MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Examining the role of the pupil-linked arousal system in perceptual decision-making using a dynamic sound localization task"

verfasst von / submitted by

Fabian Dorok, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge
Mitbetreut von / Co-Supervisor:	

Table of Contents

1. Introduction

1.1 Perception under uncertainty

The perceptual organism is often confronted with sensory uncertainty (Aitchison & Lengyel, 2017; Bejjanki et al., 2016; Knill & Pouget, 2004; Charlton et al., 2022; Ma, 2019; Yu & Dayan, 2005). Sensory uncertainty can, for example, arise due blurry visual stimuli (Stocker & Simoncelli, 2006), variability of audio-visual stimulus locations (Krishnamurthy et al., 2017) or internal noise of the perceptual system (Ma, 2019). How can we achieve optimal perception despite unfavourable perceptual conditions?

1.1.1 Perception as the integration of prior beliefs and sensory input

The brain achieves high perceptual accuracy despite sensory uncertainty by making use of prior beliefs (*priors*) about statistical regularities of the sensory environment (Charlton et al., 2022; Clark, 2013; Krishnamurthy et al., 2017; Nassar et al., 2012; Seriès & Seitz, 2013; Yon & Frith, 2021). These statistical regularities can, for example, be the probability of a certain number occurring (Nassar et al., 2012) or the probability of a stimulus occurring at a certain location (Filipowicz et al., 2020; Krishnamurthy et al., 2017). Importantly, perception seems to be the integrate of sensory input and priors (Aitchison & Lengyel, 2017; Knill & Pouget, 2004; Krishnamurthy et al., 2017; Schröger & Roeber, 2021; Sedley et al., 2016; Skerritt-Davis & Elhilali, 2021). Imagine looking into the mirror after a steamy shower. By making use of the prior about essential features of your face, you can still make out most of it, even though the mirror is foggy. Hence, you combined the prior about your face's features with the (imprecise) sensory input obtained from looking in the foggy mirror, resulting in a precise perception of your face. The perceptual system can therefore reach high perceptual accuracy by making use of prior beliefs.

The idea that priors influence perception dates back to Helmholtz, though sophisticated computational models that describe perception based on this notion emerged only in the past decade (de Lange et al., 2018). One computational approach that has gained traction in recent years is *Bayesian inference* (Achtziger et al., 2014; Aitchison & Lengyel, 2017; de Lange et al., 2018; Krishnamurthy et al., 2017). According to this approach, perception is a process of probabilistic inference (Clark, 2013) where the perceptual decision-maker decides on the most probable state of the world based on the *posterior distribution* (de Lange et al., 2018; Ma, 2019). The posterior distribution is proportional to the product of a *likelihood distribution* (i.e., internal estimates about the accuracy of the sensory input, Yon & Frith, 2021) and a *prior distribution* (i.e., estimates about the statistical regularities in the sensory environment; de Lange et al., 2018; Ma, 2019). This Bayesian inference model describes perceptual optimality by reducing perceptual uncertainty (i.e., the width of the posterior distribution, Tenenbaum et al., 2011; Ma, 2019; Yon & Frith, 2021; Yu & Dayan, 2005). The reduction of perceptual uncertainty is achieved by making appropriate use of the prior distribution (Petzschner et al., 2015; Ma, 2019). Hence, Bayesian inference is the statistically optimal way of combining sensory input with prior beliefs (for a review, see Aitchison & Lengyel, 2017; de Lange et al., 2018; Tenenbaum et al., 2011).

The mirror example can also be described in Bayesian terminology (see equation 1). By multiplying the likelihood distribution (i.e., what you see given the imprecise sensory input) and the prior (i.e., your expectations about your facial features), the posterior distribution gives you an accurate perception of your face.

$$p(\mathrm{s}|x_{mirror}) = \frac{p(x_{mirror}|s)p(s)}{p(x_{mirror})}$$

Equation 1: Mirror example in Bayesian terminology (adapted from Ma, 2019). The posterior probability distribution $p(\mathrm{s}|x_{mirror})$ for the true appearance s of your face given your observation in the mirror (x_{mirror}) is the product of the likelihood distribution $p(x_{mirror}|s)$ and the prior distribution $p(s)$, normalized by a constant factor $p(x_{mirror})$.

Several other computational models can describe perception, such as predictive coding (Schröger & Roeber, 2021) that is closely related to the Bayesian inference theory or pattern theory (de Lange et al., 2018). While a taxonomy of these computational models is beyond the scope of this thesis (but see Aitchison & Lengyel, 2017; de Lange et al., 2018), it is worth mentioning that alternative models exist. There are several reasons to choose a Bayesian approach, though. First, Bayesian inference is not new to psychological research (Etz & Vandekerckhove, 2018). In fact, it has already been introduced in the past century as an alternative to classical significance tests (Edwards et al., 1963; Edwards, 1995). Thus, although its application in perception is relatively new (de Lange et al., 2018; Petzschner et al., 2015), the method itself is already quite well established. Second, as illustrated above, perceptual inference naturally translates into Bayesian inference. This conceptual adjacency is underlined by the fact that Bayesian inference has been good at describing human and animal behaviour in past research (Ma, 2019; Mathys et al., 2011).

To conclude, the perceptual system combines sensory information with prior beliefs about the statistical regularities of the sensory environment to achieve accurate perceptual judgements (Charlton et al., 2022; Krishnamurthy et al., 2017; Nassar et al., 2012; Seriès & Seitz, 2013; Yon & Frith, 2021). This combination can be described with the computational framework of Bayesian inference (Ma, 2019).

1.1.2 Belief updating in dynamic environments

The Bayesian inference framework has been well-tested in static environments where statistical regularities (e.g., most probable sound location) remain stable (de Lange et al., 2018; Knill & Pouget, 2004; Summerfield & Egner, 2009). However, knowledge about how the perceptual organism can achieve high perceptual accuracy in dynamic environments in which the prior quality can readily change is sparser (Filipowicz et al., 2022; Krishnamurthy et al., 2017; Nassar et al., 2012)

Generally, the prior is weighted according to its relevance and reliability (Knill & Pouget, 2004; Krishnamurthy et al., 2017; Petzschner et al., 2015; Ma, 2019). Hence, the more reliable the prior is, the heavier its weight on the posterior (de Lange et al., 2018; Gold et al., 2008; Petzschner et al., 2015; Yon & Frith, 2021), but only if the prior is relevant (Charlton et al., 2022; Krishnamurthy et al., 2017; Lawson et al., 2021). *Prior reliability* (or better: *prior precision*) refers to the inverse variance of the prior distribution (Battaglia et al., 2003; Ma, 2019; Petzschner et al., 2015). The posterior of the current observation turns into the prior of a subsequent observation (de Lange et al., 2018; Knill & Pouget, 2004). Consequently, while perceptual uncertainty is being reduced as we accumulate sensory evidence (Aitchison et al., 2020; Brunyé & Gardony, 2017; Gold et al., 2008; Ma, 2019; Petzschner et al., 2015), the prior becomes increasingly precise as well (Bejjanki et al., 2016; Charlton et al., 2022; Knill & Pouget, 2004; Krishnamurthy et al., 2017).

In *dynamic environments,* statistical regularities of the sensory environment can suddenly change (Alamia et al., 2019; Charlton et al., 2022; Krishnamurthy et al., 2017; Lawson et al., 2021; Yon & Frith, 2021), which renders previously held priors irrelevant (Charlton et al., 2022; Krishnamurthy et al., 2017; Yon & Frith, 2021). *Prior relevance* reflects the probability that the true state of the environment is (still) in line with one's prior beliefs (Charlton et al., 2022; Krishnamurthy et al., 2017). Hence, the observer must update their beliefs according to the sensory input (Achtziger et al., 2014; Berniker et al., 2010; Payzan-LeNestour & Bossaerts, 2011; Zhang et al., 2014). *Belief updating* thus involves ignoring imprecise and/or irrelevant priors in favour of the likelihood (Charlton et al., 2022; Filipowicz et al., 2020; Lawson et al., 2021). By focussing on the likelihood, a new relevant prior will be built that will be rather imprecise at first (Charlton et al., 2022; Krishnamurthy et al., 2017; Yon & Frith, 2021). The prior will become increasingly precise by accumulating consistent sensory evidence (Krishnamurthy et al., 2017). Since a mismatch between the prior and sensory input is generally perceived as surprising (Clark, 2013; de Gee et al., 2021; Sedley et al., 2016; Summerfield & Egner, 2009), *surprisal* can induce belief updating, which makes it a viable indicator of it (Charlton et al., 2022; Filipowicz et al., 2020; Yon & Frith 2021). Surprisal is inversely related to prior relevance and is especially high if the prior that turned out to be irrelevant was particularly precise (Burlingham et al., 2022; Filipowicz et al., 2020; Preuschoff et al., 2011; Urai et al., 2017).

To illustrate this definition of belief updating, consider the mirror example again. Since you see your face almost every day, you probably have exact expectations of how your face looks (i.e., precise prior), and this prior is most often in line with the actual features of your face (i.e., high prior relevance). But now, imagine someone painted a small red dot on your forehead without your knowledge. Suddenly, your prior -specifically, expectations about features of your forehead- does not match the image in the mirror (i.e., low prior relevance), even though this prior is still very precise. This will surprise you, and you will likely ignore the prior in favour of the sensory input. In other words, your posterior will be primarily based on the imprecise sensory input. Since the posterior of a previous observation turns into the prior of a subsequent observation (de Lange et al., 2018; Knill & Pouget, 2004), you will build a new relevant prior based on the sensory input. This prior will then include the red dot. However, since you have seen the dot just once, you probably do not know its exact features (e.g., diameter, position). Due to this undersampling, this new prior will be imprecise at first (Krishnamurthy et al., 2017), and you will still weight the sensory input more heavily than this new prior. However, the more often you observe the red dot, the more precise your prior will become. Consequently, the prior will have a stronger influence on your perceptual judgement. Of course, once the dot washes off, you will once again adapt your prior by excluding the red dot as a facial feature.

In conclusion, priors are weighted by their precision and relevance. When priors become irrelevant (which is surprising), the posterior will be based mainly on the likelihood, which will lead to the building of a new and relevant prior (Yon & Frith, 2021; Krishnamurthy et al., 2017). As sensory evidence is being accumulated, this prior will become increasingly precise (Aitchison et al., 2020; Ma, 2019; Petzschner et al., 2015). In return, it will exert an increasingly strong influence on perception (Charlton et al., 2022; Krishnamurthy et al., 2017).

1.2 Neurophysiological mechanisms behind belief updating

Since the Bayesian framework of belief updating has been used effectively in describing human and animal behaviour (Ma, 2019; Mathys et al., 2011), neurophysiological manifestations should exist (Krishnamurthy et al., 2017). Therefore, the next logical question is: What neurophysiological mechanisms drive belief updating?

The past few decades of research suggest that the *locus coeruleus noradrenaline* (LC-NA) arousal system mediates cognitive and behavioural adaptation in dynamic environments (Bouret & Sara, 2005; Devauges & Sara, 1990; Yu & Dayan, 2005). According to the *adaptive gain theory*, the LC-NA system balances the exploiting of well-established choices and exploring new alternatives (Aston-Jones & Cohen, 2005), which is crucial for optimal behaviour in dynamic environments (Jepma & Nieuwenhuis, 2011). This is conceptually adjacent to belief updating, which allows accurate perceptual judgements despite unexpected changes in the statistical regularities of the sensory environment (Charlton et al., 2022; Filipowicz et al., 2020; Lawson et al., 2021). Hence, it could be that the LC-NA system mediates belief updating. A few studies have tested this notion using pupil diameter as a proxy of the LC-NA system (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2021; Nassar et al., 2012).

1.2.1 The locus coeruleus noradrenaline system as a neural mechanism behind optimal behaviour in dynamic environments

The idea that the LC-NA system drives cognitive and behavioural adaptation is not new. Devauges and Sara (1990) already showed over three decades ago that injecting the LC-boosting drug idazoxan in rats improved cognitive flexibility in learning the spatial configuration of a maze. According to the adaptive gain theory popularized in the early 2000s, the LC-NA system balances the trade-off between the facilitation and disengagement of task-specific processes by regulating sensory processing (for an extensive review, see Aston-Jones & Cohen, 2005). This, in turn, enables behavioural optimization and adaptation. In line with this, Jepma and Nieuwenhuis (2011) showed that LC-NA activity reflected exploration in a task where participants must switch back and forth between four slot-machines to maximize profit.

Bouret and Sara (2005) suggest that the LC-NA system achieves this balance of exploiting well-known options and exploring new ones by activating in response to unexpected changes in the environment, possibly by reorganizing existing functional neural networks. This is corroborated by Joshi et al. (2016) who found that the LC modulates neural gain by releasing NA and marking surprisal in rhesus monkeys.

In conclusion, these findings suggest that the LC-NA system enables the processing of and learning from unexpected and novel events by suppressing behavioural and cognitive habits (Yu & Dayan, 2005) in favour of sensory evidence (Krishnamurthy et al., 2017), possibly via reorganizing existing functional networks in the brain (Bouret & Sara, 2005).

1.2.2 Contributions of the locus coeruleus noradrenaline system to belief updating

To reiterate, belief updating involves the suppression of existing priors in favour of sensory input due to sudden changes in the statistical regularities of the environment. In Bayesian terms, the prior must be weighted according to its precision and relevance (Charlton et al., 2022; Filipowicz et al., 2020; Lawson et al., 2021). A mismatch between prior beliefs and sensory evidence elicits surprisal (Clark, 2013; de Gee et al., 2021; Yon & Frith, 2021), which leads to a state of heightened arousal induced by the LC through the release of NA (de Gee et al., 2021, de Gee et al., 2014). Since surprisal can induce belief updating (Charlton et al., 2022; Filipowicz et al., 2020; Yon & Frith 2021), the LC-NA system seems to mediate belief updating (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2017).

In line with this notion, Yon and Frith (2021) argue that the weighting of priors according to their precision and relevance is achieved through the release of NA. Lawson et al. (2021) also corroborate this notion by showing that administering the NA antagonist propranolol in humans leads to slower belief updating in response to changes in the probabilistic relationships between a cue sound and a target image. Furthermore, this inefficient belief updating resulted in higher surprisal. In other words, surprisal is induced by the LC-NA system (Bouret & Sara, 2005), which signals a mismatch between prior and sensory evidence. This results in ignoring the prior in favour of sensory input (Krishnamurthy et al., 2017; Ma. 2019).

These findings suggest that the LC-NA system allows for the adjustment of prior use in perceptual judgements according to its relevance and precision (Krishnamurthy et al., 2017; Lawson et al., 2021; Preuschoff et al., 2011). When a prior turns out to be irrelevant and/or imprecise, the LC-NA system reduces the prior's influence on the posterior in favour of the likelihood (Krishnamurthy et al., 2017; Ma, 2019; Lawson et al., 2021). Hence, the LC-NA system mediates belief updating by inducing a state of heightened arousal in response to surprising changes in the statistical regularities of the sensory environment (Krishnamurthy et al., 2017; Preuschoff et al., 2011; Yon & Frith, 2021).

1.2.3 Pupil diameter as an indicator of arousal-mediated belief updating

The LC can also drive pupil dilation by releasing NA (Aston-Jones & Cohen, 2005; Breton-Provencher & Sur, 2019; de Gee et al., 2021; Joshi et al., 2016; Strauch et al., 2022). Since the LC-NA system seems to mediate belief updating, pupil diameter offers a non-invasive and economical tool for assessing the contributions of the arousal system in belief updating (de Gee et al., 2014; Joshi et al., 2016; Joshi & Gold, 2020; Krishnamurthy et al., 2017; Lawson et al., 2021).

Pupil diameter has been used as a neurophysiological readout of cognitive and neural processes since the beginning of psychological and neuroscientific research (de Gee et al., 2021; Denison et al., 2020; Joshi & Gold, 2020). Strauch et al. (2022) identified different factors and neural pathways that drive pupil dilation, including the LC-NA system. This notion is backed up by many studies that found a tight link between pupil diameter and the LC-NA system (for an extensive review, see Aston-Jones & Cohen, 2005; Joshi & Gold, 2020). For instance, it has been shown that blood-oxygen-level (BOLD) activity of the LC-NA system in humans covaries with pupil diameter (Murphy et al., 2014). In line with this, the administration of the NA antagonist propranolol diminishes otherwise observed pupil dilation in response to sudden changes in the statistical regularities of the environment (Lawson et al., 2021). Furthermore, electrical microstimulation of the LC-NA system evokes pupil dilation in monkeys (Joshi et al., 2016) and mice (Reimer et al., 2016). These findings suggest that pupil diameter can be used as an indirect and non-invasive measurement of LC-NA system activity (Joshi & Gold, 2020). For this reason, the LC-NA arousal system will also be referred to as the *pupil-linked arousal system* from here on out.

Consequently, pupil diameter should indicate the contributions of the pupil-linked arousal system in belief updating. The pupil has been shown to dilate during perceptual decision-making (Lempert et al., 2015; De Gee et al., 2014), decision uncertainty (Lempert et al., 2015; Preuschoff et al., 2011; Satterthaite et al., 2007), and surprisal (Alamia et al., 2019; de Gee et al., 2021; Filipowicz et al., 2020). In line with the adaptive gain theory of the LC-NA system (Aston-Jones & Cohen, 2005), pupil dilation can predict exploratory choices (Jepma & Nieuwenhuis, 2011). Furthermore, Nassar et al. (2012) found that pupil dilation is related to learning rate and prediction error. Similarly, Urai et al. (2017) found that pupil dilation reduces the tendency to repeat perceptual choices in response to increased reliability of sensory input. This suggests that the pupil-linked arousal system mediates belief updating (Filipowicz et al., 2020).

Krishnamurthy et al. (2017) provide direct support for this notion. In their task, participants had to predict the location of the final sound of a sequence (*probe*). Participants were able to extrapolate the average location from the sound sequence to build spatial priors. These spatial priors, in turn, helped them predict the probe location. However, sudden changes in the average sound location decreased prior relevance, which was reflected by pupil dilation. Critically, pupil dilation was also associated with the reduced influence of spatial priors in the location prediction. This corroborates the assumption that the pupil-linked arousal system mediates belief updating.

In conclusion, belief updating is regulated by the pupil-linked arousal system (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012). This means that pupil dilation predicts a reduction of prior use in favour of sensory input in response to sudden changes in the statistical regularities of the environment (Filipowicz et al., 2020; Lawson et al., 2021; Urai et al., 2017). Specifically, pupil diameter seems to reflect the use of priors in accordance with their relevance and precision (Charlton et al., 2022; Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012).

1.3 The changepoint paradigm of sound localization

Belief updating is commonly examined using *change point paradigms* (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Ma, 2019; Nassar et al., 2012). In those paradigms, changes in the underlying statistical regularities of stimuli (*change points*) occur with a certain probability (*hazard* rate) set by the experimenter. The main assumption is that participants would adjust the use of priors extrapolated from previous stimulus observations (Knill & Pouget, 2004) following the occurrence of a change point. A change point renders the prior irrelevant, thereby decreasing its influence on perception

(Krishnamurthy et al., 2017; Yon & Frith, 2021). By simultaneously measuring pupil size, studies have examined the role of the pupil-linked arousal system in this belief updating process (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012). In our study, we focused on arousal-mediated belief updating in spatial localization using the change point paradigm. Specifically, we used an adapted version of the dynamic sound localization task of Krishnamurthy et al. (2017). While their design shed light on many aspects of arousal-mediated belief updating in spatial localization, it still left open some questions we wanted to address.

Krishnamurthy et al. (2017) presented participants with a sequence of audio-visual stimuli along the horizontal axis. The sequence would stop randomly before the final stimulus (*probe*) occurred, and participants must then predict where the probe will most likely come from. Krishnamurthy et al. (2017) assumed that participants would build a spatial prior by extrapolating the mean stimulus location (*generative mean*) from previous observations. Participants should then base their prediction of the probe location on this spatial prior. Of course, the more stimuli were sampled from this generative mean, the more precise the prior would become (Aitchison et al., 2020; Krishnamurthy et al., 2017). However, sudden changes in the generative mean (change points) rendered spatial priors based on the previous generative mean irrelevant, which decreased their influence on the probe location prediction. Crucially, Krishnamurthy et al. (2017) observed pupil dilation in response to irrelevant caused by changepoints. This corroborates the notion that the pupil-linked arousal system signals sudden changes in the statistical regularities of the sensory environment (Lawson et al., 2021; Nassar et al., 2012; Yon & Frith, 2021), thereby mediating belief updating.

However, Krishnamurthy et al. (2017) examined this arousal-mediated belief updating that required explicit perceptual judgements. Specifically, participants were required to predict the probe location before it was presented. Priors had to be built to make predictions as there was no sensory information of the probe to rely on. Consequently, Krishnamurthy et al. (2017) left open the question of whether this arousal-mediated belief updating could also be observed in more conventional situations on an implicit and automatic level.

Furthermore, Krishnamurthy et al. (2017) used audio-visual stimuli for their task. In general, bimodal integration improves the accuracy of spatial localization (Alais & Burr, 2014; Cai et al., 2018), thereby reducing perceptual uncertainty (Bejjanki et al., 2016).

Hence, Krishnamurthy et al. (2017) did not address whether arousal-mediated belief updating is exclusive to the visual domain or is also present for spatial priors based on purely auditory stimuli.

1.4 Our research questions and hypotheses

Empirical research about arousal-mediated belief updating in spatial localization is currently inconclusive. Studies so far have primarily examined the relation between pupil diameter and the use of spatial priors using visual (Charlton et al., 2022; de Gee et al., 2021) or audio-visual stimuli (Krishnamurthy et al., 2017). To our knowledge, arousal-mediated belief updating in spatial localization has not been examined yet using purely auditory stimuli. Since spatial priors built on visual stimuli are generally more precise than those built on auditory stimuli (Yon & Frith, 2021), it is unclear whether the relationship between pupil diameter and belief updating is visually dominated. Furthermore, most studies explicitly required the use of priors when making predictions (Filipowicz et al., 2020; Krishnamurthy., 2017; Nassar et al., 2012). Hence, it remains unclear whether belief updating could also be present on an automatic level before explicit behavioural responses. If so, does the pupil-linked arousal system also contribute to this automatic and implicit weighting of spatial priors?

We seek to close this gap in the literature, thereby expanding our knowledge about the role of the pupil-linked arousal system in spatial perceptual decision-making. To this end, we devised a dynamic sound localization task inspired by the one described in Krishnamurthy et al. (2017). Crucially, however, we did not require participants to predict the probe position. Instead, participants were only required to *estimate* the probe location *after* it was presented. Since participants could, in theory, rely exclusively on sensory input to make their estimates, we omitted the explicit encouragement to use and update spatial priors. Under the assumption that we automatically use priors in combination with sensory input to improve perceptual judgements (Yon & Frith, 2021), we expected to find link between the pupil-linked arousal system and belief updating using this approach. To shed light on whether arousal-mediated belief updating is exclusive to the visual domain, we introduced an audio-visual (AV) and auditory (A) within-subject condition.

Specifically, we expected to find a positive relation between surprisal and pupil diameter since surprisal can induce belief updating (Charlton et al., 2022; Filipowicz et al., 2020; Yon & Frith, 2021), and pupil dilation signals surprisal (Dayan & Yu, 2006;

Filipowicz et al., 2020; Nassar et al., 2012). This would show that arousal-mediated belief updating in spatial localization constitutes an implicit perceptual process. Furthermore, we expected this link to be present in both the AV and the A conditions. This would corroborate the notion that arousal-mediated belief updating is -at least in spatial localization- not exclusive to the visual domain but supramodal.

To summarize, we examined whether we can find evidence for automatic belief updating that is mediated by the pupil-linked arousal system. Furthermore, we wanted to address whether arousal-mediated belief updating in spatial perception is exclusive to the visual domain or supramodal. Taken together, our study aimed at an important expansion to our understanding of the contributions of the pupil-linked arousal system in spatial perceptual decision-making.

2. Methods

2.1 Participants

We invited 32 participants (15 female) via the Vienna Cognitive Science Hub to participate in our study at the Acoustics Research Institute of the Austrian Academy of Sciences. Participants received monetary compensation of ten Euros per hour. We chose this sample size based on Krishnamurthy et al. (2017), who found a significant relation between pupil diameter and indicators of belief updating (prior relevance and precision) using a similar sample size. Participants were 18-35 years old (*M* = 24.4 years, *SD* = 3.6), spoke German or English, and were right-handed. Participants had no history of neurological or psychiatric disorders (e.g., claustrophobia or epilepsy) or excessive drug use (including alcohol/cigarettes) at the time of testing or the day before. Participants had no impaired hearing and normal or corrected-to-normal eyesight. We excluded participants wearing glasses (not contact lenses) since glasses have caused problems with our pupillometry setup in the past. Participants were naïve to the purpose of the study and gave written consent before the experiment. After the experiment, we debriefed them in detail about the purpose of the study.

2.2 Experimental setup

The experiment was run in MATLAB using the Psychophysics toolbox extensions (Kleiner et al., 2007). It was presented on an LCD monitor (48 x 27 cm) with a refresh rate of 60 Hz. The operating system was Windows 10. Participants sat in a dimly lit sound-attenuated room and rested their heads on a chin rest 75 cm away from the monitor. During the experiment, participants wore earphones (ER2; Etymotic Research, Elk Grove Village, IL) while we monitored pupil gaze and pupil diameter using an eye tracker (Eye-Link 1000 Plus; SR Research, Osgoode, Ontario, Canada). Crucially, we kept luminance levels and focal distance between the participant and the computer screen constant to avoid confounding the pupillometry (Joshi et al., 2016; Nassar et al., 2012; Strauch et al., 2022). We also measured brain activity data using an electroencephalogram (EEG) with 128 active electrodes (actiCAP with actiCHamp; Brain Products GmbH, Gilching, Germany) for a separate study.

2.3 Procedure

Our study consisted of ten training blocks and seven main task blocks that took about 10 hours in total, split between two days with no more than 72 hours in between. The only exception was one participant returning to the second session 12 days later. However, they did two extra training blocks on the second session to avoid training effects during the main task. On the first day, participants did eight training blocks and one main task block. Each block consisted of 50 trials. We only collected EEG and pupillometry data on the second day. Participants were free to choose breaks between the blocks in the first session. In the second session (consisting of two training blocks and six main task blocks), there were two mandatory breaks of at least 15 minutes, one after the training blocks and one after the third experimental block. In both tasks, participants were required to estimate the location of a target sound. We only used the six main task blocks of the second session for further analysis.

2.4.1 Training task

The training task is illustrated in Figure 1. At the beginning of each trial, participants were asked to fixate a black dot with 2 pixels (px) in diameter on the centre of the screen. After 750-1000 ms, participants heard a single 50 ms pink noise burst with 10 ms on/off raised cosine ramps, high pass filtered with a 4^{th} order Butterworth at 250 hertz (Hz) cut-off. The sound was presented at a 48 kHz sampling rate at 80 decibel (dB) instantaneous peak sound pressure level (65 dB short-time averaged with a "fast" time constant of 125 ms; Bruel & Kjaer 2260 investigator sound level meter). The sounds were spatialized by convolution with the participant's individually measured head-related impulse response (HRIR) to ensure spatial immersion. The azimuth of each sound was sampled from a Gaussian distribution with an *SD* of 10° (*experimental* noise) and a *generative mean* taken from a uniform distribution between -60° and +60°. The elevation was fixed to 0°. To minimize data loss of the pupillometry, participants were asked to keep fixating the dot and minimize blinking during this entire period. The possible sound locations were visually represented to the participant by a black semi-circular arc (width: 2 px) on the computer screen with a radius of 0.75° visual angle around the centre of the screen. 950 ms after the sound was presented, the fixation dot would double in size (diameter: 4 px), indicating that the participants were now required to estimate the position of the sound. By moving a computer mouse, the fixation dot would turn into a black response line (width: 2px) that originated from the centre of the screen and crossed the semi-circular arc. Its outermost point was at a 0.95° visual angle. Participants moved this response line along the semi-circular arc to indicate their location estimate. After confirming their estimate by left-clicking, participants indicated their estimation confidence by adjusting the width of an area around the response line to capture the true sound location in about 80% of all trials. To make the area wider (i.e., low estimation confidence), participants moved the mouse away from their initial response line, and vice versa to make it narrower (i.e., high estimation confidence). During the estimation and confidence response, participants were no longer required to remain still and were told they could take as much time as needed for their response. After they confirmed their confidence response, the response line within the area appeared in green. Participants would then receive feedback about the actual location of the sound, represented by a red line (width: 2 px) that ran from 0.55° to 0.95° visual angle in a straight line from the centre of the screen that crossed the semi-circular arc. Afterwards, a new trial would start where participants were again required to fixate on the dot on the screen and minimize eye movements until the estimation and confidence responses were prompted. Crucially, we distinguished between two within-subject conditions that alternated at the end of each trial. In the audio-visual (AV) condition, the sound would be accompanied by a visual representation of its spatial origin, consisting of a black line (width: 2 px) that ran from 0.55° to 0.95° visual angle. On the other hand, the auditory (A) condition did not provide any visual aid during the sequence.

At the end of a training block, participants were shown their *mean absolute localization error* (*MAE*; i.e., mean degree difference between all estimations and true sound locations) and the mean percent of how often their confidence area captured the true sound location (*capture rate*). Two MAEs (one for each 25 AV and 25 A trials) and two capture rates were then presented. If an MAE was above 25°, the experimenter encouraged participants to be more accurate. Furthermore, if one of the capture rates was far off from 80%, the experimenter asked the participant to adjust the width of the confidence area accordingly. Specifically, participants were asked to narrow the confidence area if the capture rate was far above 80% and to make it broader if it was far below 80%. Participants who failed to reach MAEs below 25° for both AV and A in at least one training block during the first session were not eligible for the second session and were thus excluded from our analyses.

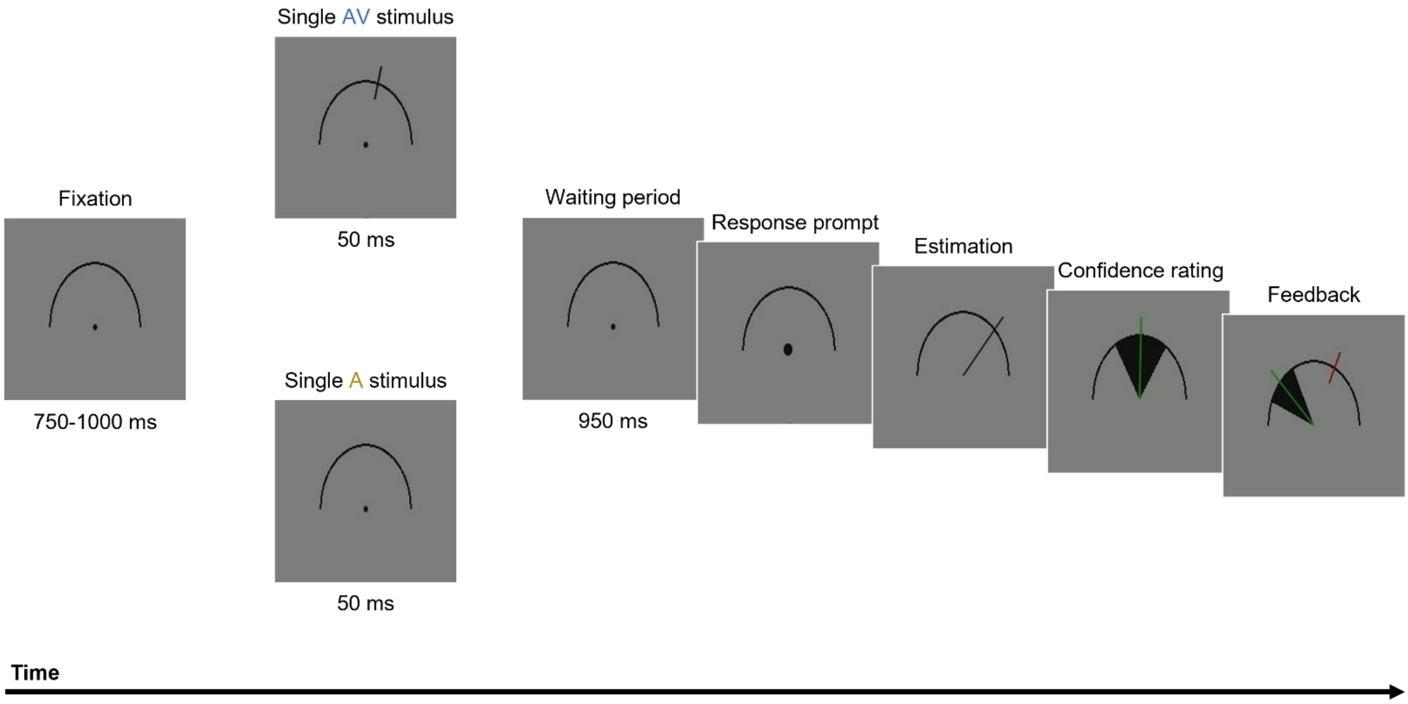

Figure 1. Sequence of a training trial. After fixating a dot on the computer screen for 750 ms, a sound was presented that was located somewhere along the semi-circular arc on the screen. In the audio-visual (AV) condition, the sound's location was visually indicated by a line that crossed the arc. 950 ms after the sound, the fixation dot doubled in size, prompting the participants to estimate the sound's location by moving a response line along the arc. After confirming their estimate, participants were required to indicate their estimation confidence by adjusting the width of an area around the response line. Finally, participants were informed about the true location, indicated by a red line.

2.4.2 Main task

Figure 2 visually illustrates the procedure of our main task. It uses the same stimuli as described in the training task. A trial would start with the participant fixating on the dot at the centre of the screen. After 750-1000 ms, the stimuli would be presented. In the main task, each trial usually consisted of a sequence of sounds rather than a single sound. There was an interstimulus interval of 450 ms (i.e., 500 ms stimulus onset asynchrony). As during training, the azimuth of each sound was sampled from a Gaussian distribution with an *SD* of 10° (*experimental noise*) and a generative mean taken from a uniform distribution between -60° and +60°. The elevation was fixed to 0°. Crucially, there was a one in six chance (*hazard* rate) for a *change point*. A change point meant that the previous generative mean was discarded and was again randomly drawn from the uniform distribution. Every sound had a one in 12 chance of being the final sound of the sequence (*probe*). Hence, single-sound trials (*localizer* trials) occasionally occurred. 1000 ms after the probe has been presented, the fixation dot would again double in size, indicating the requirement to estimate the probe location. Again, participants moved the response line along the semi-circular arc with a computer mouse to give their estimate. Since any sound of the sequence could be the probe, participants were strongly encouraged to pay attention throughout the entire sequence. After confirming their estimation, participants would again indicate their estimation confidence by adjusting the width of the area around the response line. Like in the training task, participants were free to take as long as they liked for their estimation and confidence response. Furthermore, they were free to move their eyes and blink as much as they wanted during the estimation and confidence response until the beginning of the subsequent trial. Importantly, participants were not given feedback about the true probe location at the end of a trial. Furthermore, we again distinguished between AV and A condition for all sounds preceding the probe. Hence in the AV condition, the location of all sounds *but* the probe was again indicated by a black line that crosses the semi-circular arc. In the A condition, on the other hand, we omitted visual aid. Crucially, the probe was purely auditory in both conditions. AV and A condition alternated each trial (i.e., n = 25 trials per condition in each block).

At the end of a main task block, participants were again informed about their AV and A MAEs and capture rates. Like in training, participants were encouraged to be more precise if either of the two MAEs was above 25°. They were also asked to adjust the width of their confidence areas if the capture rates were far above or below 80%.

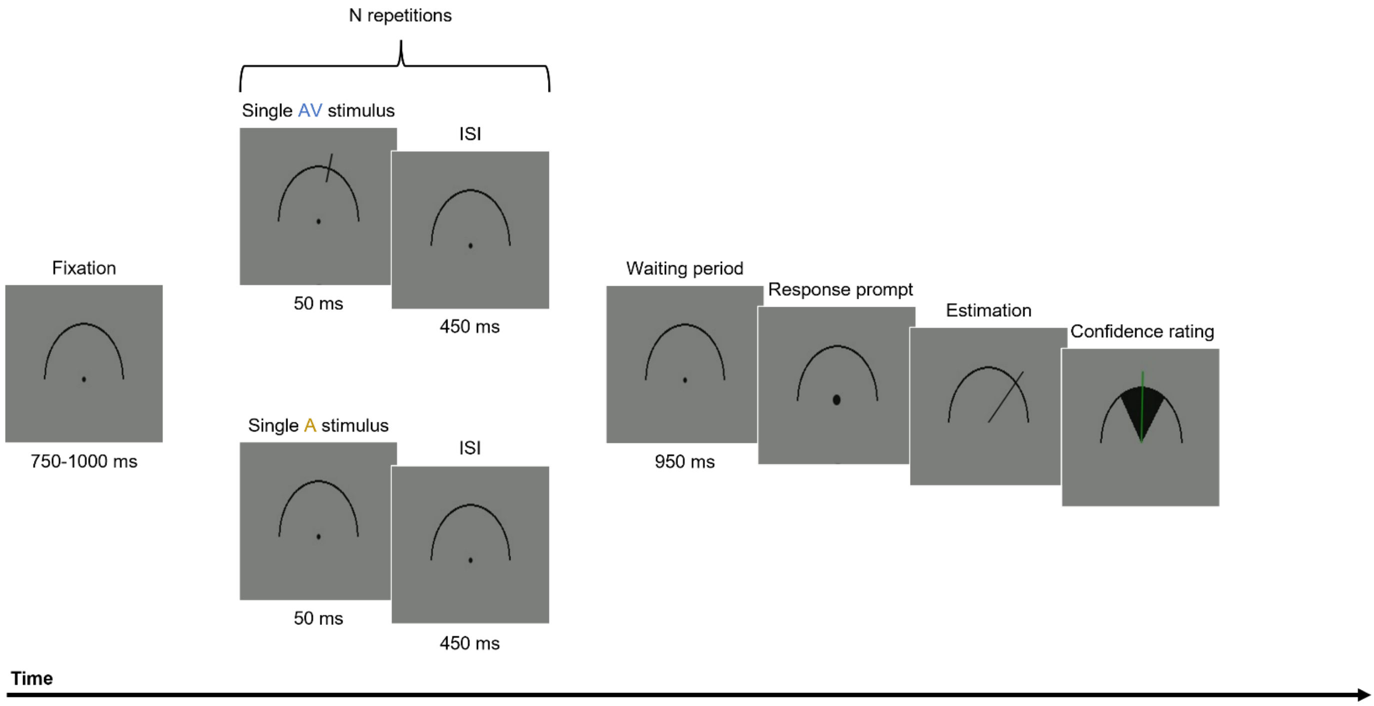

Figure 2. Sequence of a main task trial. A sequence of sounds (12 on average) with an interstimulus interval (ISI) of 450 ms was presented. In the audio-visual (AV) condition, every sound but the last one was accompanied by a visual indication of its location. In the auditory (A) condition, visual aid was omitted altogether. Participants must estimate the location of the final sound (probe) and indicate their estimation confidence as they did in the training task.

2.4 Outcome measures

We extracted two variables directly from the data of our six main blocks, separated by condition (AV and A). First, we extracted the MAE across all trials with more than one sound. The MAE reflects the mean absolute localization error and indicates participants' overall localization performance. To reduce intraindividual differences, we subtracted the MAE across all localizer trials from the mean non-localizer MAE for each participant, separated by condition (AV and A). Furthermore, we categorized each sound of every trial in terms of sound after change point (SAC; Krishnamurthy et al., 2017) levels by counting the number of sounds being presented since the last change point. For example, an SAC level of five indicated a change point five sounds before. Since change points render prior beliefs irrelevant and newly built priors increase in precision as consistent sensory evidence is being accumulated (Krishnamurthy et al., 2017), SAC levels indicated prior relevance and precision (Krishnamurthy et al., 2017). We assumed that MAE would decrease as a function of SAC level, implying that participants efficiently used their spatial priors by weighting them accordingly (Krishnamurthy et al., 2017).

2.5 Surprisal: Latent variable from a Bayesian inference model

Surprisal is defined as the negative log probability of a sensory event given all prior knowledge (i.e., the probability comes from the prior distribution; Alamia et al., 2019; Filipowicz et al., 2020; Knill & Pouget, 2004; Sedley et al., 2016). We extracted surprisal as a latent variable from a Bayesian inference model for each stimulus of every trial of the six main task blocks from the second session.

We used the Bayesian inference model described in Barumerli et al. (2022) to extract surprisal. This model is similar to the reduced model by Nassar et al. (2010), which Krishnamurthy et al. (2017) used but offers several improvements. It describes the statistically optimal way of perceptual decision-making in our change point paradigm. Similar to the online change point detector described in Adams and MacKay (2007), our model estimates the subjective probability of a change point having occurred for every stimulus (*subjective hazard rate*) based on the behavioural responses of our participants. The model then computes two conditional posterior distributions: The first assumes that a change point has occurred, which would require ignoring the spatial prior based on the previous generative mean. The other posterior distribution assumes no change point has occurred, so the prior would still be relevant (Krishnamurthy et al., 2017). To accurately estimate the location of a subsequent stimulus, the model then combines both conditional posterior distributions as a weighted average (*full posterior distribution*). The weights of the conditional posteriors depend on the subjective hazard rate. Hence, the full posterior distribution can flexibly change in accordance with the probability of a change point and thereby flexibly update its beliefs.

Importantly, the model consists of a set of parameters and latent variables. Using the MATLAB toolbox BADS described in Acerbi and Ma (2017), these parameters and latent variables were fitted to the behavioural responses of each stimulus of every trial via maximum log-likelihood estimation. Using this approach, we extracted surprisal as a latent variable (among other latent variables beyond the scope of this paper). We logarithmized surprisal to normalize the dataset. Afterwards, we z-scored surprisal separately for condition (AV and A). Furthermore, we distinguished between non-probe (*ongoing*) and *probe* surprisal.

2.6 Pupillometry

We analyzed the pupil trace of each trial of the six main task blocks (*n* = 300 trials) from the second session for each participant, separated by condition (i.e., *n* = 150 trials per condition). After pre-processing, we fitted a deconvolution-based model to obtain a pupil gain parameter for every sound in a sequence. Pupil gain is a latent variable that indicates event-related arousal (Denison et al., 2020). The steps of the pre-processing and the model fitting are described separately below.

2.6.1 Pre-processing

We first identified blinks of the pupil trace from the beginning of a trial until the response prompt. For this, we first used EyeLink's pupil annotation algorithm that marks time points during which the pupil diameter is 0 mm as blinks if the time point is preceded by a gaze downward (i.e., closing the eyelid) and followed by a gaze upward (i.e., opening the eyelid). Any time point where the pupil diameter was 0 mm but not preceded and followed by pupil gaze was considered missing data due to technical difficulties. We then added the criterium that the annotated blink event must last between 50-500 ms to be considered an actual blink, as event durations above or below likely reflected technical difficulties instead. Consequently, annotated blink events that did not fall within this duration range were also considered missing data. We then interpolated blinks and missing data using the cubic-spline function described in Mathôt (2013) to reconstruct the pupil trace without blinking and missing data events. Crucially, no more than 20% of the pupil trace during the trial must contain blinks and/or missing data. If a trial exceeded this threshold, it was excluded from further analyses to avoid creating too much artificial data. Furthermore, in the AV condition, if a blink event overlapped with the presentation of the audio-visual stimulus, we excluded that trial, as we could no longer assume the participant processed the audio-visual stimuli. We rejected a mean of 14.4% of all AV trials across participants (*SD* = 13.0%; range = 0% - 57.3%). Rejections were similar for A trials (*M* = 14.1%; *SD* = 13.4%; range = 0% - 55.3%). We then excluded participants who had more than 20% of the trials in both conditions rejected. This led to the exclusion of eight participants, leaving us with a final sample size of *N* = 24 participants for our main analyses. For our final sample, we excluded a mean of 8.6% of AV trials (*SD* = 6.2%; range = 0% - 19.3%) and a similar amount of A trials (*M* = 8.1%; *SD* = 6.5%; range = 0% - 19.3%).

2.6.2 Deconvolution-based linear model

Due to the slow response dynamics of the pupil size (around 2 s; Denison et al., 2020) and since we had a fixed interstimulus interval (ISI) of just 500 ms, we used the MATLAB *Pupil Response Estimation Toolbox* (PRET; Version 1.0) as described in Denison et al., 2020. The PRET fits a deconvolution-based linear model to the continuous pupil traces (Denison et al., 2020). Specifically, we used delta functions that were time-aligned to the stimulus onsets and (Denison et al., 2020) convolved these with the *Pupil Response Function* (PuRF) as described in Denison et al. (2020). The weighted sum of these convolutions was then fitted to the recorded pupil traces for each stimulus of all trials by minimizing the sum of the squared errors. Importantly, we used the default settings for the PuRF (Hoeks & Levelt, 1993). Thus, the only free model parameters were the amplitude of the delta function and the y-intercept (Denison et al., 2020). *Delta amplitude* indicates the percentual increase in pupil size attributable to stimulus-evoked arousal (Denison et al., 2020) and can also be referred to as pupil gain (Burlingham et al., 2022). Positive pupil gain values indicate arousal-related pupil dilation, and negative values indicate pupil constriction (Denison et al., 2020). We fitted the pupil gain for each stimulus of every trial. *y-intercept* reflects a shift in the baseline pupil diameter (Denison et al., 2020). Since pupil diameter at the beginning of a trial could vary between trials, this parameter served as baseline normalization in our analyses. Consequently, the y-intercept was fitted once per trial.

In conclusion, we extracted delta amplitude as a latent variable from a deconvolution-based linear model to fit each trial’s recorded pupil trace of each stimulus of every trial. Delta amplitude reflects pupil gain and indicates task-evoked arousal (Burlingham et al., 2022; Denison et al., 2020). We z-scored pupil gain separately for the two conditions (AV and A) for our main analyses. Furthermore, we distinguished between non-probe (*ongoing*) and *probe* pupil gain.

2.7 Statistical analyses

We conducted our analyses in MATLAB (Version 2022; The MathWorks, Natick, MA) and JASP (Version 0.17.1; JASP Team, 2023). Note that we excluded the first stimulus of every trial in our analyses. Hence, we excluded localizer trials, as these trials consisted only of the probe, which made it impossible to build and use spatial priors.

We first ran four separate two-way repeated measures analyses of variance (A-NOVAs) with the factors modality (AV and A) and SAC 1-6 on MAE, ongoing pupil gain, as well as ongoing and probe surprisal (all variables were averaged across participants). Using a two-tailed significance threshold of $p < .05$, we checked for the main effects of SAC and the interaction between modality and SAC. Note that since both ongoing and probe surprisal and pupil gain were z-scored separately by modality, we did not examine any main effects of modality except for the one on MAE. In case of a significant main effect of SAC, we used posthoc comparisons (Bonferroni corrected) to determine the direction of the effect. Note that using SAC as a factor violated the assumption of sphericity in all four variables of interest. Hence, we applied Greenhouse-Geisser correction for the main effects of SAC and the interactions between modality and SAC. In the first ANOVA, we assessed whether there was a difference in MAE between the two modalities and between SAC 1-6. We expected performance to be better in the AV modality compared to the A modality. Following the Bayesian inference theory, we also expected the beneficial effects of SAC on estimation performance (i.e., negative effect of SAC on MAE). In the second ANOVA, we addressed whether pupil gain tracked the occurrence of a change-point (i.e., negative effect of SAC on pupil gain). This would provide preliminary support for the crucial assumption that the pupil-linked arousal system mediates automatic belief updating in our location estimation task. Since surprisal was extracted from a Bayesian inference model designed specifically for our change point paradigm, the third and fourth ANOVA on ongoing and probe surprisal were conducted to illustrate how the model works. Hence, we expected strong negative effects of SAC on both types of surprisal. Finally, to address whether the effects of SAC on any of these variables were comparable between the two modalities, we examined the interaction between modality and SAC in all four ANOVAs. In case of a significant interaction, we then examined simple main effects to assess whether the effects of SAC were significant in both modality.

Our two research questions were whether we can find automatic belief updating that is mediated by the pupil-linked arousal system and whether belief updating is supramodal. To directly test these questions, we regressed surprisal against pupil gain across all stimuli of all trials with the factors modality (AV and A) and stimulus type (ongoing and probe). Using this 2x2 factor design, we had four regression coefficients for each participant. Since we only had one probe per trial, there were far fewer data points for the probe coefficients than for the ongoing ones. To test whether surprisal significantly predicted pupil gain on a group level, we then ran four separate one-sample *t*-tests with the mean coefficients of each factor. We used Bonferroni correction on the standard significance threshold of $p < .05$. This resulted in a two-tailed significance threshold of $p < .0125$ for our 2x2 factorial design. We expected the link between surprisal and pupil gain to be present throughout all factors, which would corroborate our research questions. Finally, we ran two paired sample *t*-tests to examine whether there was a difference in the relation between surprisal and pupil gain between the two modalities. We ran one paired sample *t*-test for the coefficients of ongoing stimulus sequences and one for the coefficients of probe stimuli. Using this approach, we wanted to assess whether the contributions of the pupil-linked arousal system on belief updating were comparable between the two modalities.

3. Results

3.1 Effects of sound after change point level and modality

3.1.1 Effects of modality and sound after change point level on mean absolute localization error

According to Bayesian inference theory, sound after change point (SAC) levels have a beneficial effect on spatial accuracy. Hence, the mean absolute localization error (MAE) should decrease as a function of SAC, and we tested this with a two-way repeated measures ANOVA. Figure 3 shows the MAE as a function SAC 1-6, separately for AV and A modality, with the standard error of the mean (*SEM*) across participants ($N = 24$). We did not find a significant difference between AV modality ($M = 12.69$, $SEM = 0.57$) and the A modality ($M = 14.54$, $SEM = 0.81$) $F(1, 23) = 2.49$, $p < .13$, $\eta_p^2 = .10$. However, we found a significant main effect of SAC level on MAE, $F(3.05, 70.09) = 10.63$, $p < .001$, $\eta_p^2 = .32$. As can be seen in Figure 3, MAEs were particularly high at SAC1 but sharply fell off thereafter. Posthoc comparisons (see Appendix, Table 1) confirmed this notion by showing that the MAEs of SAC1 were significantly higher than any other SAC level. Hence, as we would expect from Bayesian inference, estimation performance generally benefitted from higher SAC levels. Furthermore, there was a significant interaction between modality and SAC, $F(3.83, 70.08) = 5.35$, $p < .001$, $\eta_p^2 = 0.19$. As Figure 3 shows, there was generally a steeper drop-off of MAE values between SAC levels in AV compared to A modality. Still, simple main effects analysis confirmed that the effect of SAC on MAE was significant in both the AV ($F(6) = 11.23$, $p < .001$) and the A modality ($F(6) = 2.45$, $p = .038$).

In conclusion, estimation accuracy was comparable between AV and A. Accuracy improved as a function of SAC in both modalities. However, this effect was stronger in the AV compared to the A modality.

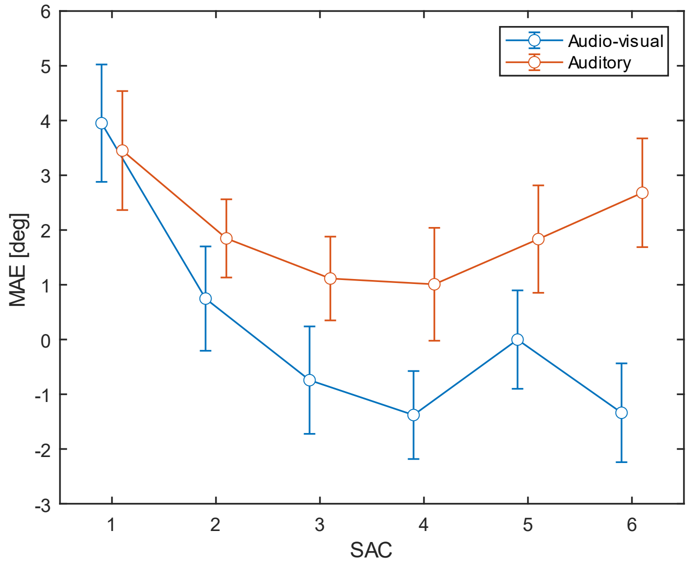

Figure 3. Error bars showing the mean absolute localization error (MAE) as a function of sound after change point (SAC) level across participants. MAE values were subtracted by the localizer MAE separately for audio-visual (blue) and auditory (orange). Positive values indicate higher MAE values compared to localizer trials. The lengths of the error bars reflect the standard error of the mean (*SEM*) per SAC level.

3.1.2 Effects of sound after change point level on pupil gain

If the pupil-link arousal system mediates belief updating, pupil gain should track change points. Hence, pupil gain should decrease as a function of SAC. To test this, we used a two-way repeated measures ANOVA. Figure 4 shows ongoing pupil gain as a function of SAC. There was a significant main effect of SAC level on pupil gain, $F(2.25, 51.77) = 4.24$, $p = .016$, $\eta_p^2 = .16$. As shown in Figure 4, pupil gain was high at SAC1 but fell sharply after that. Hence, pupil size apparently tracked the occurrence of a change point as we expected. Posthoc comparisons (see Appendix, Table 2) partially confirmed this by showing that pupil gain was much higher at SAC1 than at SAC4 and SAC6. Furthermore, there was a (although not significant) trend for higher pupil gain at SAC2 compared to SAC6. Interestingly, there was no significant interaction between modality and

SAC, $F(3.91, 51.77) = 0.80$, $p = .53$, $\eta_p^2 = 0.03$. This suggests that the effect of SAC on ongoing pupil gain was comparable between AV and A modalities.

These results suggest that ongoing pupil gain decreased as a function of SAC. This effect of SAC on ongoing pupil gain was comparable between the two modalities.

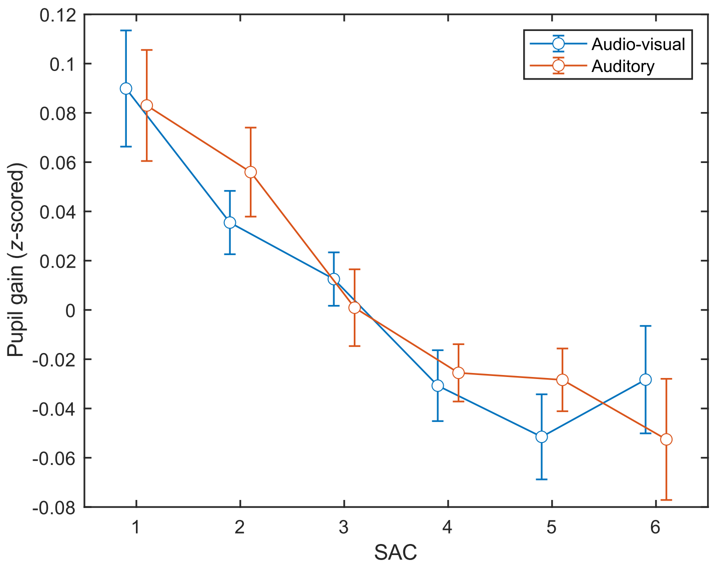

Figure 4. Error bars showing ongoing pupil gain as a function of sound after change point (SAC) level across participants. Pupil gain values were *z*-scored. The lengths of the error bars reflect the standard error of the mean (*SEM*) per SAC level. Plotted separately for the audio-visual (blue) and auditory (orange) modality.

3.1.3 Effects of sound after change point level on surprisal

The Bayesian inference model we used was specifically designed for our change point paradigm. Hence, we expected both ongoing and probe surprisal to decrease drastically as a function of SAC. To test this, we conducted two separate two-way repeated measure ANOVAs. Figure 5 shows the effect of SAV level and modality on surprisal for both ongoing and probe surprisal. There were strong main effects of SAC on both ongoing ($F(1.05, 24.06) = 251.97$, $p < .001$, $\eta_p^2 = .90$) and probe surprisal ($F(1.07, 24.52) = 126.11$, $p < .001$, $\eta_p^2 = .85$). Figure 5 illustrates how our Bayesian inference model based on the change point paradigm worked. Specifically, both ongoing and probe surprisal were much higher at SAC1 and SAC2 but fell off steeply thereafter. Posthoc comparisons (see Appendix, Table 3 and Table 4) corroborated this finding by showing significantly larger ongoing and probe surprisal at SAC1 and SAC2 compared to any other SAC level. Furthermore, probe surprisal was higher at SAC3 than at SAC6. Importantly, there was a significant interaction between modality and SAC in both ongoing (F(1.11, 24.06) = 15.57, p < .001, η_p^2 = 0.40) and probe surprisal (F(1.27, 24.51) = 17.28, p < .001, η_p^2 = 0.43). Simple main effects analyses revealed that the effect of SAC on ongoing surprisal was significant in both the AV (F(5) = 399.87, p < .001) and A modalities (F(5) = 67.74, p < .001). We found similar patterns for probe surprisal in both the AV (F(5) = 158.55, p < .001) and A (F(5) 47.10, p < .001) modalities.

In summary, surprisal decreased sharply as a function of SAC level both for ongoing stimulus sequences and probe stimuli. This is true for both the AV and A modalities, although the effect of SAC was stronger in the AV modality for both ongoing and probe surprisal.

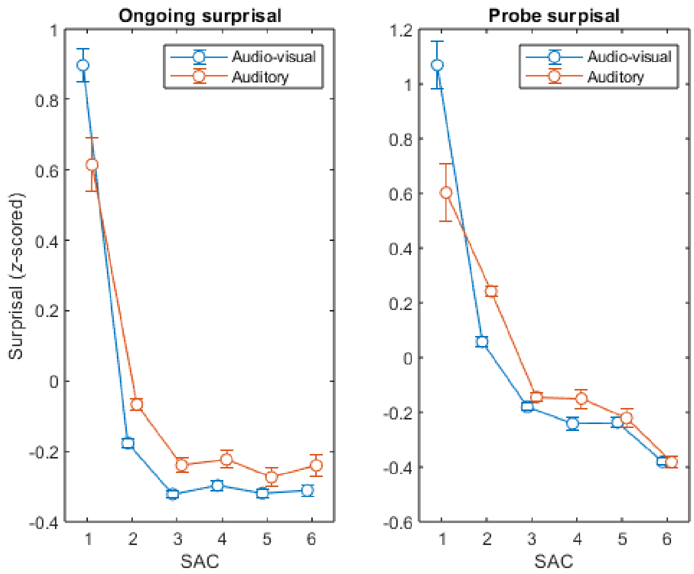

Figure 5. Error bars showing surprisal as a function of sound after change point (SAC) level across participants. The left panel show the effect of SAC on ongoing surprisal, and the right panel shows the effect of SAC on probe surprisal. The lengths of the error bars reflect the standard error of the mean per SAC level. Plotted separately for the audio-visual (blue) and auditory (orange) modality.

3.2 Relation between surprisal and pupil gain

Our main research questions were whether automatic arousal-mediated belief updating is present in our spatial estimation task, and whether this belief updating is supramodal. Hence, we expected a link between surprisal and pupil gain. To address this assumption, we ran a regression of surprisal on pupil gain across all trials for each participant, separated by modality (AV and A) and stimulus type (ongoing and probe). Figure 6 shows the distributions of the four regression coefficients across participants. Crucially, one-sample *t*-tests with a Bonferroni corrected alpha of .0125 revealed that the mean coefficient of the ongoing audio-visual modality (*M* = 0.03, *SEM* = 0.01) was significantly different from zero, $t(23) = 3.60$, $p = .002$, $d = 0.74$. The same was true for the mean coefficient of the ongoing auditory modality (*M* = 0.04, *SEM* = 0.01), t(23) = 4.44, $p < .001$, $d = 0.91$. Interestingly, the mean coefficient of probe audio-visual stimuli (*M* = 0.01, *SEM* = 0.02) was not significantly different from zero, $t(23) = 0.56$, $p = .58$, $d = 0.11$, neither was the mean coefficient of the probe auditory modality (*M* = 0.02, *SEM* = 0.02), $t(23) = 1.26$, $p = .22$, $d = 0.26$. However, the predictive power R^2 of surprisal on pupil gain was generally rather small (range: < 0.00 – 0.03).

Still, the significance of the ongoing mean coefficients suggests that surprisal predicted pupil gain in both the ongoing AV and A modality. Paired *t*-test revealed that this effect was not significantly different between the two modalities, $t(23) = -1.59$, $p = .13$, $d = -0.32$. There was also no significant difference in the mean coefficients between the two probe modalities, $t(23) = -0.50$, $p = .63$, $d = 0.27$.

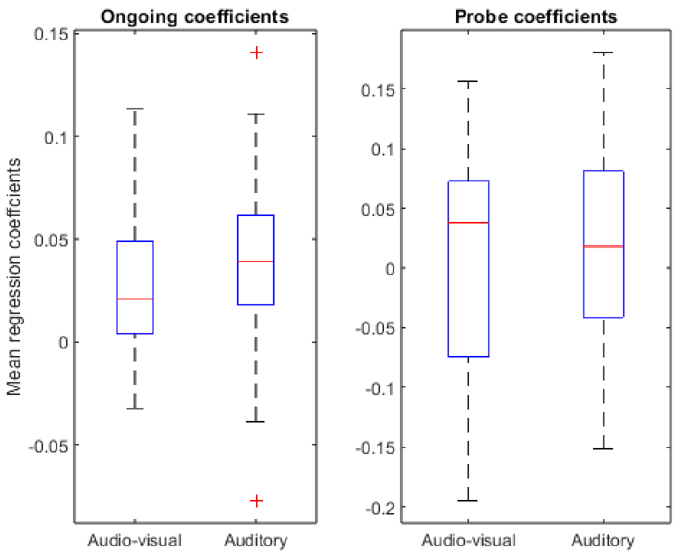

Figure 6. Boxplots showing the mean regression coefficients between surprisal and pupil gain across participants. The left panel show the regression coefficients of ongoing stimulus sequences, the right panel shows the regression coefficients of probe stimuli. Plotted separately for modality (audio-visual and auditory). The red line in the middle of each boxplot reflects the median, the bottom edge the 25th, and the top edge the 75th percentile. The red plus signs represent outliers.

4. Discussion and limitations

We found evidence for arousal-mediated belief updating because pupil gain, an indicator of arousal (Denison et al., 2020), changed in response to the occurrence of a change point and was predicted by surprisal, an indicator of belief updating extracted from a Bayesian inference model (Charlton et al., 2022; Filipowicz et al., 2020; Krishnamurthy et al., 2017). Since our estimation task never explicitly required belief updating, we have shown for the first time that arousal-mediated belief updating during spatial localization does not require conscious effort. Instead, it seems to be an implicit process of perception. Furthermore, we found evidence for arousal-mediated belief updating during purely auditory stimulus sequences. Hence, arousal-mediated belief updating during spatial localization is apparently supramodal.

4.1 Belief updating does not require conscious effort

4.1.1 Estimation accuracy depended on the occurrence of a change point

In contrast to previous studies (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012), we examined arousal-mediated belief updating using an *estimation* task that does not explicitly require the use and updating of spatial priors. Several findings support our assumption that arousal-mediated belief updating is an automatic process that does not require conscious effort and thus constitutes an implicit perceptual process. On a behavioural level, we found that MAE decreased as a function of SAC level. Estimation errors were particularly strong at SAC1 but sharply decreased thereafter. In line with Bayesian inference theory, this finding suggests that participants built spatial priors and used them for their localization by weighting them in accordance with the occurrence of a change point.

However, we normalized MAEs by subtracting them from the MAE across all localizer trials. Hence, positive values in Figure 3 mean worse performance in the corresponding SAC compared to localizer trial performance, and negative values indicate better performance. Thus, even though estimation accuracy generally improved as a function of SAC, overall performance in the A modality was apparently worse than localizer trial performance. One possible reason for this might be the *attentional repulsion effect* (Suzuki & Cavanagh, 1997). It has been shown that briefly presented (< 200 ms) visual stimuli are perceived as displaced from an attended location (Suziki & Cavanagh, 1997). Similar effects have also been observed in spatial hearing (Pecka et al., 2020). Hence, it could be that the repulsion effect caused the probe to be perceived as farther away from the previous stimulus than it actually was in our paradigm. Thus, it seems likely that the repulsion effect impaired the overall performance of non-localizer trials even though performance generally improved as a function of SAC. This is in contrast with the assumptions of the Bayesian inference model that assumes that the perceptual system approximates the generative mean with increasing precision as consistent sensory evidence is accumulated (Krishnamurthy et al., 2017) which should result in an attraction effect. Future studies should examine whether there is a repulsion or attraction effect on the behavioural level. The goodness of fit of the Bayesian model to the behavioural data could also serve as an indicate the presence (or lack thereof) of an attraction effect.

Despite this limitation, Bayesian models have been shown to describe human and animal behaviour accurately (Ma, 2019; Mathys et al., 2011). Hence, even though the performance was worse than in localizer trials, the fact that MAE decreased as a function of SAC still imply that participants made built spatial priors and used them in accordance with the occurrence of a change point. In other words, it appears that belief updating does not require conscious effort. Still, the effects of SAC on MAE were quite inconsistent, as there was no difference between SAC2-6. This merits further research about the perceptual benefits of arousal-mediated belief updating during spatial localization.

4.1.2 Pupil gain tracked the occurrence of a change point

While the effects of SAC on MAE provide some preliminary evidence for the presence of belief updating in our paradigm, it does not address the role of the pupil-linked arousal system in belief updating (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2021). The pupil-linked arousal system, also called the LC-NA system, has been argued to be the neurophysiological mechanism that mediates belief updating (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2021) by inducing a state of increased arousal (Dayan & Yu, 2006; Nassar et al., 2012; Yon & Frith, 2021). Crucially, we found evidence for this notion by showing that pupil dilation tracked the occurrence of a change point. Specifically, we found that pupil gain, an indicator of arousal-induced pupil dilation (Denison et al., 2020), decreased as a function of SAC. Hence, participants efficiently used their spatial priors, resulting in progressively smaller pupil diameters. This provides preliminary evidence that pupil gain is indicative of arousal-mediated belief updating that constitutes an implicit perceptual process and does not require conscious effort.

4.1.3 Surprisal predicted pupil gain in ongoing stimulus sequences

Research suggests that belief updating is triggered by surprising and unexpected changes in the statistical regularities of the environment (Charlton et al., 2022; Filipowicz et al., 2020; Lawson et al., 2021). Hence, surprisal seems to be a viable indicator of belief updating (Charlton et al., 2022; Filipowicz et al., 2020; Yon & Frith, 2021). Since the Bayesian inference model we used was designed specifically for our change point paradigm, it is unsurprising that we found a substantial decrease of surprisal as a function of SAC. This finding corroborates the notion that surprisal signals change points and hence irrelevant priors (Yon & Frith, 2021) and thus confirms surprisal as a viable indicator of belief updating in our paradigm. Specifically, it suggests that participants, in accordance with the Bayesian inference model, made efficient use of their spatial priors in accordance with the occurrence of a change point, thereby effectively reducing surprisal. Importantly, the effects of SAC on surprisal were similar to the effects of SAC on pupil gain. This further hints at a relationship between surprisal and pupil gain, which would be the most direct evidence for the presence of arousal-mediated belief updating in our paradigm.

Indeed, we found that surprisal predicted pupil gain during ongoing stimulus sequences. Since we established surprisal as an indicator of belief updating, this link between ongoing surprisal and pupil gain is in line with previous suggestions that have argued that the pupil-linked arousal system mediates belief updating (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2021; Nassar et al., 2012). It seems plausible that surprisal signalled change points that rendered priors irrelevant (Yon & Frith, 2021), and induced states of high arousal, as indicated by pupil dilation (Dayan & Yu, 2006; Filipowicz et al., 2020; Nassar et al., 2012). This arousal state might then allow us to focus more on sensory evidence compared to the (now irrelevant) spatial prior (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Lawson et al., 2017; Ma, 2019). Hence, it seems plausible that the pupil-linked arousal system mediates flexible updating of beliefs in accordance with change points possible.

In contrast, we did not find a relation between probe surprisal and pupil gain. Note, however, that we regressed surprisal against pupil gain on a trial-by-trial basis for each subject. Since a probe trial consisted of only a single data point, both probe pupil gain and surprisal were rather noisy to begin with, especially since many other factors besides belief updating can drive pupil dilation. For instance, Denison et al. (2020) argue for the possibility of a pupil response before the onset of a stimulus. In line with this, Krishnamurthy et al. (2017) argue that pupil dilation can be driven by the anticipation of a monetary reward participants would receive after a correct answer. While we did not reward correct answers monetarily, we provided verbal encouragement to improve performance at the end of every block. Hence, perhaps this anticipation of a verbal reward (or lack thereof) affected pupil responses to probes. It seems likely that a study that controls for these confounds and has a higher total number of probe trials would find a link between probe surprisal and pupil gain.

Taken together, our results provide strong evidence for automatic belief updating in spatial localization, which seems to be mediated by the LC-NA arousal system (Krishnamurthy et al., 2017; Lawson et al., 2021; Preuschoff et al., 2011; Nassar et al., 2012), and is indicated by pupil dilation. Specifically, these findings suggest that arousal-mediated belief updating constitutes and implicit perceptual process. However, using pupil dilation as an indicator of brain activity has a few limitations, most notably the so-called *inverse problem* (Joshi & Gold, 2020). Given that several neural pathways drive pupil dilation, how can we infer LC-NA system activity from pupil dilation alone? Many other complex and interconnected neural pathways modulate pupil size (for a review, see Joshi & Gold, 2020; Strauch et al., 2022). While we could control for confounding factors such as light levels (Brunyê & Gardony, 2017; Strauch et al., 2022) or focal distance (Strauch et al., 2022), other cognitive and neural processes besides arousal-mediated belief updating, such as anticipation of a reward (Krishnamurthy et al., 2017) could have possibly driven pupil dilation. Furthermore, the LC-NA system seems to be part of a complex reciprocal neural network by sending and receiving information from various cortical and subcortical areas (Joshi & Gold, 2020). Hence, we must be careful not to oversimplify and only consider the LC-NA system as the neural basis of belief updating, as the underlying neural mechanisms behind belief updating are likely to be much more complex (also see Strauch et al., 2022). Also note that pupil gain is not a direct measure of pupil diameter, but a latent variable extracted from a deconvolution-based linear model (Denison et al., 2020). This approach allowed us to disentangle overlapping pupil traces of ongoing stimulus sequences. While it has become quite common to describe pupil dilation as the linear transformation of event-related arousal (Denison et al., 2020), this approach is currently criticized for oversimplification (Burlingham et al., 2022).

Hence, even though our results strongly suggest a central role of the LC-NA system in belief updating, the actual neurophysiological mechanisms behind it are likely to be much more complex. It should be noted, however, that our observed link between pupil dilation and the LC-NA system and its role in belief updating are backed up by several psychopharmacological and electrophysiological studies. For instance, Lawson et al. (2021) found that administering the NA antagonist propranolol reduced the updating of beliefs and pupil responsivity to changes in the statistical probabilities of a sound cue. Furthermore, de Gee et al. (2021) showed that rapid pupil dilation and the slow-wave event-related potential (ERP) component reflected changes in arousal and were linked to surprisal. They argue for the slow-wave ERP component as another indicator of belief updating and further corroborate the notion that belief updating is arousal-mediated. The EEG data we collected during our study will hopefully shed more light on this topic. Still, future studies should assess arousal-mediated belief updating in our adapted version of the Krishnamurthy et al. (2017) paradigm using imaging techniques such as functional magnetic resonance imaging (fMRI) or psychopharmacological interventions to provide more direct evidence for the role of the LC-NA system in belief updating.

Finally, it is important to note that we modelled the link between surprisal and pupil gain as a linear relationship. While this facilitates interpretation, it also simplifies the actual processes of arousal-mediated belief updating. Our findings on the effects of SAC on pupil gain and surprisal may hint at a non-linear relation. Specifically, while pupil gain decreased linearly as a function of SAC, we found a strong, almost exponential decrease of surprisal from SAC1 to SAC2. It may be helpful to use non-linear approaches to describe the processes of arousal-mediated belief updating in perceptual decision-making more accurately in the future.

In conclusion, the finding that estimation performance depended on the occurrence of a change point in our experiment provides ample evidence that belief updating does not require conscious effort and instead constitutes an implicit perceptual process. This belief updating appears to be mediated by the pupil-linked arousal system because pupil gain tracked the occurrence of a change point and surprisal predicted pupil gain of ongoing stimulus sequences.

4.2. Arousal-mediated belief updating is supramodal

4.2.1 Beneficial effects of sound after change point level on estimation accuracy are not exclusive to the visual domain

We found that the effect of SAC on MAE was also present for auditory stimulus sequences. This corroborates our notion that belief updating in spatial localization is not exclusive to the visual domain. However, MAE decreased stronger as a function of SAC for audio-visual than for auditory stimulus sequences. This makes sense under the premise that we usually bias spatial judgements toward the visual domain (Charlton et al., 2022) due to the overall higher accuracy of spatial localization in the visual system compared to the auditory system (for a review, see Opoku-Baah et al., 2021). Perhaps the higher accuracy in locating audio-visual stimuli then also allowed for a more efficient learning of statistical regularities of audio-visual compared to auditory stimuli. This might explain why we found stronger effects of SAC on spatial accuracy in the AV modality. Consequently, it may be that the perceptual system benefitted more strongly from audio-visual stimuli during perceptual inference, which might also explain why MAEs in the A condition were higher than the MAEs of localizer trials. In other words, even though estimation performance followed a pattern that is expected by the Bayesian inference theory, the pattern may not be strong enough to buffer out the deteriorating effects of the attentional repulsion effect (Suzuki & Cavanagh, 1997).

Either way, participants still appeared to be able to extrapolate the underlying generative mean of auditory stimuli, even if they were less successful than for audio-visual stimuli in doing so. This finding implies that participants used and updated spatial priors based on both audio-visual and auditory stimuli, providing preliminary support that belief updating may not be exclusive to the visual domain during spatial localization.

4.2.2 Pupil gain tracked the occurrence of change points regardless of stimulus modality

Again, while the effects of SAC on MAE imply the presence of arousal-mediated belief updating during spatial localization regardless of modality, they do not address the role of the pupil-linked arousal system in it. However, we found that pupil gain decreased as a function of SAC in both the AV and A modalities. Hence, the pupil-linked arousal system tracked the occurrence of a change point through pupil gain regardless of stimulus modality. Interestingly, this effect was equally strong for both AV and A modalities. Since pupil dilation by itself, of course, cannot enhance the perception of auditory stimuli, this challenges previous suggestions that the pupil-linked arousal system increases the receptiveness toward sensory information of visual stimuli (de Gee et al., 2014; Yu & Dayan, 2005). Instead, it seems that the pupil-linked arousal decreased the weight of spatial priors rather than facilitated sensory input processing (Lawson et al., 2021; Urai et al., 2017).

In any case, the fact that pupil gain was also sensitive to change points during auditory stimulus sequences supports our assumption that arousal-mediated belief updating is not exclusive to the visual domain.

4.2.3 Surprisal predicted pupil gain in ongoing stimulus sequences regardless of modality

The relation between surprisal and pupil gain is the most direct approach to address our assumption that arousal-mediated belief updating is supramodal. As mentioned above, our Bayesian inference model was explicitly designed for the change point paradigm, making surprisal a viable indicator of belief updating (Charlton et al., 2022; Filipowicz et al., 2020; Yon & Frith, 2021). Importantly, our results suggest that this is true for both AV and A modalities. Specifically, we have shown that the strong effect of SAC on both ongoing and probe surprisal was present for both modalities. In other words, surprisal decreased strongly as a function of SAC, regardless of stimulus modality. This provides support for the notion that surprisal can be used as an indicator of belief updating in both modalities.

Crucially, we found that the relation between ongoing surprisal and pupil gain was comparable between the AV and A modality. In other words, surprisal predicted pupil gain regardless of stimulus modality. This suggests that the ability of the perceptual system to detect change points and extrapolate the underlying generative mean of a stimulus sequence is not exclusive to the visual domain. In other words, these findings further corroborate our assumption that arousal-mediated belief updating during spatial perception is supramodal. Again, this is a very interesting finding, as it suggests that the role of the pupil-linked arousal system goes beyond enhancing the receptiveness of the pupil toward visual information. Instead, the pupil-linked arousal system seems to mediate belief updating by suppressing priors in response to the occurrence of a change point (Lawson et al., 2021; Urai et al., 2017), thereby increasing the relative weight of sensory information on the perceptual judgement (Charlton et al., 2022; Filipowciz et al., 2020; Lawson et al., 2021). Also note that there was no difference in the link of surprisal and pupil gain between modalities, suggesting that the sensitivity of the perceptual system toward change points is comparable between audio-visual and auditory stimuli.

Our study expands existing literature by showing that arousal-mediated belief updating in spatial localization is not exclusive to the visual domain but can also be found for auditory stimulus sequences. Even though our results imply that spatial priors based on audio-visual stimuli were used and updated more efficiently than those based on auditory stimuli to improve localization accuracy, they strongly suggest that belief updating in spatial localization is supramodal. Furthermore, it seems that the pupil-linked arousal system does not distinguish between modalities when signalling the occurrence of a change point, which implies that its contributions to belief updating in spatial localization does not depend on stimulus modality.

4.3 Conclusion

Our results extend to previous studies about arousal-mediated belief updating in spatial localization. Specifically, it seems that arousal-mediated belief updating does not require conscious effort but constitutes an implicit perceptual process. Furthermore, our findings suggest that the perceptual system cannot just extrapolate statistical regularities of the sensory environment from audio-visual (Krishnamurthy et al., 2017) or visual stimuli (Charlton et al., 2022; De Gee et al., 2021; Knill & Pouget, 2007), but also from purely auditory stimuli. Thus, arousal-mediated belief updating appears to be a supramodal component of perceptual decision-making. Therefore, our findings closed a current gap in the literature and should help deepen our understanding of how the perceptual system can make accurate perceptual judgements despite sensory uncertainty.

Acknowledgement

This thesis was written as part of a bigger project of the young independent researcher group Dynamates of the Austrian Academy of Sciences. The expertise and inputs of Dr. Robert Baumgartner and Dr. David Meijer (Acoustics Research Institute of the Austrian Academy of Sciences, Vienna) have been invaluable throughout the entire thesis. Specifically, they designed the study and provided important input on the analyses, as well as the structuring of the thesis.

5. References

Acerbi, L. & Ma, J. W. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Neural Information Processing Systems, 30*, 1834–1844. https://doi.org/10.48550/arXiv.1705.04405

Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S. & Steinhauser, M. (2014). The neural basis of belief updating and rational decision making. *Social Cognitive and Affective Neuroscience, 9*(1), 55–62. https://doi.org/10.1093/scan/nss099

Aitchison, L. & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology, 46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A. & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience, 39*(27), 5369–5376. https://doi.org/10.1523/jneurosci.3010-18.2019

Aston-Jones, G. & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

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

Barumerli, R.; Meijer, D.; Baumgartner, R. (2022) *Joint modeling confirms pupil dilation as neurophysiological marker of Bayesian spatial inference in dynamic auditory environments* [Conference session]. ICA 2022 (24th International Congress on Acoustics), Gyeongju.

Bejjanki, V. R., Knill, D. C. & Aslin, R. N. (2016). Learning and inference using complex generative models in a spatial localization task. *Journal of Vision, 16*(5), Article 9. https://doi.org/10.1167/16.5.9

Berniker, M., Voss, M. & Kording, K. (2010). Learning priors for Bayesian computations in the nervous system. *PLoS ONE, 5*(9), Article e12686. https://doi.org/10.1371/journal.pone.0012686

Bouret, S. & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences, 28*(11), 574–582. https://doi.org/10.1016/j.tins.2005.09.002

Breton-Provencher, V. & Sur, M. (2019). Active control of arousal by a locus coeruleus GABAergic circuit. *Nature Neuroscience, 22*(2), 218–228. https://doi.org/10.1038/s41593-018-0305-z

Brunyé, T. T. & Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *International Journal of Psychophysiology, 120*, 60–68. https://doi.org/10.1016/j.ijpsycho.2017.07.008

Burlingham, C. S., Mirbagheri, S. & Heeger, D. J. (2022). A unified model of the task-evoked pupil response. *Science Advances, 8*(16), Article eabi9979. https://doi.org/10.1126/sciadv.abi9979

Cai, Y., Chen, G., Zhong, X., Yu, G., Mo, H., Jiang, J., Chen, X., Zhao, F. & Zheng, Y. (2018). Influence of audio-visual training on horizontal sound localization and its related ERP response. *Frontiers in Human Neuroscience, 12*, Article 423. https://doi.org/10.3389/fnhum.2018.00423

Charlton, J. A., Młynarski, W. F., Bai, Y. H., Hermundstad, A. M. & Goris, R. L. T. (2022). *Perceptual decisions exhibit hallmarks of dynamic Bayesian inference*. bioRxiv. https://doi.org/10.1101/2022.05.23.493109

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Dayan, P. & Yu, A. J. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems, 17*(4), 335–350. https://doi.org/10.1080/09548980601004024

de Gee, J. W., Correa, C. M. C., Weaver, M., Donner, T. H. & van Gaal, S. (2021). Pupil dilation and the slow wave ERP reflect surprise about choice outcome resulting from intrinsic variability in decision confidence. *Cerebral Cortex, 31*(7), 3565–3578. https://doi.org/10.1093/cercor/bhab032

de Gee, J. W., Knapen, T. & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences, 111*(5), 618-625. https://doi.org/10.1073/pnas.1317557111 de Lange, F. P., Heilbron, M. & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Denison, R. N., Parker, J. A. & Carrasco, M. (2020). Modeling pupil responses to rapid sequential events. *Behavior Research Methods, 52*(5), 1991–2007. https://doi.org/10.3758/s13428-020-01368-6

Devauges, V. & Sara, S. J. (1990). Activation of the noradrenergic system facilitates an attentional shift in the rat. *Behavioural Brain Research, 39*(1), 19–28. https://doi.org/10.1016/0166-4328(90)90118-x

Edwards, W., Lindman, H. & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*(3), 193–242. https://doi.org/10.1037/h0044139

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology, 2*(2), 312–329. https://doi.org/10.1016/0022-2496(65)90007-6

Etz, A. & Vandekerckhove, J. (2017). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review, 25*(1), 5–34. https://doi.org/10.3758/s13423-017-1262-3

Fischer, B. J. & Peña, J. L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience, 14*(8), 1061–1066. https://doi.org/10.1038/nn.2872

Filipowicz, A. L., Glaze, C. M., Kable, J. W. & Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife, 9*, Article e57872. https://doi.org/10.7554/elife.57872

Gold, J. I., Law, C. T., Connolly, P. & Bennur, S. (2008). The relative influences of priors and sensory evidence on an oculomotor decision variable during perceptual learning. *Journal of Neurophysiology, 100*(5), 2653–2668. https://doi.org/10.1152/jn.90629.2008

Hoeks, B. & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers, 25*(1), 16–26. https://doi.org/10.3758/bf03204445

JASP Team (2023). *JASP* (Version 0.17.1) [Computer software. https://jasp-stats.org

Jepma, M. & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience, 23*(7), 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Joshi, S. & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences, 24*(6), 466–480. https://doi.org/10.1016/j.tics.2020.03.005

Joshi, S., Li, Y., Kalwani, R. & Gold, J. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kleiner, M., Brainard, D. H. & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1–16. https://journals.sagepub.com/doi/pdf/10.1177/03010066070360S101

Knill, D. C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

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

Krishnamurthy, K., Nassar, M. R., Sarode, S. & Gold, J. I. (2017). Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour, 1*(6), Article 0107. https://doi.org/10.1038/s41562-017-0107

Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N. & Rees, G. (2021). The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. *Current Biology, 31*(1), 163-172.e4. https://doi.org/10.1016/j.cub.2020.10.043

Lempert, K. M., Chen, Y. L. & Fleming, S. M. (2015). Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS ONE, 10*(5), Article e0126588. https://doi.org/10.1371/journal.pone.0126588

Ma, W. J. (2019). Bayesian decision models: A primer. *Neuron, 104*(1), 164–175. https://doi.org/10.1016/j.neuron.2019.09.037

Mathôt, S. (2013). A simple way to reconstruct pupil size during eye blinks. *figshare*. Journal contribution. https://doi.org/10.6084/m9.figshare.688001.v1

Mathys, C., Daunizeau, J., Friston, K., Klaas, S. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience, 5*, Article 39. https://doi.org/10.3389/fnhum.2011.00039

MATLAB. (2022). *Version 2022* (R2022a) [Computer software]. Natick, Massachusetts: The MathWorks Inc.

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping, 35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B. & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience, 15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience, 30*(37), 12366–12378. https://doi.org/10.1523/jneurosci.0822-10.2010

Opoku-Baah, C., Schoenhaut, A., Vassall, S. G., Tovar, D. A., Ramachandran, R. & Wallace, M. S. (2021b). Visual influences on auditory behavioral, neural, and perceptual processes: A Review. *Jaro-journal of The Association for Research in Otolaryngology, 22*(4), 365–386. https://doi.org/10.1007/s10162-021-00789-0

Payzan-LeNestour, E. & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Computational Biology, 7*(1), Article e1001048. https://doi.org/10.1371/journal.pcbi.1001048

Pecka, M., Leibold, C. & Grothe, B. (2020). Biological aspects of perceptual space formation. *Modern Acoustics and Signal Processing*, 151–171. https://doi.org/10.1007/978-3-030-00386-9_6

Petzschner, F. H., Glasauer, S. & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Preuschoff, K., 't Hart, B. M., Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*, Article 115. https://doi.org/10.3389/fnins.2011.00115

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A. & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications, 7*(1), Article 13289. https://doi.org/10.1038/ncomms13289

Sara, S. & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron, 76*(1), 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M. & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage, 37*(3), 1017–1031. https://doi.org/10.1016/j.neuroimage.2007.04.066

Schröger, E. & Roeber, U. (2021). Encoding of deterministic and stochastic auditory rules in the human brain: The mismatch negativity mechanism does not reflect basic probability. *Hearing Research, 399*, Article 107907. https://doi.org/10.1016/j.heares.2020.107907

Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., Howard, M. A. & Griffiths, T. D. (2016). Neural signatures of perceptual inference. *eLife, 5*, Article e11476. https://doi.org/10.7554/elife.11476

Seriès, P. & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience, 7*, Article 668. https://doi.org/10.3389/fnhum.2013.00668

Skerritt-Davis, B. & Elhilali, M. (2021). Neural encoding of auditory statistics. *The Journal of Neuroscience, 41*(31), 6726–6739. https://doi.org/10.1523/jneurosci.1887-20.2021

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578–585. https://doi.org/10.1038/nn1669

Strauch, C., Wang, C. A., Einhäuser, W., Van der Stigchel, S. & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences, 45*(8), 635–647. https://doi.org/10.1016/j.tins.2022.05.003

Summerfield, C. & Egner, T. (2009b). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Urai, A. E., Braun, A. & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications, 8*(1), Article 14637. https://doi.org/10.1038/ncomms14637

Yon, D. & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology, 31*(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044

Yu, A. J. & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zhang, S., Huang, H. C. & Yu, A. J. (2014). Sequential effects: A Bayesian analysis of prior bias on reaction time and behavioral choice. *Cognitive Science, 36*(36), 1844-1849. https://escholarship.org/content/qt9jp3j5c7/qt9jp3j5c7.pdf?t=op9xyz

6. Table of figures

7. List of abbreviations

Abbreviation	Definition
A	Auditory
ANOVA	Analysis of variance
AV	Audio-visual
BOLD	Blood oxygen level dependent
EEG	Electroencephalogram
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
HRIR	Head-related impulse function
ISI	Interstimulus-interval
LC-NA system	Locus coeruleus noradrenaline system
MAE	Mean absolute localization error
PuRF	Pupil response function
px	Pixel
SAC	Sound after changepoint
SEM	Standard error of the mean

8. Appendix

English Abstract

The perceptual system can accurately localize stimuli despite sensory uncertainty by flexibly adjusting prior expectations about the most probable stimulus location (belief updating). Previous research has suggested that an increase in arousal, indicated by pupil dilation, mediates belief updating. We addressed two open questions: (1) Does arousal-mediated belief updating during spatial localization constitute an implicit perceptual process, and (2) is it exclusive to the visual domain? To this end, 24 healthy individuals performed a dynamic sound localization task during which we measured pupil diameter. During the task, a stochastic sequence of stimuli whose average location could change randomly (change point) was presented. In contrast to previous studies, participants only estimated the position of the last sound after it had been presented and did not explicitly predict it beforehand. Furthermore, in addition to an audio-visual condition as in previous studies, we also included a purely auditory condition. In both conditions, we found that estimation error was highest immediately after a change point but fell off if the change point occurred more than one sound ago. Furthermore, pupil-gain, an indicator of arousal estimated by a deconvolution-based linear model, tracked the occurrence of change points, and gradually decreased the further away the change point was. Crucially, pupil gain was also predicted by surprisal, an indicator of belief updating, in both conditions. Together, these findings indicate the presence of arousal-mediated belief updating during spatial perception. Hence, our findings suggest that arousal-mediated belief updating during spatial localization constitutes an implicit perceptual process and is supramodal.

Deutsches Abstract

Das perzeptuelle System kann Reize trotz sensorischer Unsicherheit präzise räumlich lokalisieren, indem es Vorerwartungen über die wahrscheinlichste Reizposition flexibel anpasst (Erwartungsanpassung). Bisherige Forschung hat nahegelegt, dass erhöhtes Erregungsniveau, angezeigt durch Pupillenerweiterung, Erwartungsanpassung mediiert. Wir haben zwei offene Fragen beantwortet: (1) Ist Erwartungsanpassung, mediiert durch Erregungsniveau, ein impliziter Wahrnehmungsprozess und (2) ist Erwartungsanpassung exklusiv in der visuellen Modalität? Dafür haben 24 gesunde Versuchspersonen eine dynamische Ton-Lokalisierungs-Aufgabe durchgeführt, während wir Pupillendurchmesser erhoben haben. Während der Aufgabe wurde eine stochastische Sequenz an Reizen präsentiert, dessen durchschnittliche Position sich zufällig ändern konnte (Änderungspunkt). Im Gegensatz zu vorigen Studien mussten Versuchspersonen lediglich die Position des letzten Tons schätzen, nachdem er präsentiert wurde, statt seine Position davor explizit vorherzusagen. Außerdem führten wir neben einer audio-visuellen Bedingung aus anderen Studien auch eine rein auditive Bedingung ein. Wir fanden in beiden Bedingungen, dass der Schätzfehler direkt nach einem Änderungspunkt am höchsten

war, aber abfiel, wenn der Änderungspunkt mehr als ein Ton her war. Außerdem zeigte Pupillenzunahme, ein Indikator von Erregungsniveau den wir durch ein dekonvolutions-basiertes lineares Modell schätzten, das Auftreten eines Änderungspunkts an und sank graduell je länger der Änderungspunkt her war. Am wichtigsten war der Befund, dass Pupillenzunahme durch Überraschung, einem Indikator von Erwartungsanpassung, in beiden Bedingungen vorhergesagt wurde. Zusammen deuten diese Ergebnisse auf die Anwesenheit von Erwartungsanpassung, mediiert durch Erregungsniveau. Daher legen unsere Ergebnisse nahe, dass durch Erregungsniveau mediierte Erwartungsanpassung während räumlicher Lokalisierung einen impliziten Wahrnehmungsprozess darstellt, der supramodal ist.

Table 1.
Post hoc comparisons between SAC levels in MAE

		t	Cohen's *d*	p_{bonf}
SAC1	**SAC2**	**4.02**	**0.523**	**.002**
	SAC3	**5.88**	**0.764**	**< .001**
	SAC4	**6.50**	**0.845**	**< .001**
	SAC5	**4.66**	**0.606**	**< .001**
	SAC6	**5.06**	**0.659**	**< .001**
SAC2	SAC3	1.86	0.242	.988
	SAC4	2.48	0.322	.220
	SAC5	0.64	0.083	1.000
	SAC6	1.05	0.136	1.000
SAC3	SAC4	0.62	0.081	1.000
	SAC5	-1.22	-0.159	1.000
	SAC6	-0.81	-0.106	1.000
SAC4	SAC5	-1.84	-0.240	1.000
	SAC6	-1,43	-0.186	1.000
SAC5	SAC6	0.41	0.053	1.000

Notes. Bold rows indicate significant comparisons ($p < .05$, two-tailed). *MAE,* mean absolute error, *SAC*, sound after change point level; p_{bonf}, Bonferroni corrected *p*-values.

Table 2.
Post hoc comparisons between SAC levels in ongoing pupil gain

		t	Cohen's *d*	p_{bonf}
SAC1	SAC2	0.68	0,183	1.000
	SAC3	2.40	0.652	.266
	SAC4	**3.22**	**0.873**	**.025**
	SAC5	2.95	0.799	.058
	SAC6	**3.54**	**0.960**	**.009**
SAC2	SAC3	1.73	0.469	1.000
	SAC4	2.55	0.690	.183
	SAC5	2.27	0.616	.373
	SAC6	2.87	0.777	.074
SAC3	SAC4	0.82	0.221	1.000
	SAC5	0.54	0.147	1.000
	SAC6	1.14	0.308	1.000
SAC4	SAC5	-0.27	-0.074	1.000
	SAC6	0.32	0.087	1.000
SAC5	SAC6	0.59	0.161	1.000

Notes. Bold rows indicate significant comparisons ($p < .05$, two-tailed). *MAE,* mean absolute error, *SAC*, sound after change point level; p_{bonf}, Bonferroni corrected *p*-values

Table 3.
Post hoc comparisons between SAC levels in ongoing surprisal

		t	Cohen's *d*	p_{bonf}
SAC1	**SAC2**	**21.76**	**5.671**	**< .001**
	SAC3	**25.70**	**6.697**	**< .001**
	SAC4	**25.19**	**6.565**	**< .001**
	SAC5	**26.08**	**6.797**	**< .001**
	SAC6	**25.57**	**6.663**	**< .001**
SAC2	**SAC3**	**3.94**	**1.026**	**.002**
	SAC4	**3.43**	**0.894**	**.013**
	SAC5	**4.32**	**1.123**	**< .001**
	SAC6	**3.80**	**0.992**	**.003**
SAC3	SAC4	-0.51	-0.132	1.000
	SAC5	0.38	0.100	1.000
	SAC6	-0.13	-0.034	1.000

SAC4	SAC5	0.89	0.232	1.000
	SAC6	0.38	0.098	1.000
SAC5	SAC6	-0.51	-0.134	1.000

Notes. Bold rows indicate significant comparisons (p < .05, two-tailed). *SAC*, sound after change point level; p_{bonf}, Bonferroni corrected p-values.

Table 4.

Post hoc comparisons between SAC levels in probe surprisal

		t	Cohen's d	p_{bonf}
SAC1	**SAC2**	**12.28**	**3.179**	**< .001**
	SAC3	**17.88**	**4.628**	**< .001**
	SAC4	**18.47**	**4.782**	**< .001**
	SAC5	**19.06**	**4.936**	**< .001**
	SAC6	**21.79**	**5.642**	**< .001**
SAC2	**SAC3**	**5.60**	**1.449**	**< .001**
	SAC4	**6.19**	**1.603**	**< .001**
	SAC5	**6.78**	**1.757**	**< .001**
	SAC6	**9.51**	**2.462**	**< .001**
SAC3	SAC4	0.59	0.154	1.000
	SAC5	1.19	0.308	1.000
	SAC6	**3.91**	**1.013**	**.002**
SAC4	SAC5	0.59	0.154	1.000
	SAC6	3.32	0.860	.018
SAC5	SAC6	2.73	0.706	.111

Notes. Bold rows indicate significant comparisons (p < .05, two-tailed). *SAC*, sound after change point level; p_{bonf}, Bonferroni corrected p-values.