delong, P., Aller, M., Giani, A. S., Rohe, T., Conrad, V., Watanabe, M., & Noppeney, U. (2018). Invisible Flashes Alter Perceived Sound Location. *Scientific Reports*, *8*(1), 12376. https://doi.org/10.1038/s41598-018-30773-3

Denève, S., Duhamel, J. R., & Pouget, A. (2007). Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of Kalman filters. *Journal of Neuroscience*, *27*(21), 5744–5756. https://doi.org/10.1523/JNEUROSCI.3985-06.2007

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*(1), 241–265. https://doi.org/10.1113/jphysiol.1984.sp015499

Dhruv, N. T., Tailby, C., Sokol, S. H., & Lennie, P. (2011). Multiple adaptable mechanisms early in the primate visual pathway. *Journal of Neuroscience*, *31*(42), 15016–15025. https://doi.org/10.1523/JNEUROSCI.0890-11.2011

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal* *of Vision*, *9*(12), 7. https://doi.org/10.1167/9.12.7

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/BF03195006

Dingle, R. N., Hall, S. E., & Phillips, D. P. (2012). The three-channel model of sound localization mechanisms: Interaural level differences. *The Journal of the Acoustical Society of America*, *131*(5), 4023–4029. https://doi.org/10.1121/1.3701877

Dokka, K., Park, H., Jansen, M., DeAngelis, G. C., & Angelaki, D. E. (2019). Causal inference accounts for heading perception in the presence of object motion. *Proceedings of the National Academy of Sciences*, *116*(18), 9060–9065. https://doi.org/10.1073/pnas.1820373116

Drécourt, J. P., Madsen, H., & Rosbjerg, D. (2006). Bias aware Kalman filters: Comparison and improvements. *Advances in Water Resources*, *29*(5), 707–718. https://doi.org/10.1016/j.advwatres.2005.07.006

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1319–1331. https://doi.org/10.1098/rstb.1998.0286

Ege, R., van Opstal, A. J., & van Wanrooij, M. M. (2019). Frequency Transfer of the Ventriloquism Aftereffect. *PDF Hosted at the Radboud Repository of the Radboud University Nijmegen*, 125. https://doi.org/https://doi.org/10.1101/2021.12.22.473801

Epstein, W. (1975). Recalibration by pairing: a process of perceptual learning. *Perception*, *4*(1), 59–72. https://doi.org/10.1068/p040059

Eramudugolla, R., Kamke, M. R., Soto-Faraco, S., & Mattingley, J. B. (2011). Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition*, *118*(1), 62–74. https://doi.org/10.1016/J.COGNITION.2010.09.009

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Luca, M. Di. (2011). Multisensory perception: From integration to remapping. In J. Trommershäuser, K. P. Körding, & M. S. Landy (Eds.), *Sensory Cue Integration* (pp. 224–250). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.003.0012

Fang, Y., Yu, Z., Liu, J. K., & Chen, F. (2019). A unified neural circuit of causal inference and multisensory integration. *Neurocomputing*, *358*, 355–368. https://doi.org/10.1016/j.neucom.2019.05.067

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

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–154. https://doi.org/10.1038/nn.2983

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, *29*(49), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009

Fortenbaugh, F. C., Sanghvi, S., Silver, M. A., & Robertson, L. C. (2012). Exploring the edges of visual space: The influence of visual boundaries on peripheral localization. *Journal of Vision*, *12*(2), 19–19. https://doi.org/10.1167/12.2.19

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage.

Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*(3), 425–442. https://doi.org/10.1016/j.neuron.2011.10.006

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. https://doi.org/10.1007/s00221-002-1262-y

Frens, M. A., & Van Opstal, A. J. (1995). A quantitative study of auditory-evoked saccadic eye movements in two dimensions. *Experimental Brain Research*, *107*(1), 103–117. https://doi.org/10.1007/BF00228022

Friedland, B. (1969). Treatment of Bias in Recursive Filtering. *IEEE Transactions on Automatic Control*, *AC-14*(4), 359–367. https://doi.org/10.1109/TAC.1969.1099223

Frissen, I., Vroomen, J., & de Gelder, B. (2012). The Aftereffects of Ventriloquism: The Time Course of the Visual Recalibration of Auditory Localization. *Seeing and Perceiving*, *25*(1), 1–14. https://doi.org/10.1163/187847611X620883

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, *113*(3), 315–327. https://doi.org/10.1016/S0001-6918(03)00043-X

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: Generalization across sound-frequencies. *Acta Psychologica*, *118*(1–2), 93–100. https://doi.org/10.1016/j.actpsy.2004.10.004

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity*, *1*(1), 13–36. https://doi.org/10.1089/brain.2011.0008

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

Garcia, S. E., Jones, P. R., Rubin, G. S., & Nardini, M. (2017). Auditory localisation biases increase with sensory uncertainty. *Scientific Reports*, *7*(1), 40567. https://doi.org/10.1038/srep40567

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nature Neuroscience*, *22*(4), 514–523. https://doi.org/10.1038/s41593-019-0340-4

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., & Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. *Neuron*, *47*(4), 607–620. https://doi.org/10.1016/j.neuron.2005.07.016

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, *1*(7), 775. https://doi.org/10.1364/josaa.1.000775

Getzmann, S. (2004). Spatial discrimination of sound sources in the horizontal plane following an adapter sound. *Hearing Research*, *191*(1–2), 14–20. https://doi.org/10.1016/j.heares.2003.12.020

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In *Advances in Psychology* (Vol. 119, Issue C, pp. 117–147). Elsevier. https://doi.org/10.1016/S0166-4115(97)80006-4

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The Neural Basis of Perceptual Learning. *Neuron*, *31*(5), 681–697. https://doi.org/10.1016/S0896-6273(01)00424-X

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience*, *6*, 77. https://doi.org/10.3389/fnint.2012.00077

Grantham, D. W., Hornsby, B. W. Y., & Erpenbeck, E. A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *The Journal of the Acoustical Society of America*, *114*(2), 1009–1022. https://doi.org/10.1121/1.1590970

Groh, J. M., Kelly, K. A., & Underhill, A. M. (2003). A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus. *Journal of Cognitive Neuroscience*, *15*(8), 1217–1231.

https://doi.org/10.1162/089892903322598166

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of Sound Localization in Mammals. *Physiological Reviews*, *90*(3), 983–1012. https://doi.org/10.1152/physrev.00026.2009

Grubbs, F. E. (1973). Errors of Measurement, Precision, Accuracy and the Statistical Comparison of Measuring Instruments. *Technometrics*, *15*(1), 53–66. https://doi.org/10.1080/00401706.1973.10489010

Grzywacz, N. M., & De Juan, J. (2003). Sensory adaptation as Kalman filtering: Theory and illustration with contrast adaptation. *Network: Computation in Neural Systems*, *14*(3), 465–482. https://doi.org/10.1088/0954-898X_14_3_305

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, *11*(10), 1201–1210. https://doi.org/10.1038/nn.2191

Habets, B., Bruns, P., & Röder, B. (2017). Experience with crossmodal statistics reduces the sensitivity for audio-visual temporal asynchrony. *Scientific Reports*. https://doi.org/10.1038/s41598-017-01252-y

Han, S., Chen, Y.-C., Maurer, D., Shore, D. I., Lewis, T. L., Stanley, B. M., & Alais, D. (2022). The development of audio–visual temporal precision precedes its rapid recalibration. *Scientific Reports*, *12*(1), 21591. https://doi.org/10.1038/s41598-022-25392-y

Harvey, A. C. (1986). Analysis and Generalisation of a Multivariate Exponential Smoothing Model. *Management Science*, *32*(3), 374–380. https://doi.org/10.1287/mnsc.32.3.374

Hastie, T. J., & Pregibon, D. (2017). Generalized Linear Models. In *Statistical Models in S* (pp. 195–247). Routledge. https://doi.org/10.1201/9780203738535-6

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493. https://doi.org/10.1038/s41586-021-04129-3

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, *179*(4), 595–606. https://doi.org/10.1007/s00221-006-0814-y

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141. https://doi.org/10.1016/S0042-6989(97)00043-6

Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, *1*(5), 417–421. https://doi.org/10.1038/1633

Holmes, N. P., & Spence, C. (2005). Multisensory Integration: Space, Time and Superadditivity. *Current Biology*, *15*(18), R762–R764. https://doi.org/10.1016/j.cub.2005.08.058

Hong, F., Badde, S., & Landy, M. S. (2021). Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. *PLoS Computational Biology*, *17*(11). https://doi.org/10.1371/journal.pcbi.1008877

Howard, I. P., & Templeton, W. B. (1966). Human spatial orientation. In *Human spatial orientation*. John Wiley & Sons.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Inoue, M., Uchimura, M., Karibe, A., O'Shea, J., Rossetti, Y., & Kitazawa, S. (2015). Three timescales in prism adaptation. *Journal of Neurophysiology*, *113*(1), 328–338. https://doi.org/10.1152/jn.00803.2013

Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, *7*(3), e1002012. https://doi.org/10.1371/journal.pcbi.1002012

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex*, *15*(10), 1609–1620. https://doi.org/10.1093/cercor/bhi039

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. https://doi.org/10.1017/S0140525X10003134

Julian, R., Abel, R., Brouwer, A. E., Colbourn, C. J., & Dinitz, J. H. (1996). Mutually Orthogona Latin Squares (MOLS). In C. J. Colbour & J. H. Dinitz (Eds.), *The CRC Handbook of Combinatorial Designs* (pp. 111–142). CRC Press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, Transactions of the ASME*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kashino, M., & Nishida, S. (1998). Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *The Journal of the Acoustical Society of America*, *103*(6), 3597–3604. https://doi.org/10.1121/1.423064

Keating, P., & King, A. J. (2015). Sound localization in a changing world. *Current Opinion in Neurobiology*, *35*, 35–43. https://doi.org/10.1016/j.conb.2015.06.005

Keil, J., & Senkowski, D. (2018). Neural oscillations orchestrate multisensory processing. *The Neuroscientist*, *24*(6), 609–626. https://doi.org/10.1177/1073858418755352

Keuroghlian, A. S., & Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, *82*(3), 109–121. https://doi.org/10.1016/j.pneurobio.2007.03.005

King, A. J. (2009). Visual influences on auditory spatial learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 331–339. https://doi.org/10.1098/rstb.2008.0230

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, *43*(7), 831–854. https://doi.org/10.1016/S0042-6989(03)00003-8

Knill, D. C. (2007a). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7), 5. https://doi.org/10.1167/7.7.5

Knill, D. C. (2007b). Learning Bayesian priors for depth perception. *Journal of Vision*, *7*(8), 13–13. https://doi.org/10.1167/7.8.13

Knudsen, E. I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature*, *417*(6886), 322–328. https://doi.org/10.1038/417322a

Knudsen, E. I., & Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience*, *9*(9), 3306–3313. https://doi.org/10.1523/jneurosci.09-09-03306.1989

Knudsen, E. I., & Knudsen, P. F. (1990). Sensitive and critical periods for visual calibration of sound localization by barn owls. *The Journal of Neuroscience*, *10*(1), 222–232.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. https://doi.org/10.1016/j.actpsy.2010.03.010

Kopco, N., Lin, I.-F., Shinn-Cunningham, B. G., & Groh, J. M. (2009). Reference Frame of the Ventriloquism Aftereffect. *Journal of Neuroscience*, *29*(44), 13809–13814.

https://doi.org/10.1523/JNEUROSCI.2783-09.2009

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kording, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nature Neuroscience*, *10*(6), 779–786. https://doi.org/10.1038/nn1901

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. https://doi.org/10.1016/j.tics.2006.05.003

Kral, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, *247*, 117–133. https://doi.org/10.1016/j.neuroscience.2013.05.021

Kral, A., Tillein, J., Heid, S., Hartmann, R., & Klinke, R. (2005). Postnatal cortical development in congenital auditory deprivation. *Cerebral Cortex*, *15*(5), 552–562. https://doi.org/10.1093/cercor/bhh156

Kriegstein, K. Von, & Giraud, A. (2006). *Implicit Multisensory Associations Influence Voice Recognition*. *4*(10). https://doi.org/10.1371/journal.pbio.0040326

Lee, C.-C., & Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neuroscience*, *14*(1), 108–114. https://doi.org/10.1038/nn.2713

Leo, F., Bertini, C., Di Pellegrino, G., & Làdavas, E. (2008). Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Experimental Brain Research*, *186*(1), 67–77. https://doi.org/10.1007/s00221-007-1204-9

Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning & Memory*, *9*(5), 268–278. https://doi.org/10.1101/lm.51402

Lewald, J., Dörrscheidt, G. J., & Ehrenstein, W. H. (2000). Sound localization with eccentric head position. *Behavioural Brain Research*, *108*(2), 105–125. https://doi.org/10.1016/S0166-4328(99)00141-2

Lewald, J., & Ehrenstein, W. H. (2000). Visual and proprioceptive shifts in perceived egocentric direction induced by eye-position. *Vision Research*, *40*(5), 539–547. https://doi.org/10.1016/S0042-6989(99)00197-2

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory – visual integration. *Behavioural Brain Research*, *121*(1), 69–79.

Lingner, A., Pecka, M., Leibold, C., & Grothe, B. (2018). A novel concept for dynamic adjustment of auditory space. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-26690-0

Linkenhoker, B. A., & Knudsen, E. I. (2002). Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature*, *419*(6904), 293–296. https://doi.org/10.1038/nature01002

Logothetis, N. K., & Charles, E. R. (1990). The minimum motion technique applied to determine isoluminance in psychophysical experiments with monkeys. *Vision Research*, *30*(6), 829–838. https://doi.org/10.1016/0042-6989(90)90052-M

Lu, Z.-L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America A*, *18*(9), 2331. https://doi.org/10.1364/josaa.18.002331

Ma, W. J. (2018). Identifying suboptimalities with factorial model comparison. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X18001541

Ma, W. J., & Jazayeri, M. (2014). Neural Coding of Uncertainty and Probability. *Annual*

Review of Neuroscience, *37*(1), 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Machulla, T.-K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*(1), 89–97. https://doi.org/10.1007/s00221-011-2976-5

Macmillan, N. a, & Creelman, C. D. (2004). *Detection Theory*. Psychology Press. https://doi.org/10.4324/9781410611147

Maddox, R. K., Billimoria, C. P., Perrone, B. P., Shinn-Cunningham, B. G., & Sen, K. (2012). Competing Sound Sources Reveal Spatial Effects in Cortical Processing. *PLoS Biology*, *10*(5), e1001319. https://doi.org/10.1371/journal.pbio.1001319

Magosso, E., Cona, F., & Ursino, M. (2013). A Neural Network Model Can Explain Ventriloquism Aftereffect and Its Generalization across Sound Frequencies. *BioMed Research International*, *2013*, 1–17. https://doi.org/10.1155/2013/475427

Mahani, M. A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R., & Ahmadabadi, M. N. (2017). Multisensory Perception of Contradictory Information in an Environment of Varying Reliability: Evidence for Conscious Perception and Optimal Causal Inference. *Scientific Reports*, *7*(1), 3167. https://doi.org/10.1038/s41598-017-03521-2

Maier, J. K., Hehrmann, P., Harper, N. S., Klump, G. M., Pressnitzer, D., & McAlpine, D. (2012). Adaptive coding is constrained to midline locations in a spatial listening task. *Journal of Neurophysiology*, *108*(7), 1856–1868. https://doi.org/10.1152/jn.00652.2011

Maier, J. X., & Groh, J. M. (2009). Multisensory guidance of orienting behavior. *Hearing Research*, *258*(1–2), 106–112. https://doi.org/10.1016/J.HEARES.2009.05.008

Maiworm, M., & Röder, B. (2011). Suboptimal auditory dominance in audiovisual integration of temporal cues. *Tsinghua Science and Technology*, *16*(2), 121–132. https://doi.org/10.1016/S1007-0214(11)70019-0

Maiworm, Mario, Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics*, *74*(6), 1302–1311. https://doi.org/10.3758/s13414-012-0310-3

Makin, J. G., Fellows, M. R., & Sabes, P. N. (2013). Learning multisensory integration and coordinate transformation via density estimation. *PLoS Computational Biology*, *9*(4), e1003035. https://doi.org/10.1371/journal.pcbi.1003035

Marin, J. M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, *22*(6), 1167–1180. https://doi.org/10.1007/s11222-011-9288-2

McDermott, K. C., Malkoc, G., Mulligan, J. B., & Webster, M. A. (2010). Adaptation and visual salience. *Journal of Vision*, *10*(13), 17.1-17.32. https://doi.org/10.1167/10.13.17

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568–9579. https://doi.org/10.1523/JNEUROSCI.5061-14.2015

McGovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports*, *6*(1), 24673. https://doi.org/10.1038/srep24673

Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, *119*, 74–88. https://doi.org/10.1016/j.cortex.2019.03.026

Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Frontiers in Neuroscience*, *8*, 219. https://doi.org/10.3389/fnins.2014.00219

Mendonça, C., Campos, G., Dias, P., & Santos, J. A. (2013). Learning Auditory Space: Generalization and Long-Term Effects. *PLoS ONE*, *8*(10), e77900. https://doi.org/10.1371/journal.pone.0077900

Mendonça, C., Escher, A., van de Par, S., & Colonius, H. (2015). Predicting auditory space calibration from recent multisensory experience. *Experimental Brain Research*, *233*(7), 1983–1991. https://doi.org/10.1007/s00221-015-4259-z

Mesik, J., Bao, M., & Engel, S. A. (2013). Spontaneous recovery of motion and face aftereffects. *Vision Research*, *89*, 72–78. https://doi.org/10.1016/j.visres.2013.07.004

Middlebrooks, J. C., & Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *Journal of Neuroscience*, *33*(27), 10986–11001. https://doi.org/10.1523/JNEUROSCI.1065-13.2013

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annu Rev Psychol*, *42*, 135–159. m:%5CLiteratur%5Cbr%5Cmiddlebrooks 1991 -annu rev psychol- sound localization by human listeners.pdf

Mills, A. W. (1958). On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, *30*(4), 237–246. https://doi.org/10.1121/1.1909553

Młynarski, W. (2015). The Opponent Channel Population Code of Sound Location Is an Efficient Representation of Natural Binaural Sounds. *PLoS Computational Biology*, *11*(5), 1004294. https://doi.org/10.1371/journal.pcbi.1004294

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2007). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*(1), 39–52. https://doi.org/10.1007/s00221-007-1080-3

Murai, Y., & Yotsumoto, Y. (2018). Optimal multisensory integration leads to optimal time estimation. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-31468-5

Murphy, K. P. (1998). *Switching Kalman Filters*.

Nava, E., Föcker, J., & Gori, M. (2020). Children can optimally integrate multisensory information after a short action-like mini game training. *Developmental Science*, *23*(1), e12840. https://doi.org/10.1111/desc.12840

Negen, J., Chere, B., Bird, L. A., Taylor, E., Roome, H. E., Keenaghan, S., Thaler, L., & Nardini, M. (2019). Sensory cue combination in children under 10 years of age. *Cognition*, *193*, 104014. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104014

Niehorster, D. C., Li, L., & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *I-Perception*, *8*(3), 1–23. https://doi.org/10.1177/2041669517708205

Noppeney, U. (2021). *Annual Review of Neuroscience Perceptual Inference, Learning, and Attention in a Multisensory World*. https://doi.org/10.1146/annurev-neuro-100120

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*(2), 316–333. https://doi.org/10.1037/0033-2909.97.2.316

Odegaard, B., & Shams, L. (2016). The Brain's Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*, *27*(4), 583–591. https://doi.org/10.1177/0956797616628860

Odegaard, B., Wozny, D., & Shams, L. (2015a). Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLOS Computational Biology*, *11*(12), e1004649. https://doi.org/10.1371/journal.pcbi.1004649

Odegaard, B., Wozny, D., & Shams, L. (2015b). Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Computational Biology*, *11*(12), e1004649. https://doi.org/10.1371/journal.pcbi.1004649

Odegaard, B., Wozny, D., & Shams, L. (2017). A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ*, *5*, e3143. https://doi.org/10.7717/peerj.3143

Okorokova, E., Lebedev, M., Linderman, M., & Ossadtchi, A. (2015). A dynamical model improves reconstruction of handwriting from multichannel electromyographic recordings. *Frontiers in Neuroscience*, *9*(OCT). https://doi.org/10.3389/fnins.2015.00389

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Pages, D. S., & Groh, J. M. (2013). Looking at the ventriloquist: Visual outcome of eye movements calibrates sound localization. *PLoS ONE*, *8*(8), e72562. https://doi.org/10.1371/journal.pone.0072562

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108. https://doi.org/10.1073/pnas.1322705111

Park, H., & Kayser, C. (2019). Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. *ELife*, *8*. https://doi.org/10.7554/eLife.47001.001

Park, H., & Kayser, C. (2020). Robust spatial ventriloquism effect and trial-by-trial aftereffect under memory interference. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-77730-7

Park, H., & Kayser, C. (2021). The neurophysiological basis of the trial-wise and cumulative ventriloquism aftereffects. *The Journal of Neuroscience*, *41*(5), 1068–1079. https://doi.org/10.1523/JNEUROSCI.2091-20.2020

Park, H., Nannt, J., & Kayser, C. (2021). Sensory- and memory-related drivers for altered ventriloquism effects and aftereffects in older adults. *Cortex*, *135*, 298–310. https://doi.org/10.1016/j.cortex.2020.12.001

Passamonti, C., Frissen, I., & Làdavas, E. (2009). Visual recalibration of auditory spatial perception: two separate neural circuits for perceptual learning. *European Journal of Neuroscience*, *30*(6), 1141–1150. https://doi.org/10.1111/j.1460-9568.2009.06910.x

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*(4), 160–166. https://doi.org/10.1016/J.TICS.2009.01.006

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 1–16. https://doi.org/10.1038/s41467-021-26731-9

Pleger, B., Blankenburg, F., Ruff, C. C., Driver, J., & Dolan, R. J. (2008). Reward facilitates tactile judgments and modulates hemodynamic responses in human primary somatosensory cortex. *Journal of Neuroscience*, *28*(33), 8161–8168. https://doi.org/10.1523/JNEUROSCI.1093-08.2008

Pleger, B., Ruff, C. C., Blankenburg, F., Klöppel, S., Driver, J., & Dolan, R. J. (2009). Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biology*, *7*(7), e1000164. https://doi.org/10.1371/journal.pbio.1000164

Quintero, S. I., Shams, L., & Kamal, K. (2022). *Changing the Tendency to Integrate the Senses*. https://doi.org/10.3390/brainsci12101384

Radeau, M. (1985). Signal intensity, task context, and auditory-visual interactions. *Perception*, *14*(5), 571–577. https://doi.org/10.1068/p140571

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Current Psychology of Cognition*, *13*(1), 3–51.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology*, *26*(1), 63–71. https://doi.org/10.1080/14640747408400388

Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, *20*(4), 227–235. https://doi.org/10.3758/BF03199448

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, *22*(2), 137–146. https://doi.org/10.3758/BF03198746

Radeau, M., & Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*, *23*(4), 341–343. https://doi.org/10.3758/BF03199719

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X18000936

Rajan, R., Aitkin, L. M., Irvine, D. R. F., & McKay, J. (1990). Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and the effects of variations in stimulus parameters. *Journal of Neurophysiology*, *64*(3), 872–887. https://doi.org/10.1152/jn.1990.64.3.872

Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision Research*, *39*(11), 1963–1989. https://doi.org/10.1016/S0042-6989(98)00279-X

Rauschecker, J. P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends in Neurosciences*, *18*(1), 36–43. https://doi.org/10.1016/0166-2236(95)93948-W

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 869–875. https://doi.org/10.1073/pnas.95.3.869

Recanzone, G. H. (2000). Spatial processing in the auditory cortex of the macaque monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11829–11835. https://doi.org/10.1073/pnas.97.22.11829

Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology*, *59*, 119–142. https://doi.org/10.1146/annurev.psych.59.103006.093544

Redding, G. M., & Wallace, B. (1997). *Adaptive Spatial Alignment*. Psychology Press. https://books.google.com/books?id=_xFWPePK774C&pgis=1

Redding, G. M., & Wallace, B. (2002). Strategie Calibration and Spatial Alignment: A Model From Prism Adaptation. *Journal of Motor Behavior*, *34*(2), 126–138. https://doi.org/10.1080/00222890209601935

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - Revisited. *NeuroImage*, *84*, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1598), 2159–2168. https://doi.org/10.1098/rspb.2006.3578

Robinson, F. R., Soetedjo, R., & Noto, C. (2006). Distinct short-term and long-term adaptation to reduce saccade size in monkey. *Journal of Neurophysiology*, *96*(3), 1030–1041. https://doi.org/10.1152/jn.01151.2005

Röder, B., Kekunnaya, R., & Guerreiro, M. J. S. (2021). Neural mechanisms of visual sensitive periods in humans. *Neuroscience & Biobehavioral Reviews*, *120*, 86–99. https://doi.org/10.1016/j.neubiorev.2020.10.030

Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2016). Statistically Optimal Multisensory Cue Integration: A Practical Tutorial. *Multisensory Research*, *29*(4–5), 279–317. https://doi.org/10.1163/22134808-00002510

Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, *10*(1), 1–17. https://doi.org/10.1038/s41467-019-09664-2

Rohe, T., & Noppeney, U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biology*, *13*(2), 1–18. https://doi.org/10.1371/journal.pbio.1002073

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, *26*(4), 509–514. https://doi.org/10.1016/j.cub.2015.12.056

Rohe, T., & Noppeney, U. (2018). Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Attention. *Eneuro*, *5*(1), e0315-17. https://doi.org/10.1523/ENEURO.0315-17.2018

Rohlf, S., Bruns, P., & Röder, B. (2021). The Effects of Cue Reliability on Crossmodal Recalibration in Adults and Children. *Multisensory Research*, *34*(7), 743–761. https://doi.org/10.1163/22134808-bja10053

Rohlf, S., Li, L., Bruns, P., & Röder, B. (2020). Multisensory Integration Develops Prior to Crossmodal Recalibration. *Current Biology*, *30*(9), 1726-1732.e7. https://doi.org/10.1016/j.cub.2020.02.048

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A*, *22*(5), 801. https://doi.org/10.1364/josaa.22.000801

Rosas, P., & Wichmann, F. A. (2011). Cue Combination: Beyond Optimality. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory Cue Integration* (pp. 144–152). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.003.0008

Rose, T., Jaepel, J., Hübener, M., & Bonhoeffer, T. (2016). Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science*, *352*(6291), 1319–1322. https://doi.org/10.1126/science.aad3358

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-Induced Flash Illusion is Resistant to Feedback Training. *Brain Topography*, *21*, 185–192. https://doi.org/10.1007/s10548-009-0090-9

Samad, M., & Shams, L. (2018). Recalibrating the body: visuotactile ventriloquism aftereffect. *PeerJ*, *6*, e4504. https://doi.org/10.7717/peerj.4504

Sato, Y., & Körding, K. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 1–13. https://doi.org/10.1167/14.13.13

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli. *Neural Computation*, *19*(12), 3335–3355. https://doi.org/10.1162/neco.2007.19.12.3335

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, *11*(7), 1–17. https://doi.org/10.1167/11.7.12

Schween, R., McDougle, S. D., Hegele, M., & Taylor, J. A. (2020). Assessing explicit strategies in force field adaptation. *Journal of Neurophysiology*, *123*(4), 1552–1565. https://doi.org/10.1152/jn.00427.2019

Senna, I., Piller, S., Gori, M., & Ernst, M. (2022). The power of vision: calibration of auditory space after sight restoration from congenital cataracts. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1984), 20220768. https://doi.org/10.1098/rspb.2022.0768

Shalom-Sperber, S., Chen, A., & Zaidel, A. (2022). Rapid cross-sensory adaptation of self-motion perception. *Cortex*, *148*, 14–30. https://doi.org/10.1016/j.cortex.2021.11.018

Shams, L., & Beierholm, U. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Shams, L., & Beierholm, U. (2011). From Integration to Segregation: When and How the Human Nervous System Combines Crossmodal Sensory Signals. *Sensory Cue Integration*, 251–262.

Shams, L., & Beierholm, U. (2022). Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, *137*, 104619. https://doi.org/10.1016/j.neubiorev.2022.104619

Shapiro, K. L., Egerman, B., & Klein, R. M. (1984). Effects of arousal on human visual dominance. *Perception & Psychophysics*, *35*(6), 547–552. https://doi.org/10.3758/BF03205951

Shinn-Cunningham, B. (2000). Adapting to remapped auditory localization cues: A decision-theory model. *Perception & Psychophysics*, *62*(1), 33–47. https://doi.org/10.3758/BF03212059

Simon, D. M., Nidiffer, A. R., & Wallace, M. T. (2018). Single Trial Plasticity in Evidence Accumulation Underlies Rapid Recalibration to Asynchronous Audiovisual Speech. *Scientific Reports*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-30414-9

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1760–1765. https://doi.org/10.1073/pnas.0607208104

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10. https://doi.org/10.1097/00001756-200101220-00009

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting Adaptive Processes with Different Timescales Underlie Short-Term Motor Learning. *PLoS Biology*, *4*(6), e179. https://doi.org/10.1371/journal.pbio.0040179

Stange, A., Myoga, M. H., Lingner, A., Ford, M. C., Alexandrova, O., Felmy, F., Pecka, M., Siveke, I., & Grothe, B. (2013). Adaptation in sound localization: from GABA B receptor-mediated synaptic modulation to perception. *Nature Neuroscience*, *16*(12), 1840–1847. https://doi.org/10.1038/nn.3548

Stecker, G. C., Harrington, I. A., & Middlebrooks, J. C. (2005). Location Coding by Opponent Neural Populations in the Auditory Cortex. *PLoS Biology*, *3*(3), e78. https://doi.org/10.1371/journal.pbio.0030078

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Research*, *448*(2), 355–358. https://doi.org/10.1016/0006-8993(88)91276-0

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(40), 10154–10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. *The Journal of Neuroscience*, *34*(8), 3023–3032. https://doi.org/10.1523/JNEUROSCI.3619-13.2014

Temme, L. A., Maino, J. H., & Noell, W. K. (1985). Eccentricity perception in the periphery of normal observers and those with retinitis pigmentosa. *Optometry and Vision Science*, *62*(11), 736–743. https://doi.org/10.1097/00006324-198511000-00003

Tong, J., Li, L., Bruns, P., & Röder, B. (2020). Crossmodal associations modulate multisensory spatial integration. *Attention, Perception, and Psychophysics*, *82*(7), 3490–3506. https://doi.org/10.3758/s13414-020-02083-2

Tong, J., Parisi, G. I., Wermter, S., & Röder, B. (2018). Closing the loop on multisensory interactions: A neural architecture for multisensory causal inference and recalibration. *ArXiv*, *Trr 169*, 1–30. https://doi.org/https://doi.org/10.48550/arXiv.1802.06591

Toni, T., Ozaki, Y.-I., Kirk, P., Kuroda, S., & Stumpf, M. P. H. (2012). Elucidating the in vivo phosphorylation dynamics of the ERK MAP kinase using quantitative proteomics data and Bayesian model selection. *Molecular BioSystems*, *8*(7), 1921. https://doi.org/10.1039/c2mb05493k

Toni, T., & Stumpf, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, *26*(1), 104–110. https://doi.org/10.1093/bioinformatics/btp619

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision*, *16*(3–4), 255–275. https://doi.org/10.1163/156856803322467527

Ursino, M., Cuppini, C., Magosso, E., Beierholm, U., & Shams, L. (2019). Explaining the Effect of Likelihood Manipulation and Prior Through a Neural Network of the Audiovisual Perception of Space. *Multisensory Research*, *32*(2), 111–144. https://doi.org/10.1163/22134808-20191324

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When Feeling Is More Important Than Seeing in Sensorimotor Adaptation. *Current Biology*, *12*(10), 834–837. https://doi.org/10.1016/S0960-9822(02)00836-9

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

Van Der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports*, *5*(1), 14526. https://doi.org/10.1038/srep14526

Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience*, *31*(10), 1763–1771. https://doi.org/10.1111/j.1460-9568.2010.07198.x

van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. *Frontiers in Psychology*, *4*(JUL), 388. https://doi.org/10.3389/fpsyg.2013.00388

Vercillo, T., Burr, D., Sandini, G., & Gori, M. (2015). Children do not recalibrate motor-sensory temporal order after exposure to delayed sensory feedback. *Developmental Science*, *18*(5), 703–712. https://doi.org/10.1111/desc.12247

Vigneault-MacLean, B. K., Hall, S. E., & Phillips, D. P. (2007). The effects of lateralized adaptors on lateral position judgements of tones within and across frequency channels. *Hearing Research*, *224*(1–2), 93–100. https://doi.org/10.1016/j.heares.2006.12.001

von Helmholtz, H. (1896). Handbuch der physiologischen Optik. In L. Voss (Ed.), *Allgemeine Encyklopädie der Physik* (1st ed.). Gustav Karsten.

Vroomen, J. (2001). *The ventriloquist effect does not depend on the direction of automatic visual attention*. *63*(4), 651–659.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vroomen, J., & Stekelenburg, J. J. (2014). A bias-free two-alternative forced choice procedure to examine intersensory illusions applied to the ventriloquist effect by flashes and averted eye-gazes. *European Journal of Neuroscience*, *39*(9), 1491–1498. https://doi.org/https://doi.org/10.1111/ejn.12525

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. a. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *158*(2), 252–258. https://doi.org/10.1007/s00221-004-1899-9

Watson, D. M., Akeroyd, M. A., Roach, N. W., & Webb, B. S. (2019). Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history. *Scientific*

Reports, *9*(1), 8513. https://doi.org/10.1038/s41598-019-44984-9

Weber, A. I., Krishnamurthy, K., & Fairhall, A. L. (2019). Coding Principles in Adaptation. *Annual Review of Vision Science*, *5*(1), 427–449. https://doi.org/10.1146/annurev-vision-091718-014818

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, *1*(1), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Welch, R. B. (1978). Adaptation to visual transposition. In R. B. Welch (Ed.), *Perceptual Modification : Adapting to Altered Sensory Environments.* (p. 365). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-741850-6.50001-X

Welch, R. B. (1999). Chapter 15 Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). https://doi.org/10.1016/S0166-4115(99)80036-3

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.88.3.638

Werner-Reiss, U., & Groh, J. M. (2008). A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *The Journal of Neuroscience*, *28*(14), 3747–3758. https://doi.org/10.1523/JNEUROSCI.5044-07.2008

Werner, S., & Diedrichsen, J. (2002). The time course of spatial memory distortions. *Memory and Cognition*, *30*(5), 718–730. https://doi.org/10.3758/BF03196428

Willmore, B. D. B., & King, A. J. (2023). Adaptation in auditory processing. *Physiological Reviews*, *103*(2), 1025–1058. https://doi.org/10.1152/physrev.00011.2022

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wilson, R. C., & Finkel, L. H. (2009). A Neural Implementation of the Kalman Filter. *Advances in Neural Information Processing Systems*, *22*.

Wissig, S. C., Patterson, C. A., & Kohn, A. (2013). Adaptation improves performance on a visual search task. *Journal of Vision*, *13*(2), 6–6. https://doi.org/10.1167/13.2.6

Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., & Recanzone, G. H. (2006). Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *Journal of Neurophysiology*, *96*(6), 3323–3337. https://doi.org/10.1152/jn.00392.2006

Woods, T. M., & Recanzone, G. H. (2004). Visually Induced Plasticity of Auditory Spatial Perception in Macaques. *Current Biology*, *14*(17), 1559–1564. https://doi.org/10.1016/j.cub.2004.08.059

Wozny, D., Beierholm, U., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, *6*(8), e1000871. https://doi.org/10.1371/journal.pcbi.1000871

Wozny, D., & Shams, L. (2011a). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, *5*(November), 1–11. https://doi.org/10.3389/fnint.2011.00075

Wozny, D., & Shams, L. (2011b). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *Journal of Neuroscience*, *31*(12), 4607–4612. https://doi.org/10.1523/JNEUROSCI.6079-10.2011

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zaidel, A., Laurens, J., DeAngelis, G. C., & Angelaki, D. E. (2021). Supervised Multisensory Calibration Signals Are Evident in VIP But Not MSTd. *The Journal of Neuroscience*, *41*(49), 10108–10119. https://doi.org/10.1523/JNEUROSCI.0135-21.2021

Zaidel, A., Ma, W., & Angelaki, D. E. (2013). Supervised calibration relies on the multisensory percept. *Neuron*, *80*(6), 1544–1557. https://doi.org/10.1016/j.neuron.2013.09.026

Zaidel, A., Turner, A. H., & Angelaki, D. E. (2011). Multisensory Calibration Is Independent of Cue Reliability. *The Journal of Neuroscience*, *31*(39), 13949–13962. https://doi.org/10.1523/JNEUROSCI.2732-11.2011

Zierul, B., Röder, B., Tempelmann, C., Bruns, P., & Noesselt, T. (2017). The role of auditory cortex in the spatial ventriloquism aftereffect. *NeuroImage*, *162*, 257–268. https://doi.org/10.1016/J.NEUROIMAGE.2017.09.002

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical Band Width in Loudness Summation. *The Journal of the Acoustical Society of America*, *29*(5), 548–557. https://doi.org/10.1121/1.1908963

Zwiers, M. P., Van Opstal, A. J., & Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*, *6*(2), 175–181. https://doi.org/10.1038/nn999

Appendix

Contents Appendix

Appendix A: Chapter III

Figures

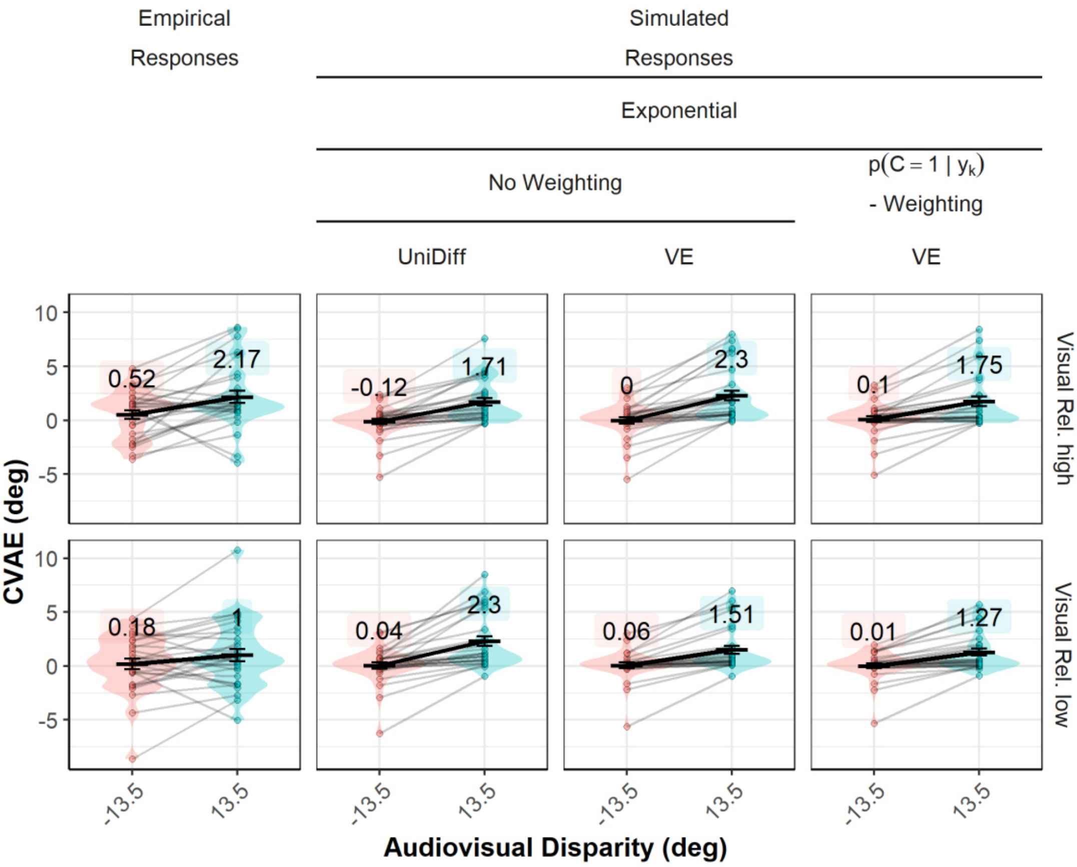


Figure A. 1. Empirical results and posterior simulations of died out CVAE models. Auditory average CVAEs as a function of visual reliability (top vs. bottom row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Model factors and abbreviations are described in detail in Chapter II.

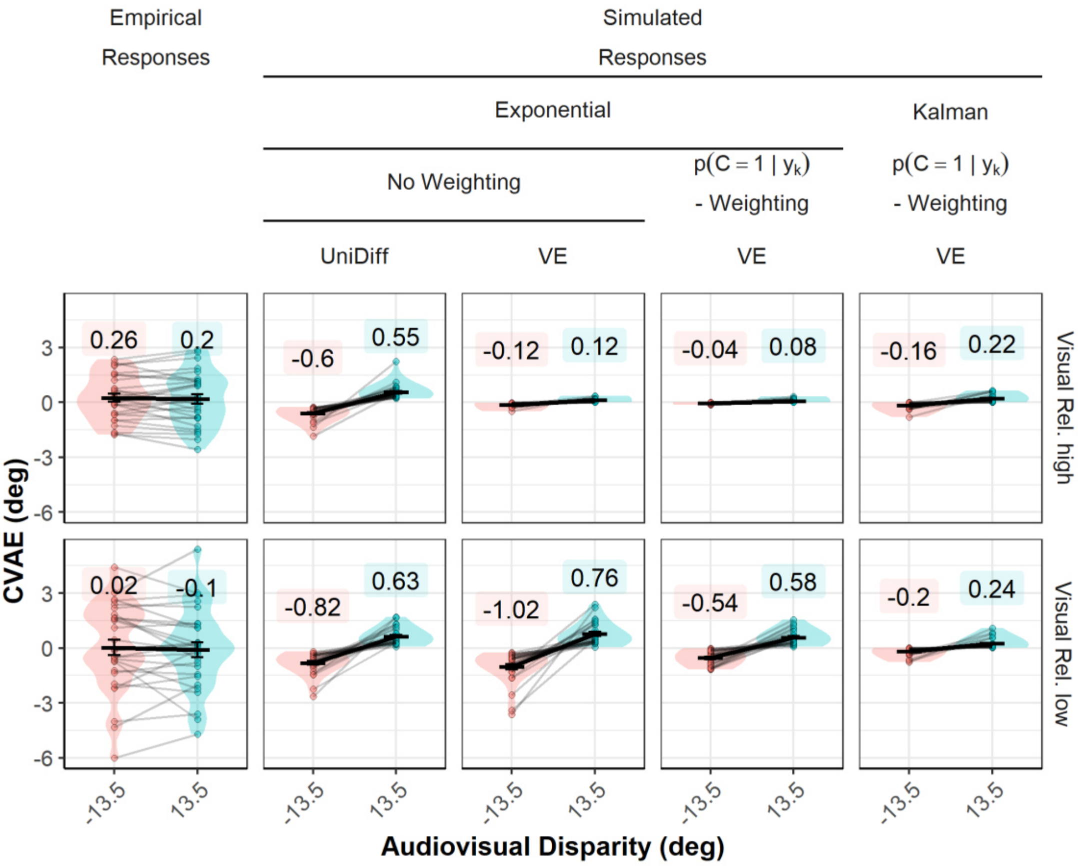


Figure A. 2. Empirical results and posterior simulations of died out CVAE models. Visual average CVAEs as a function of visual reliability (top vs. bottom row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Model factors and abbreviations are described in detail in Chapter II.

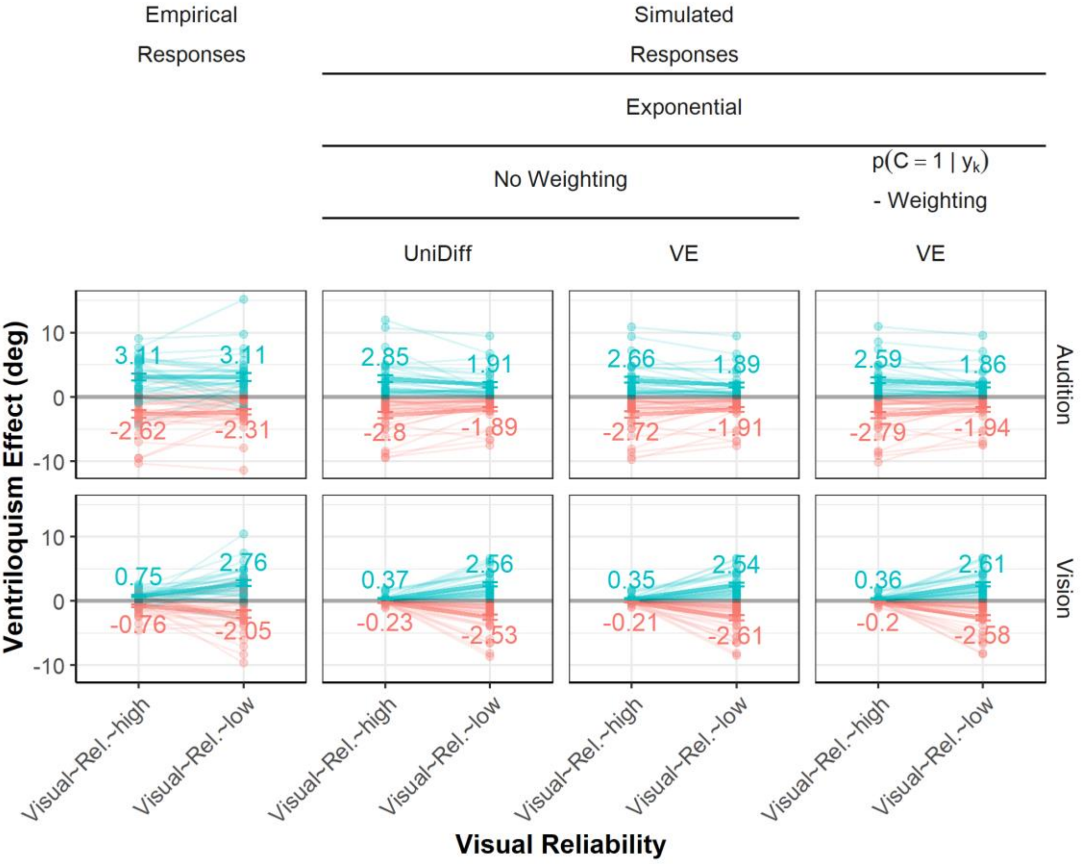


Figure A. 3. Empirical results and posterior simulations of died out CVAE models. Auditory and visual average VEs as a function of visual reliability (top vs. bottom row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Model factors and abbreviations are described in detail in Chapter II.

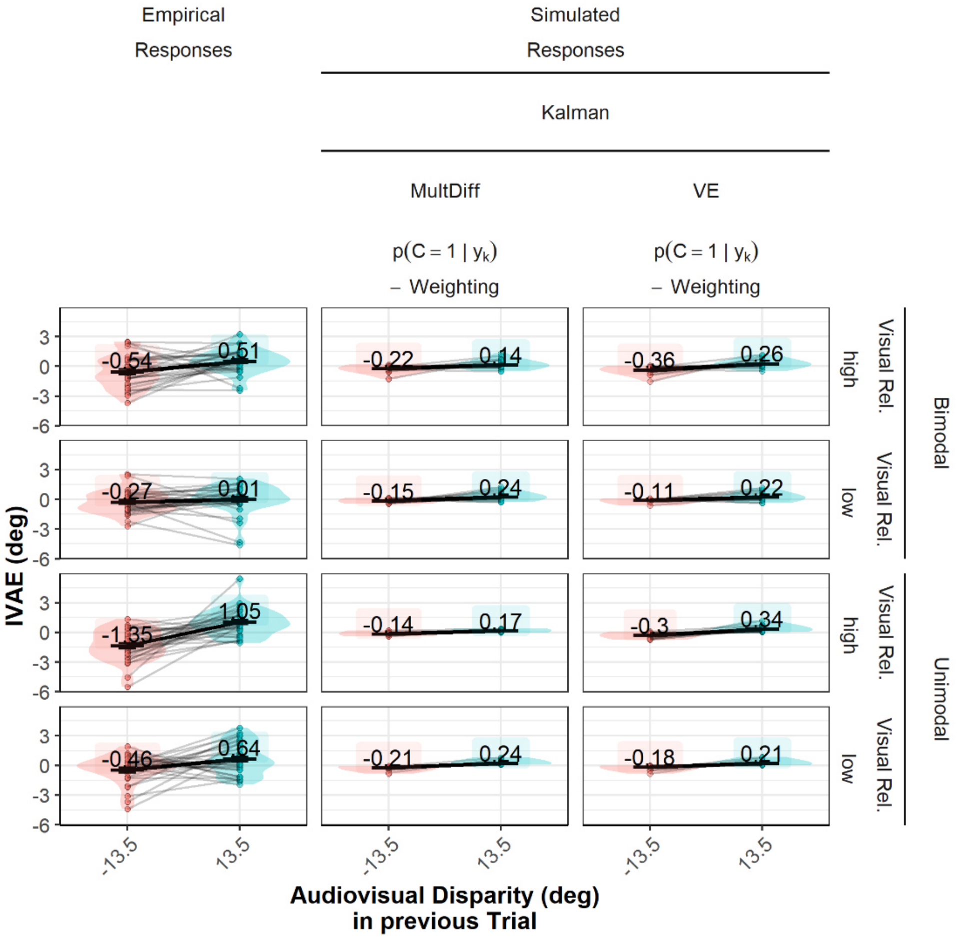


Figure A. 4. Empirical results and posterior simulations of the best fitting IVAE models with delayed update. Models are categorized by the errorterm, learning mechanism and type of CI-Weighting. All models apply delayed updating of the IVAE. Auditory average IVAEs as a function of visual reliability and trial type (top to bottom row) and audio-visual disparity (left vs. right, negative values indicate shifts to the left). Model factors and abbreviations are described in detail in Chapter II.

Tables

Table A. 1

LMM of the auditory, bimodal Res as an estimate of trial-wise bias in perceptual space

Term	*statistic*	*df*	*p.value*
reliability	0.04	1	0.846
previous disparity	0.22	1	0.638
disparity	2111.20	1	<0.001
reliability:previous disparity	0.13	1	0.718
reliability:disparity	0.06	1	0.812
previous disparity:disparity	6.38	1	0.012
reliability:previous disparity:disparity	0.41	1	0.522

Table A. 2
LMM of the visual, bimodal Res as an estimate of trial-wise bias in perceptual space

term	*statistic*	*df*	*p.value*
reliability	0.02	1	0.88
previous disparity	1.27	1	0.26
disparity	601.75	1	<0.001
reliability:previous disparity	0.59	1	0.44
reliability:disparity	678.86	1	<0.001
previous disparity:disparity	0.12	1	0.73
reliability:previous disparity:disparity	0.01	1	0.93

Table A. 3

LMM of the corrected, auditory, unimodal Res as an estimate of trial-wise bias in cue space

term	*statistic*	*df*	*p.value*
reliability	16.98	1	<0.001
previous disparity	0.48	1	0.49
disparity	86.48	1	<0.001
previous task	0.35	1	0.55
reliability:previous disparity	0.09	1	0.76
reliability:disparity	14.25	1	<0.001
previous disparity:disparity	53.91	1	<0.001
reliability:previous task	1.96	1	0.16
previous disparity:previous task	0.43	1	0.51
disparity:previous task	0.77	1	0.38
reliability:previous disparity:disparity	13.36	1	<0.001
reliability:previous disparity:previous task	0.31	1	0.58
reliability:disparity:previous task	0.22	1	0.64
previous disparity:disparity:previous task	2.21	1	0.14
reliability:previous disparity:disparity:previous task	1.50	1	0.22

Table A. 4

LMM of the corrected, visual, unimodal Res as an estimate of trial-wise bias in cue space

term	*statistic*	*df*	*p.value*
reliability	4.69	1	0.030
previous disparity	1.45	1	0.229
disparity	1.90	1	0.168
previous task	0.49	1	0.485
reliability:previous disparity	1.11	1	0.292
reliability:disparity	0.50	1	0.478
previous disparity:disparity	0.02	1	0.884
reliability:previous task	0.13	1	0.723
previous disparity:previous task	3.05	1	0.081
disparity:previous task	0.06	1	0.803
reliability:previous disparity:disparity	1.42	1	0.234
reliability:previous disparity:previous task	4.35	1	0.037
reliability:disparity:previous task	0.11	1	0.745
previous disparity:disparity:previous task	1.08	1	0.299
reliability:previous disparity:disparity:previous task	0.70	1	0.402

Table A. 5

Aggregated Parameter Estimates for all tested VE Models

Weighting	*parameter*	*MED*	*MAD*	*M*	*SEM*	*SD*
optimal	ω	0.00	0.00	0.00	0.00	0.00
optimal	$p(C = 1)$	0.40	0.42	0.43	0.04	0.33
overestimating	$\omega_{SE,V}$	0.32	1.39	-0.20	0.22	1.77
overestimating	$p(C = 1)$	0.37	0.42	0.41	0.04	0.34
overweighting	$\omega_{SW,V}$	-0.30	1.33	-0.11	0.27	2.19
overweighting	$p(C = 1)$	0.37	0.41	0.41	0.04	0.32

Note: Detailed definitions of the model factors and parameters are given in Chapter II. Importantly Weighting refers to the reliability-based weights for multisensory integration.

Table A. 6

Aggregated Parameter Estimates for all tested CVAE Models

Learning Mechanism	*Errorterm*	*CI-Weighting*	*parameter*	*MED*	*MAD*	*M*	*SEM*	*SD*
Kalman	MultDiff	No Weighting	$p(C=1)$	0.49	0.48	0.45	0.06	0.34
Kalman	MultDiff	No Weighting	$\sigma_{ßAleft}$	0.04	0.04	0.19	0.06	0.34
Kalman	MultDiff	No Weighting	$\sigma_{ßAright}$	-2.80	1.55	-2.90	0.39	2.21
Kalman	MultDiff	No Weighting	$\sigma_{ßV}$	0.13	0.15	0.16	0.02	0.14
Kalman	MultDiff	No Weighting	t_A^{CVAE}	0.35	0.51	0.44	0.06	0.36
Kalman	MultDiff	$p(C=1\|y_k)$ - Weighting	$p(C=1)$	0.31	0.44	0.44	0.06	0.36
Kalman	MultDiff	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßAleft}$	0.06	0.08	0.30	0.09	0.49
Kalman	MultDiff	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßAright}$	-0.83	1.47	-2.06	0.59	3.33
Kalman	MultDiff	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßV}$	0.33	0.42	0.31	0.05	0.28
Kalman	MultDiff	$p(C=1\|y_k)$ - Weighting	t_A^{CVAE}	0.44	0.48	0.44	0.06	0.36
Kalman	VE	$p(C=1\|y_k)$ - Weighting	$p(C=1)$	0.42	0.54	0.43	0.07	0.37
Kalman	VE	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßAleft}$	0.04	0.05	0.14	0.04	0.21
Kalman	VE	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßAright}$	-1.83	1.75	-2.54	0.46	2.59
Kalman	VE	$p(C=1\|y_k)$ - Weighting	$\sigma_{ßV}$	0.06	0.07	0.21	0.05	0.31
Kalman	VE	$p(C=1\|y_k)$ - Weighting	t_A^{CVAE}	0.36	0.45	0.38	0.06	0.35
RBA	RBA	No Weighting	$p(C=1)$	0.39	0.46	0.40	0.06	0.35
RBA	RBA	No Weighting	$\sigma_{ßAleft}$	NA	NA	NaN	NA	NA
RBA	RBA	No Weighting	$\sigma_{ßAright}$	-6.98	1.17	-7.27	0.18	1.04
RBA	RBA	No Weighting	$\sigma_{ßV}$	NA	NA	NaN	NA	NA
RBA	RBA	No Weighting	t_A^{CVAE}	0.46	0.63	0.46	0.07	0.40

Note: Parameters for models are only presented if the estimated frequency was different from zero. Detailed definitions of the model factors and parameters are given in Chapter II.

Table A. 7

Aggregated Parameter Estimates for all tested IVAE Models

Learning Mechanism	*Errorterm*	*CI-Weighting*	*Processing Level*	*Time of Update*	*Parameter*	*MED*	*MAD*	*M*	*SEM*	*SD*
Exponential	MultDiff	No Weighting	resp	i.	d_{IVAE}	0.25	0.00	0.40	0.04	0.24
Exponential	MultDiff	No Weighting	resp	i.	$p(C = 1)$	0.44	0.37	0.44	0.05	0.29
Exponential	MultDiff	No Weighting	resp	i.	$\sigma_{ßAleft}$	0.13	0.15	0.30	0.11	0.62
Exponential	MultDiff	No Weighting	resp	i.	$\sigma_{ßAright}$	-3.33	3.66	-3.86	0.68	3.86
Exponential	MultDiff	No Weighting	resp	i.	$\sigma_{ßV}$	0.20	0.14	0.26	0.05	0.31
Kalman	MultDiff	No Weighting	resp	i.	d_{IVAE}	0.33	0.11	0.40	0.04	0.21
Kalman	MultDiff	No Weighting	resp	i.	$p(C = 1)$	0.46	0.34	0.46	0.05	0.29
Kalman	MultDiff	No Weighting	resp	i.	$\sigma_{ßAleft}$	0.69	0.99	1.70	0.68	3.87
Kalman	MultDiff	No Weighting	resp	i.	$\sigma_{ßAright}$	0.10	6.28	-1.47	0.88	4.98
Kalman	MultDiff	No Weighting	resp	i.	$\sigma_{ßV}$	0.47	0.67	1.34	0.36	2.02
Kalman	PRI	$p(C = 1\|y_k)$-Weighting	percept	i.	d_{IVAE}	0.32	0.10	0.43	0.04	0.25
Kalman	PRI	$p(C = 1\|y_k)$-Weighting	percept	i.	$p(C = 1)$	0.36	0.36	0.44	0.05	0.30
Kalman	PRI	$p(C = 1\|y_k)$-Weighting	percept	i.	$\sigma_{ßAleft}$	0.64	0.92	2.69	0.79	4.49
Kalman	PRI	$p(C = 1\|y_k)$-Weighting	percept	i.	$\sigma_{ßAright}$	1.77	3.72	0.38	0.83	4.72
Kalman	PRI	$p(C = 1\|y_k)$-Weighting	percept	i.	$\sigma_{ßV}$	0.17	0.20	2.26	0.97	5.46

Note: Parameters for models are only presented if the estimated frequency was different from zero. Detailed definitions of the model factors and parameters are given in Chapter II. Processing Levels are abbreviated (resp= Response Level, percept= Perceptual Level) as well as Time of Update (i.= instantaneous).

Appendix B: Chapter IV

Tables

Table B. 1
LMM of Res across all experimental factors and stimulus types

term	*statistic*	*df*	*p.value*
association type	1.19	1	0.2745
block number	25.89	1	<0.001
absolute disparity	742.22	1	<0.001
block type	5.80	1	0.0160
stimulus type	2884.90	1	<0.001
previous disparity	123.42	1	<0.001
association type:block number	21.53	1	<0.001
association type:absolute disparity	3.98	1	0.0462
block number:absolute disparity	0.24	1	0.6240
block number:block type	10.11	1	0.0015
absolute disparity:block type	5.08	1	0.0241
association type:stimulus type	0.03	1	0.8711
block number:stimulus type	0.00	1	0.9621
absolute disparity:stimulus type	127.06	1	<0.001
association type:previous disparity	0.15	1	0.6990
absolute disparity:previous disparity	7.55	1	0.0060
stimulus type:previous disparity	13.06	1	<0.001
association type:block number:absolute disparity	2.21	1	0.1374
association type:block number:stimulus type	0.17	1	0.6767
association type:absolute disparity:stimulus type	1.26	1	0.2623
block number:absolute disparity:stimulus type	2.38	1	0.1229
association type:absolute disparity:previous disparity	4.38	1	0.0363
association type:block number:absolute disparity:stimulus type	7.66	1	0.0056

Note: Due to the large number of fixed factors we applied a stepwise procedure to reduce the number of factors (Hastie & Pregibon, 2017).

Table B. 2

Summary of free parameter in all tested models of study 2

adaptive Pcom	*CVAE errorterm*	*CVAE learning mechanism*	*CVAE posterior weighting*	*IVAE errorterm*	*IVAE learning mechanism*	*IVAE posterior weighting*	*IVAE processing stage*	d_{IVAE}	$d_{p(C=1)}$	$\lambda_{p(C=1)}$	$p(C=1)$	$\sigma_{ßCAL}$	$\sigma_{ßCAR}$	$\sigma_{ßIAL}$	$\sigma_{ßIAR}$	t_A^{CVAE}	t_A^{IVAE}
-	MultDiff	Kalman	NW	MultDiff	Kalman	NW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	MultDiff	Kalman	NW	MultDiff	Kalman	PW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	MultDiff	Kalman	NW	No IVAE				-	-	-	✓	✓	✓	-	-	✓	-
-	MultDiff	Kalman	NW	PRI	Kalman	PW	perceptual	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	MultDiff	Kalman	PW	MultDiff	Kalman	NW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	MultDiff	Kalman	PW	MultDiff	Kalman	PW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	MultDiff	Kalman	PW	No IVAE				-	-	-	✓	✓	✓	-	-	✓	-
-	MultDiff	Kalman	PW	PRI	Kalman	PW	perceptual	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	No CVAE			MultDiff	Kalman	NW	response	✓	-	-	✓	-	-	✓	✓	-	✓
-	No CVAE			MultDiff	Kalman	PW	response	✓	-	-	✓	-	-	✓	✓	-	✓
-	No CVAE			No IVAE				-	-	-	✓	-	-	-	-	-	-
-	No CVAE			PRI	Kalman	PW	perceptual	✓	-	-	✓	-	-	✓	✓	-	✓
-	VE	Kalman	PW	MultDiff	Kalman	NW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	VE	Kalman	PW	MultDiff	Kalman	PW	response	✓	-	-	✓	✓	✓	✓	✓	✓	✓
-	VE	Kalman	PW	No IVAE			perceptual	-	-	-	✓	✓	✓	-	-	✓	-
-	VE	Kalman	PW	PRI	Kalman	PW	perceptual	✓	-	-	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	NW	MultDiff	Kalman	NW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	NW	MultDiff	Kalman	PW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	NW	No IVAE				-	✓	✓	✓	✓	✓	-	-	✓	-
✓	MultDiff	Kalman	NW	PRI	Kalman	PW	perceptual	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	PW	MultDiff	Kalman	NW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	PW	MultDiff	Kalman	PW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	MultDiff	Kalman	PW	No IVAE				-	✓	✓	✓	✓	✓	-	-	✓	-
✓	MultDiff	Kalman	PW	PRI	Kalman	PW	perceptual	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	No CVAE			MultDiff	Kalman	NW	response	✓	✓	✓	✓	-	-	✓	✓	-	✓
✓	No CVAE			MultDiff	Kalman	PW	response	✓	✓	✓	✓	-	-	✓	✓	-	✓
✓	No CVAE			No IVAE				-	✓	✓	✓	-	-	-	-	-	-
✓	No CVAE			PRI	Kalman	PW	perceptual	✓	✓	✓	✓	-	-	✓	✓	-	✓
✓	VE	Kalman	PW	MultDiff	Kalman	NW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	VE	Kalman	PW	MultDiff	Kalman	PW	response	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	VE	Kalman	PW	No IVAE				-	✓	✓	✓	✓	✓	-	-	✓	-
✓	VE	Kalman	PW	PRI	Kalman	PW	perceptual	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Detailed definitions of the model factors and parameters are given in Chapter II. Additional abbreviations involve PW for $p(C=1|y_k)$-Weighting, NW for no Weighting, Pcom for $p(C=1)$, perceptual for perceptual level and response for response level.

Appendix C: Chapter V

Figures

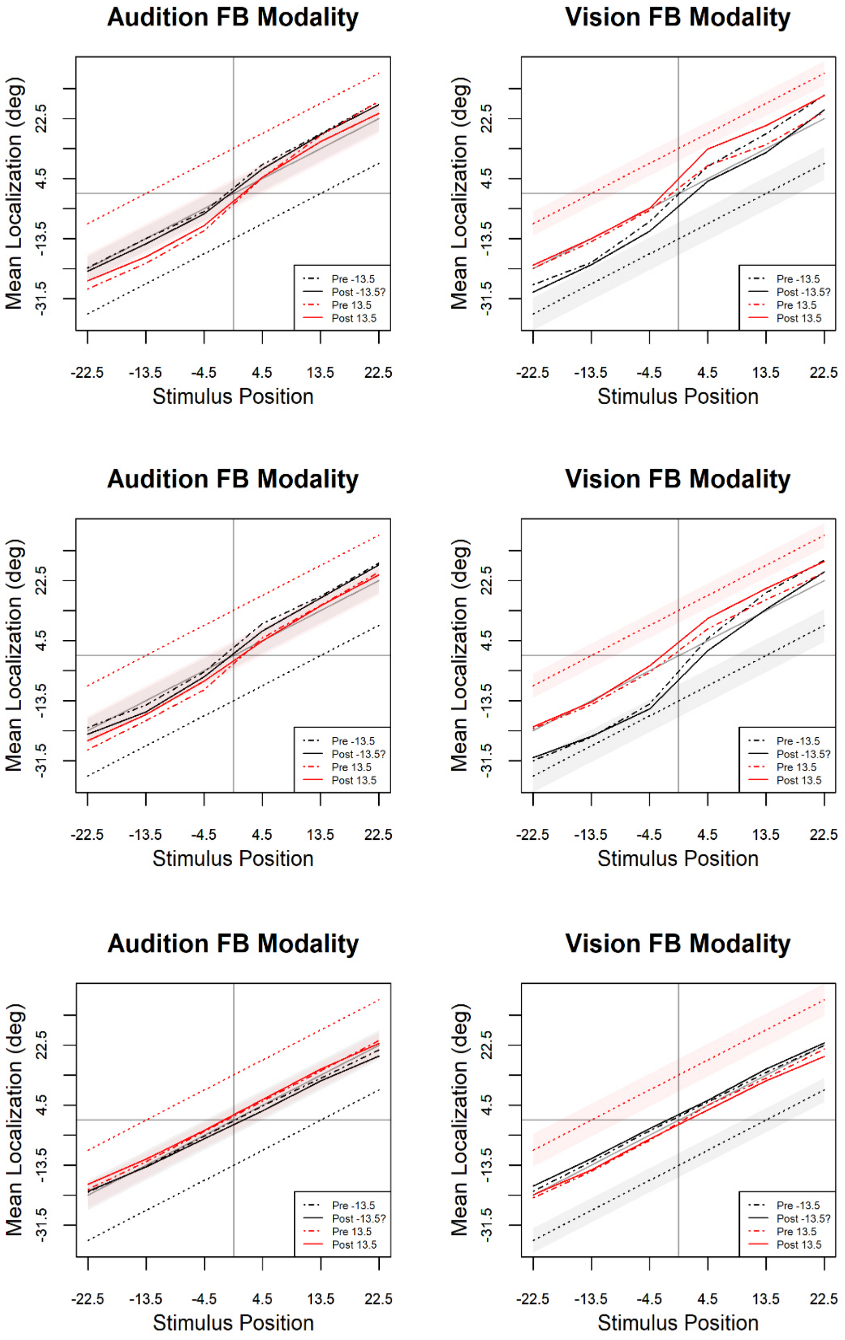


Figure C. 1. Auditory and visual localization behavior for the AS (A and B), the control sound (C and D) and the visual stimulus (E and F) in unimodal blocks. The first column shows results when audition was the feedback modality, and the second column shows the results when vision was the feedback modality. Each panel shows results separately for pretest (dashed lines) and posttest (solid lines). Red lines represent sessions where the audio-visual disparity during adaptation was to the right, and black lines show results for sessions where the audio-visual disparity was to the left. Shaded areas represent reward zones for the respective conditions.

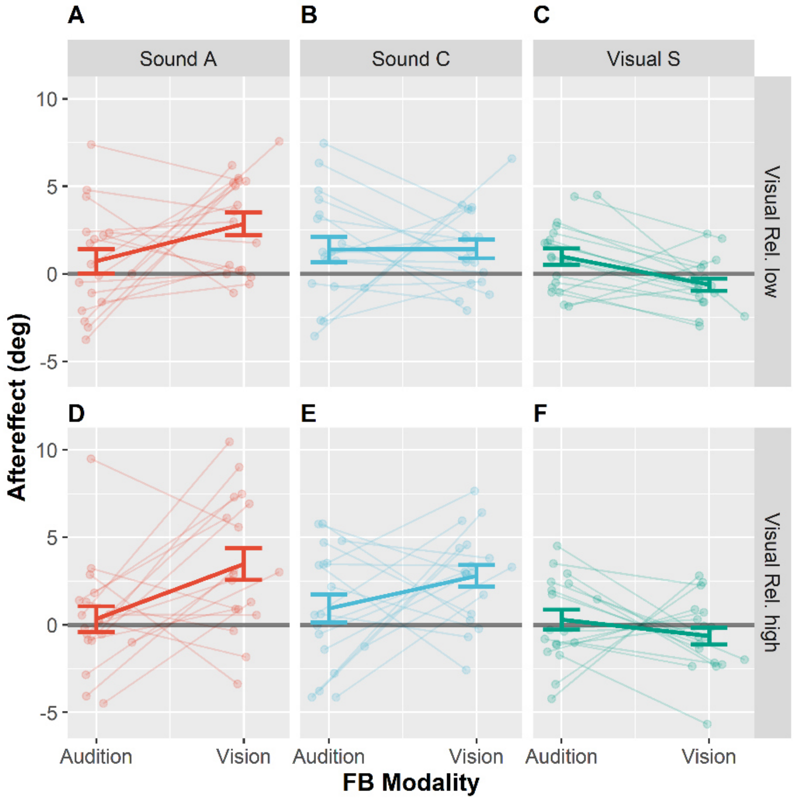


Figure C. 2. Ventriloquism aftereffects shown separately for the different stimulus types (Sound A, Sound C and Visual S) and collapsed over leftward and rightward audio-visual disparities. Mean ventriloquism aftereffects for different levels of visual reliability are depicted in separate rows (Visual Rel. low or Visual Rel. high). Each panel shows aftereffects separately for the conditions Audition FB modality and Vision FB Modality. Individual data is shown with light-colored points and lines whereas sample averages are indicated by dark-colored bold lines. Paired data points (i.e., individual data from a single participant) are connected via lines. Values were calculated as difference between pre- and posttest localization error multiplied with the sign of the audio-visual discrepancy. Thus, shifts in the direction of the competing stimulus during adaptation are positive. Error bars represent the standard error of the mean.

Tables

Table C. 1

Total number of participants for each combination of sound frequency of the adapted sound, FB modality and visual reliability.

		Audition FB Modality		Vision FB Modality	
		Visual Rel. low	Visual Rel. high	Visual Rel. low	Visual Rel. high
Sound Frequency Adapted	250 Hz	5	5	5	3
	500 Hz	4	4	6	4
	1000 Hz	5	5	3	5
	2000 Hz	4	4	4	6

Table C. 2
Total number of participants for each combination of session number, FB modality and visual reliability.

		Audition FB Modality		Vision FB Modality	
		Visual Rel. low	Visual Rel. high	Visual Rel. low	Visual Rel. high
Session Number	1	5	4	4	5
	2	3	5	5	5
	3	5	5	4	4
	4	5	4	5	4

Note: The session numbers reflect the temporal order of sessions for each participant.

Table C. 3

Total number of participants for each combination of session number, FB modality and visual reliability.

Session Number		Audition FB Modality		Vision FB Modality	
		Visual Rel. low	Visual Rel. high	Visual Rel. low	Visual Rel. high
	1	5	4	4	5
	2	3	5	5	5
	3	5	5	4	4
	4	5	4	5	4

Note: The session numbers reflect the temporal order of sessions for each participant.

Table C. 4

Total number of participants for each combination of audio-visual discrepancy, FB modality and visual reliability.

		Audition FB Modality		Vision FB Modality	
		Visual Rel. low	Visual Rel. high	Visual Rel. low	Visual Rel. high
Audio-visual Discrepancy	13.5°	5	5	5	3
	-13.5°	4	4	6	4

Acknowledgements

This thesis was supported by the German Research Foundation [Deutsche Forschungsgemeinschaft (DFG)] Grant TRR 169/A1.

First, I would like to thank Brigitte Röder for the opportunity to delve into the depths of academia. Furthermore, I would like to thank Patrick Bruns for his continuous support, for always being accessible when needed and his general effort to make things easier. Thanks go to the members of my committee: Frank Rösler, Sebastian Gluth and Timo Gerkmann. Moreover, I am grateful to Cordula Hölig and Kirsten Hötting, teaching has always been a great experience with your support. Moreover, thanks for helping with data collection to all my research assistants and especially Suong Nguyen, Berit Hecht, Lily Huber and Elisa Bußkamp as well as Dagmar Tödter and Nicola Kaczmarek. Additionally, I would like to thank Jan Mück and his workshop colleagues for helping me with the installation of the various experimental setups.

Special thanks go to my colleagues and friends Liesa Stange, Daniela Schönberger, Andreas Weiß and Madita Linke for running the last mile together. Moreover, Jonathan Tong, Lux Li and Rashi Pant provided precious feedback to diverse parts of my work.

Finally, and most importantly, I am infinitely grateful to Lena Hartung for always having my back and being endlessly patient with me. You are certainly the reason that I did not get lost along the way. And of course, I must thank Mila Hartung, for motivating me to get my act together and for being a constant source of joy.

UHH Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Erklärung gemäß *(bitte Zutreffendes ankreuzen)*

☐ **§ 4 (1c) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☒ **§ 5 (4d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich,

Alexander, Kramer (Vorname, Nachname),

dass ich mich an einer anderen Universität oder Fakultät noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer Doktorprüfung bemüht habe.

Hamburg, 18.07.2023
Ort, Datum

Unterschrift

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT

Institut für Bewegungswissenschaft
Institut für Psychologie

Eidesstattliche Erklärung nach ***(bitte Zutreffendes ankreuzen)***

□ **§ 7 (4) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☒ **§ 9 (1c und 1d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich an Eides statt,

1. dass die von mir vorgelegte Dissertation nicht Gegenstand eines anderen Prüfungsverfahrens gewesen oder in einem solchen Verfahren als ungenügend beurteilt worden ist.
2. dass ich die von mir vorgelegte Dissertation selbst verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die wörtlich oder inhaltlich übernommenen Stellen habe ich als solche kenntlich gemacht.

Hamburg, 18.07.2023
Ort, Datum

Unterschrift