universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Neural signatures of Bayesian perceptual adaptation during auditory motion discrimination“

verfasst von / submitted by

Roman Fleischmann, BSc

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

Master of Science (MSc)

Wien, 2024 / Vienna 2024

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

A. INTRODUCTION

A. 1. Perception under uncertainty

Thinking about perception, one might assume it is a straightforward process of sensory input provided by our sensory organs that is being used as-is to perceive our environment. But once you dive in a little further, problems arise. Our sensory organs are not necessarily precise (Parr et al., 2018; Plack, 2013), depending on and varying a lot with external conditions (Deneve & Pouget, 2004; Plack, 2013). Visual input can be blurry, auditory input misleading, tactile input can be mislocalized (Maij et al., 2020) and perceptual illusions from the blind spot (Paradiso & Nakayama, 1991; Ramachandran & Gregory, 1991) to the rubber hand illusion (Kammers et al., 2009) or even hallucinations (Corlett et al., 2019) are well known and defy any last notion of objectivity in perception. With sensory organs being imprecise and input being faulty, our sensory input is uncertain (Knill & Pouget, 2004), which arises two obvious issues: (1) The same sensory input could originate from different world states and (2) different sensory inputs, be it from different sensory organs or from a time series of sensory observations from a single organ, might not match, as they don't necessarily accurately reflect reality (Deneve & Pouget, 2004; Yon & Frith, 2021). Both problems require a perceptual decision to be made. The situation gets even worse if we consider that not only the sensory input is uncertain, but also the environment that we are trying to perceive itself, as it can be subject to change. This begs the question: How are different sensory inputs integrated and how can we achieve precise perception under these uncertain conditions?

A. 1.1. Bayesian inference

An example could be an audio-visual mismatch, where a sound associated with an event appears to come from a different direction, than the event. This might be due to noisy sensory perception, where we hear the correct sound but misjudge the location it came from (Plack, 2013), or due to mistakenly connecting the wrong sound to the wrong event. In both cases we end up with one event being linked to two different locations in the respective different sensory modalities. Merging them into one coherent perceived location is a classical multisensory integration problem (Deneve & Pouget, 2004; Yon & Frith, 2021). While we could just average out both locations, this is obviously not the best way to go, as common sense tells us that a normal-hearing and normal-seeing person

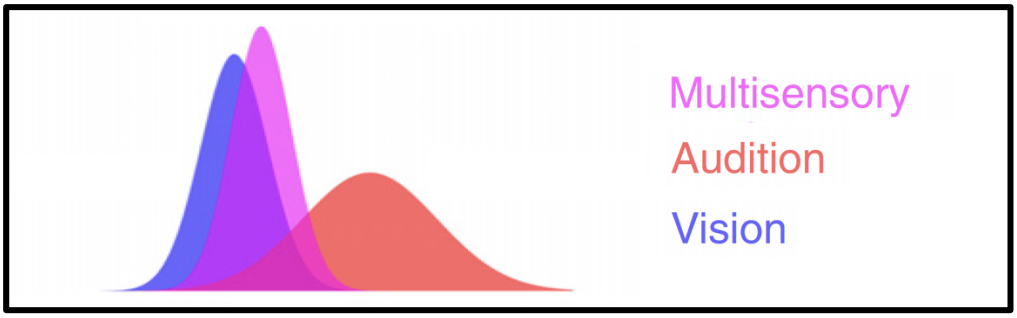
should generally trust their eyes over their ears. We instinctively know that our auditory input is to some degree uncertain, at least when compared to our vision.

One way to resolve this issue in a better way is *Bayesian inference* (Clark, 2013; Deneve & Pouget, 2004; Knill & Pouget, 2004; Yon, 2021; Yon & Frith, 2021). Ernst and Banks (2002) find humans to integrate different signals close to a statistically optimal fashion and propose a model that assumes different signals to be processed based on their precision or reliability. Perception is seen as a process of probabilistic inference, where an agent estimates the most probable state of the world based on incoming signals, (optimally) accounting for each signal's uncertainty (Clark, 2013).

Let's take a meowing cat as an example. Both sensory inputs (visual image of the cat, sound of the meowing) are represented by probability density functions which are used to infer an estimate for the real position. The product of both functions is the perceptual inference, or within our example: The optimal guess for the cat's location. The width of each function represents the precision or reliability of the sensory input, essentially adding the subjective uncertainty in form of weights to the multiplication (Knill & Pouget, 2004; Yon & Frith, 2021). To stay with our example, visual input would have a narrower distribution than auditive input, as we see the location of the cat precisely, while we hear the meow more from a general direction. An illustration of this multisensory integration can be seen in Figure 1.

Figure 1

Bayesian multisensory integration (Yon & Frith, 2021)



Note. Multisensory integration as a process of Bayesian inference weighs the incoming sensory inputs by their precision or reliability, represented in the width of the respective distribution, to infer the most probable state of the world.

With decreasing visual conditions, for example darkness, the visual estimate becomes less clear, hence, less reliable, less precise. Any vaguely cat-sized dark spot could possibly be it, so if we now hear a meow from a general direction, this is where we

are searching for our lost cat. With increasing darkness, the weighting of sound and image by their reliability has changed. We are now attributing a higher reliability to the sound (that we still perceive clearly) than to the visual image (that became uncertain), and we have adjusted our optimal guess (Yon & Frith, 2021).

A. 1.2. Prior belief

The example dealt with integration over different modalities to be merged. Bayesian inference suggests that not only differences between bottom-up sensory signals are resolved this way, but also differences between bottom-up sensory signals and top-down expectations (Clark, 2013; Ernst & Banks, 2002; Knill & Pouget, 2004; Yon & Frith, 2021). Or phrased differently: Top-down expectations influence our perception (de Lange et al., 2018; Yon & Frith, 2021).

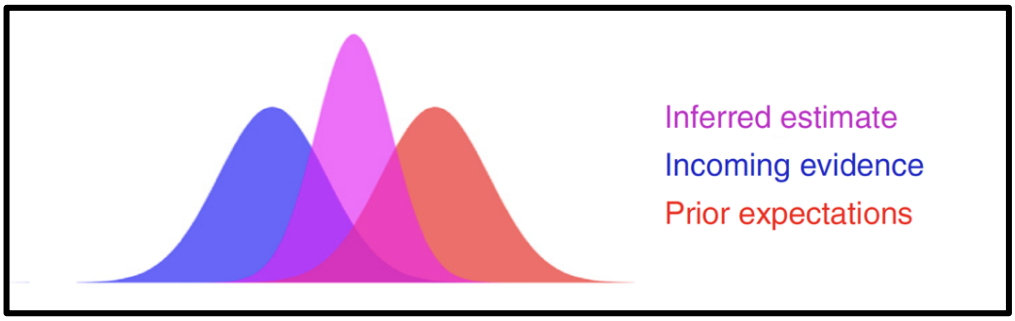
This becomes especially relevant regarding the abovementioned differing sensory inputs in a timeseries of inputs from one singular sensory organ. A first sensory input will translate into a belief about the environment, which will translate into an expectation about future sensory input. The actual following sensory input might under uncertain conditions not match the preceding one, or the expectation. This mismatch is postulated to be resolved in the same fashion as in multisensory integration. In Bayesian terms the sensory input is called the *likelihood*, while the belief about the environment is called *prior*. Their product, called the *posterior*, is the perceptual inference (Clark, 2013; Ma, 2019; Yon & Frith, 2021).

Let's stay with the problem of finding a cat in the dark as an example. In this scenario we are relying on auditory input only; it is difficult to hear, but we have heard a first meowing a second ago. This past somewhat reliable input now represents our top-down belief about the cat's whereabouts. It is our prior. If we were to hear another meow, this is where we would it expect to come from. The next actual meow we are hearing is our likelihood, the bottom-up sensory input, which might not be perceived at the exact same location. Both prior and likelihood are then used to infer the posterior, our optimal estimate, which is now our current belief about the world state of interest (i. e. the cat's location). For a possible third meow, this posterior would serve as the new prior, together with a new likelihood. With every additional meow the prior becomes increasingly more reliable, given the cat did not move (Knill & Pouget, 2004; Ma, 2019; Yon & Frith, 2021).

Despite the bottom-up auditory input staying constant, we are increasingly surer about the cat's location, by decreasing uncertainty (see Figure 2).

Figure 2

Bayesian integration of sensory input and prior expectations (Yon & Frith, 2021)



Note. According to Bayesian perceptual inference both bottom-up incoming evidence (likelihood) and top-down expectations (prior) are utilized to infer the true state of the environment.

A. 1.3. The ideal Bayesian observer

The exact definition of the posterior is given by Bayes' theorem (Ma, 2019), which, formulated for our exemplary cat, would look like this (Eq. 1):

$$p(s|x_{cat}) = \frac{p(x_{cat}|s)p(s)}{p(x_{cat})} \qquad \text{(Equation 1)}$$

The posterior for the cat's true location s given our observation x_{cat} is the product of the prior $p(s)$ and the likelihood $p(x_{cat}|s)$, divided by $p(x_{cat})$. The posterior of the previous observation serves as the prior of the current observation. In a series of observations under static conditions this leads to an accumulation of information in the prior with every additional observation. In cases with no preceding beliefs a prior can be represented by a distribution holding no or very little information, like a uniform distribution (Ma, 2019).

Bayes' theorem describes a statistically optimal inference about the cat's location. This assumption is commonly illustrated as the ideal Bayesian observer, a postulated ideal rationalized agent acting solely upon Bayes' theorem and achieving statistically optimal performance (Kersten & Mamassian, 2009). Whether or not and to what extent humans function like the ideal Bayesian observer is topic of debate (Rahnev, 2019) and exceeds by far the limitations of this thesis.

For the context of this thesis it is important to state that while there are alternatives like pattern theory (de Lange et al., 2018) or the closely related predictive coding (Sedley

et al., 2016), which are also able to model perception, there is no such thing as proof for any of these theories. There are good reasons to work within the Bayesian framework when assessing perception, nevertheless. As an approach to model human perception Bayesian inference has proven successful across a large number of papers yielding sensible results in describing human perception and behavior in existing research (Ma, 2019; Yon & Frith, 2021) and building computational theories for perception (Knill & Pouget, 2004). Despite being a rather recent application, gaining traction in the last decades (de Lange et al., 2018; Petzschner et al., 2015), the method itself has been introduced to psychology much earlier as a possible alternative to classical hypothesis testing (Edwards et al., 1963; Fornacon-Wood et al., 2022) and is hence well established.

A. 1.4. Dynamic environments

In theory Bayes' theorem can be applied and repeated indefinitely, endlessly accumulating information in the prior and improving the precision with every new observation. That is, given the additional observations are roughly consistent with the initial one (Ma, 2019). Most studies within the Bayesian framework are conducted under the assumption of a static environment (Kolossa et al., 2015; Visalli et al., 2021).

A dynamic environment poses a challenge to this. Sudden change in the environment can render all information held in the prior obsolete (Knill & Pouget, 2004; Ma, 2019; Nassar et al., 2012). To stay with the last example, if the cat unpredictably and inaudibly moves to a different spot, the location held in the prior is either useless or even detrimental for our task to find the cat, no matter how precise it eventually got. In this case perception should suddenly only rely on the sensory input, or likelihood in Bayesian terms. An attempt to make use of the prior anyways would result in a bias towards the cat's former favorite spot. If this likelihood is precise enough the influence of the now obsolete prior and therefore the bias will likely be negligible, but if the perception is uncertain the prior's relative weight increases and the bias becomes stronger. This could possibly lead to misjudgment and identification of any random noise as the cat, if coming from the cat's former general direction (Yon & Frith, 2021). A series of new observations will establish a new informative prior, if sufficiently consistent.

Experimental evidence suggests Bayesian inference to achieve high perceptual accuracy in dynamic environments with volatile priors as well (Krishnamurthy et al., 2017; Nassar et al., 2012). An ideal Bayesian observer is postulated to learn and anticipate the

volatility of the environment and adapt the usage of priors for possible changes (Knill & Pouget, 2004; Nassar et al., 2012). At its core Bayesian inference is about the anticipation and reduction of uncertainty; the volatility of the environment simply adds an additional layer of uncertainty on top of existing ones (Clark, 2013; Knill & Pouget, 2004).

A. 2. Surprisal

A measure commonly associated with Bayesian perceptual inference is surprisal (Kolossa et al., 2015; Visalli et al., 2021). Note that the term surprisal does not refer to the emotional state of “being surprised”, but to a measure of information theory, although they are closely related.

Across literature surprisal can be found under different terms, such as surprise (Sedley et al., 2016; Visalli et al., 2021), information content (Baldi & Itti, 2010; Visalli et al., 2021), predictive surprise (Kolossa et al., 2015), Shannon surprise (Modirshanechi et al., 2022), Shannon information (Cole et al., 2021), Bayesian surprise (Itti & Baldi, 2009), as well as under different mathematical definitions (Modirshanechi et al., 2022). This sometimes leads to confusion regarding nomenclature and concepts when comparing different literature (see Visalli et al., 2021, commenting on Kolossa et al., 2015), although most definitions can be traced back to only two different computational concepts: the Kullback-Leibler divergence and the Shannon information (Visalli et al., 2021). The present thesis will focus on the latter.

In their core the different concepts of surprisal across literature refer to (or are at least closely related to) information content, which bridges the gap to the emotional state of “being surprised”. Events are considered more informative if they are improbable and precise (Sedley et al., 2016; Visalli et al., 2021). The connection to the emotional state becomes clear if seen through a subjective perspective, as a particularly informative piece of data, for example a sound that is very precise and very unexpected under current beliefs, would certainly leave you feeling surprised.

In neuroscientific and psychologic literature surprise or surprisal seems to be often understood intuitively, referring to either the emotional state or a definition from information theory or even both, sometimes without clear differentiation (Modirshanechi et al., 2023; Reisenzein et al., 2019).

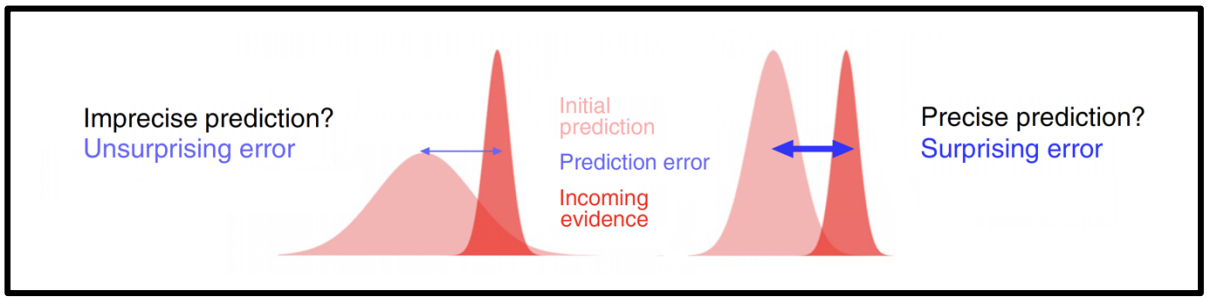
A. 2.1. Surprisal as prediction error

Surprisal can be seen as one possible quantification of the prediction error, specifically the prediction error weighted by its reliability (Sedley et al., 2016; Yon & Frith, 2021), a definition falling into place especially nicely in the context of sudden changes and volatile environments. The prior accumulates information from previous sensory experience as a quantification of belief, accounting for the volatility of the environment and translating into predictions about future sensory input. Due to unpredictability of the environment and sensory noise the actual incoming sensory input can differ wildly from predictions, with the extent of the prediction error mirroring the content of information stored in the new piece of data. A greater prediction error is more informative about the true current state of the world than a lesser one, and vice versa. Going back to the example of the cat: finding it in a different spot than predicted by listening would certainly both add new information to our belief about its whereabouts and elicit a feeling of emotional surprise. How much of both exactly depends on the difference to the prediction: One meter would maybe even go unnoticed; fifty meters would make us seriously doubt ourselves.

The common computational quantification of surprisal is the Shannon information defined as the negative log probability of a sensory event, given the prior (Modirshanechi et al., 2022; Sedley et al., 2016). This definition also accounts for the prior's precision, eliciting greater surprisal if the prior is very precise and vice versa (Sedley et al., 2016). In case of the cat this would mean greater surprisal if we mislocated it after 30 minutes of non-stop meowing, versus a mislocalization after only two meows (see Figure 3).

Figure 3

Surprisal depends on prior precision (Yon & Frith, 2021)



Note. Priors translate into predictions about future sensory input. Surprisal, the reliability-weighted prediction error, depends on the prior's precision. The same incoming evidence will elicit more surprise if the initial prediction is very precise.

A. 2.2. Change point paradigm

A common practice to examine surprisal and Bayesian inference in dynamic environments is within a so-called change point (CP) paradigm (Krishnamurthy et al., 2017; Ma, 2019; Nassar et al., 2012). A time-series of stimuli is generated and presented to subjects. The stimuli contain a certain variability according to a set of underlying rules. As already established, in Bayesian inference subjects form expectations in the form of priors about the occurrence of stimuli, increasing in precision with every additional stimulus. The stimuli vary within the bounds of the sensory noise, which can be seen as a fixed underlying set of rules (Ma, 2019; Nassar et al., 2012). CPs introduce a second level of variability, namely volatility, changing the underlying rules and therefore rendering any expectations based on earlier observed stimuli meaningless and the prior irrelevant, even if the prior already accounted for a certain degree of variability (Knill & Pouget, 2004; Ma, 2019; Nassar et al., 2012). This essentially simulates the circumstances of a dynamic environment, as opposed to a static environment without CPs. The CPs occur with a certain probability called hazard rate. A key assumption is that subjects adapt to the occurrence of CPs by adjusting their use of priors for the possibility of a CP, essentially learning the hazard rate and accounting for it like an ideal Bayesian observer would (Clark, 2013; Knill & Pouget, 2004; Nassar et al., 2012).

The practicality of the CP paradigm for assessing surprisal is obvious: CPs evoke less probable stimuli. Surprisal is therefore greatest immediately following a CP, as this is where presented stimuli deviate the most from probabilistic predictions (Baldi & Itti, 2010), hence the prediction violation is the greatest. This improbable stimulus carries the most information for the new prior. Surprisal is therefore central in the adaptive process of dealing with a volatile environment, as it marks the process of updating expectations in reaction to a changed environment. Using a Bayesian model, surprisal can be estimated based on the probability of an occurring stimulus under the momentarily established prior (Baldi & Itti, 2010; Visalli et al., 2021).

A. 3. The P300 as a neural correlate for surprisal

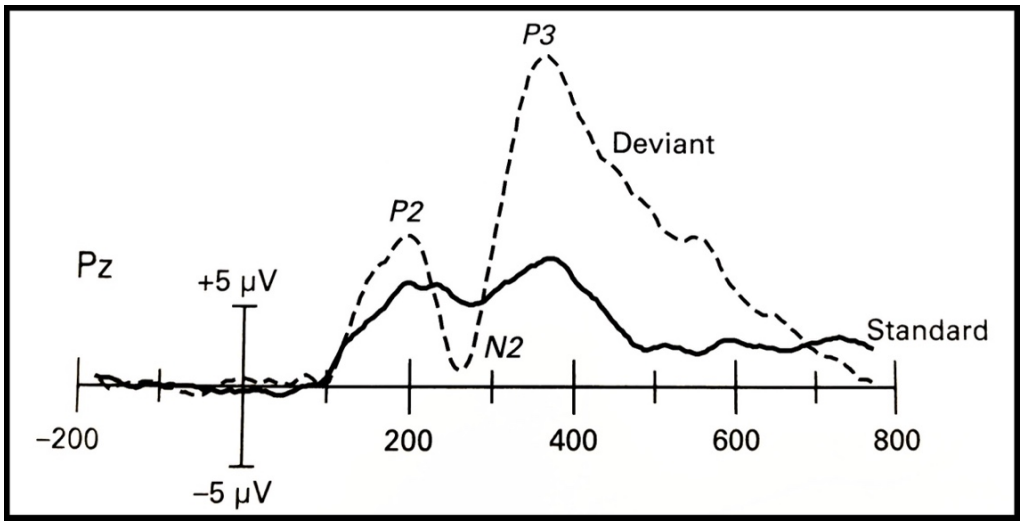
With a plethora of existing literature successfully describing human perception with Bayesian models (Yon, 2021; Yon & Frith, 2021), neurophysiological correlates for Bayesian inference are to be expected. One commonly observed neural correlate for surprisal is the P3b (Kolossa et al., 2015; Visalli et al., 2021), a subcomponent of the P300

event-related potential (ERP) (Luck, 2005; Polich, 2007). The P300 is a positive deflection peaking within a time frame of 250 – 500 ms (Polich, 2007) or even as late as 850 ms (Luck, 2005) after stimulus or event onset. The earlier and shorter frontally maximal subcomponent is known as the P3a, followed by the later and longer sustained centro-parietally distributed (parietally maximal) P3b (Luck, 2005; Polich, 2007).

Since the onset time of the P300 shows much variability, it is often simply referred to as P3, for being the third major positive deflection. Confusingly, much of the literature refers to P3 when really talking about the P3b specifically (Luck, 2005). An exemplary P3b can be seen in Figure 4.

Figure 4

P3b deflection at the Pz electrode in oddball task (Luck, 2005)



Note. An event-related potential (ERP) elicited in a visual oddball experiment at the Pz electrode. The target ('Deviant', low probability) shows a clear and greater P3b (marked as P3) in comparison to the 'Standard' (higher probability), illustrating the inverse relation between P3b and probability.

The P300 complex, especially the P3b, is among the most intensively studied ERPs, with thousands of published experiments, dating back all the way to 1964 (Chapman & Bragdon, 1964). Already early on, the P3b has been shown to be inversely related to stimulus probabilities (Donchin & Coles, 1988) and tied to a variety of hypotheses regarding event expectancy (Squires et al., 1975), uncertainty (Sutton et al., 1965) or surprise (Donchin, 1981).

Despite the long history of research, understanding of the P3b is limited, with no clear consensus about underlying neural or cognitive processes. A widely cited theory (despite scarce experimental evidence) is from Emanuel Donchin, who proposed the P3b

to be related to “context updating”, with context being understood as broad representations of the overall state of the environment (Donchin, 1981; Luck, 2005). Donchin further proposed the P300 to be reflecting a strategic process, meaning a process that happens in order to prepare for a future event, rather than dealing with a current event. He argued for P3b amplitude variability to predict later memory for the eliciting stimulus (Donchin, 1981; Luck, 2005). Some later research cautiously tied the P3b to the updating of working memory (Luck, 2005; Vogel & Luck, 2002).

Despite uncertainty about the meaning of the P3b, its responses to certain manipulations are well established and replicable. A key factor is its (inverse) relation to stimulus probability: A smaller stimulus probability leads to a larger P3b amplitude (see Figure 4). This is true both in overall probability (target in comparison to all stimuli) and local probability (target in comparison only to the preceding stimuli since the last target). The effect is exclusive to the task-defined stimulus category, not the overall stimulus; all stimuli can be unique, but in a task targeting broader categories of those stimuli the P3b amplitude will be sensitive to those categories’ probabilities of occurrence (Luck, 2005).

More recent studies on the P300 complex employing a Bayesian approach and focusing on Bayesian inference aimed to disentangle the subcomponents P3a and P3b (Kolossa et al., 2013, 2015; Visalli et al., 2021), specifically linking surprisal to P3b amplitude variability, while linking the P3a to belief updating (Kolossa et al., 2015; Visalli et al., 2021).

A. 4. Research question and hypotheses

Existing literature shows Bayesian inference to accurately describe human perception in different situations both in visual motion perception and sound localization in static environments (McLachlan et al., 2021; Yang et al., 2021), and even sound localization in adaptation to dynamic environments (Krishnamurthy et al., 2017). To my knowledge, auditory motion perception has so far not been examined as process of Bayesian inference. We are attempting to close this gap in literature by investigating whether auditory motion perception in a dynamic environment shows hallmarks of Bayesian inference on a behavioral and neurophysiological level.

We employed a change point (CP) paradigm, presenting sequences of auditory motions with random lengths and directional CPs in the horizontal plane and asking participants to judge and report the perceived motions at the end of each sequence. By

doing so, we strategically built up priors and rendered established priors obsolete by the occurrence of directional CPs within the sequence, with the goal of forcing adaptation and eliciting surprisal. The task's overall difficulty was chosen to reach sufficiently high levels of uncertainty in motion perception, giving relative weight to the priors. The task was further designed so motions switch between two opposite directions, rendering priors following a CP not only obsolete, but actively detrimental, further increasing their influence.

The hypotheses consist of two dimensions, behavioral and physiological. On a behavioral level, if humans employ Bayesian inference for auditory motion perception, we should be able to observe the influence of priors. Specifically, we expect participants' behavioral answers to reflect their usage of momentarily established priors on direction due to uncertainty. This should increase participants' accuracy when priors are more informative and decrease their accuracy directly following a CP, when the prior is detrimental. As an extension of the behavioral dimension, we expect estimates of surprisal, extracted from a Bayesian CP model fitted to the participants answers, to reflect the usage of momentarily established priors. Specifically, we expect surprisal to be highest directly following a CP, and to decrease with the accumulation of new evidence in the prior. On a physiological level, we predict these momentary estimates of surprisal to correlate with measures obtained by electroencephalography (EEG) during the motion perception task, specifically with P3b amplitude variability.

B. METHODS

B. 1. Participants

The study was conducted in the laboratory of the Acoustics Research Institute of the Austrian Academy of Sciences. The sample consists of 26 participants (13 female, 13 male), recruited via the Vienna Cognitive Science Hub. At the day of testing, participants were between 20 and 33 years old (M = 24, SD = 3.1). Participants were naïve to the purpose of the study, right-handed, spoke German or English and had normal or corrected-to-normal eyesight. The participants reported to be healthy, to have no hearing impairments (hearing aides, cochlea implants, tinnitus), to have no neuronal disorders (like epilepsy or claustrophobia), to have had no concussions, skull fractures, comas or operations on head or vascular system in the past, to have taken no mood-altering medication (psychopharmaceuticals, hormones) and to not have engaged in drug use on the day of testing and the day prior (including alcohol).

Further exclusion criteria were thresholds in the *minimal audible angle* (MAA) measurement before entering the main experiment. We had a total of 14 exclusions, 11 of which were excluded due to these thresholds (see *B. 3.5. Minimal audible angle task*). One further participant was excluded after completion of the entire experiment due to a technical error in the EEG that rendered the data unusable. Two participants dropped out between sessions without indicating a reason. The final sample size of 26 is comparable to similar studies (Krishnamurthy et al., 2017; Visalli et al., 2021).

Participants received monetary compensation (10€/h) and gave written consent after an initial introduction by the experimenter. After the experiment they were debriefed about the study's purpose.

B. 2. Experimental setup

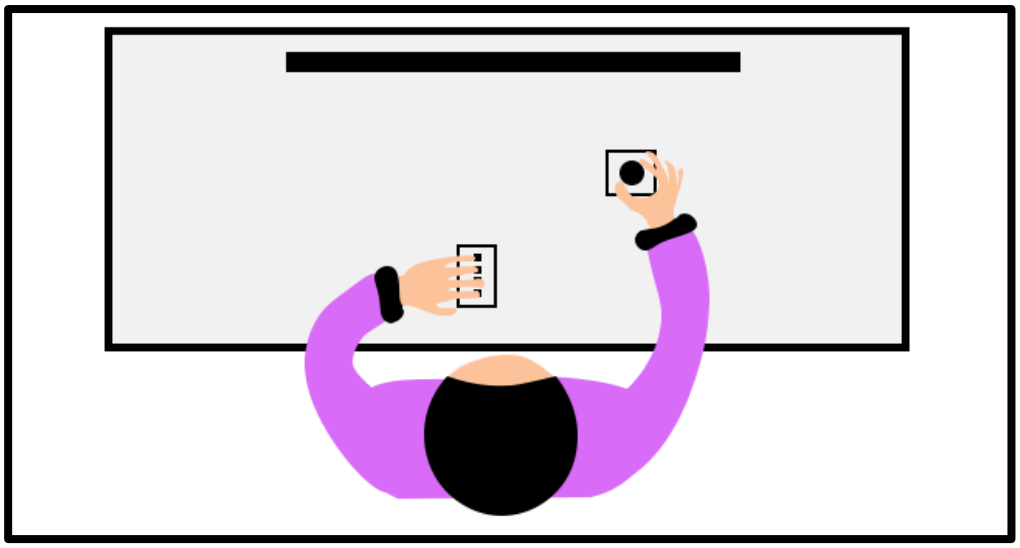
Participants were seated at a distance of 75 cm from a computer monitor (48 x 27 cm, 1.280 x 1.024 pixels, refresh rate of 60 Hz) in a dimly lit, sound attenuated room. Their heads were supported by a chin rest, preventing motion, and ensuring constant distance and eye-level. With their right hand the participants were controlling a digital rotating wheel. Their left hand was placed on a keypad with four marked vertical keys. The Keypad was placed in front of the body and could be moved by the participant to ensure a

comfortable arm position, but participants were instructed to always keep the marked keys vertical (see Figure 5).

In the second session, participants‘ brain activity was monitored via a 128-electrode EEG (actiCAP with actiCHamp; Brain Products GmbH, Gilching, Germany). During EEG we also collected eye-tracking data for a different study interested in pupil dilation, utilizing an eye tracker (Eye-Link 1000 Plus; SR Research, Osgoode, Ontario, Canada). Auditory stimuli were presented via insert earphones (ER-2, Etymotic Research, Inc., Grove Village, Illinois) at 48kHz sampling rate. The experiment was run on Windows 10 in MATLAB, using the Psychophysics Toolbox extensions (Kleiner et al., 2007).

Figure 5

Positioning of the subjects



Note. Subject's right hand is positioned on the rotating wheel, while left hand is placed on the keypad. The participants are free to move keypad and rotating wheel around to find a comfortable position but are instructed to keep their arms in this general position.

B. 3. Task and procedure

The experiment was split in two sessions that took part on two separate days with a maximum of 72 hours in between. The first session started with a familiarization task (1 block) to first introduce the general experimental setup and trial design. The same task in small alterations was then used to measure the individual MAAs (3 blocks). The session ended with one block of the main task, called last direction discrimination (LDD), which was another alteration of the same design. The second session repeated the familiarization task (1 block), followed by the LDD task (4 blocks). After the second LDD block we scheduled a mandatory break of at least 10 minutes, but participants were

offered breaks after every block throughout both sessions, to take as needed. Throughout the second session participant's brain activity was monitored via EEG.

B. 3.1. Trial overview

One trial consisted of a sequence of motions moving along the azimuth (bounds: ±40°), with a fixed elevation of 0°. One motion was presented as two distinct stimuli, marking its starting- and endpoint. When a sequence was presented the endpoint of the previous motion was the starting point of the following motion, therefore a trial always consisted of one more stimulus than motions. Every motion had two defining properties: direction and velocity. The direction was either clockwise, or counterclockwise. The velocity is the distance between the two stimuli in degrees azimuth divided by the time between two stimuli, the stimulus onset asynchrony (SOA). The SOA was kept constant; hence the distance was always proportional to the velocity. Together direction and velocity exactly defined the endpoint of a motion in relation to the starting point, hence the locations of the stimuli.

During familiarization trials and MAA trials the "sequences" always had a length of only one singular motion (two stimuli), while LDD trials also included sequences with more motions. During all experimental blocks (familiarization, MAA, LDD) the participants had the same task: indicate whether the last motion's direction was clockwise or counterclockwise in a two-alternative forced choice design (2AFC). Participants responded by rotating the digital wheel. A subsequent prompt with a question mark and one arrow in each direction asked the participants to indicate the perceived direction. After the response was given the arrows changed to indicating the direction of the given response and a vertical Likert scale appeared on screen. Participants rated the confidence in their response via vertical keys corresponding to the scale ("Very uncertain", "Somewhat uncertain", "Somewhat certain", "Very certain"). Until a response on confidence was given participants were able to change their answer on perceived direction. The confidence-response ended the trial and automatically started the next one. During stimulus presentation (incl. silent periods) the screen was grey and empty, except for a fixation dot at the center of the screen. Participants were instructed to keep fixating the dot throughout the trial to minimize eye movement, up until the response prompt appeared.

B. 3.2. Stimuli and spatialization

Each stimulus consisted of a 50 ms pink noise burst (10 ms on/off raised cosine ramps, high-pass filtered from 250 Hz with 4th order Butterworth), presented at around 80 dB instantaneous peak SPL (65 dB short-time averaged; Bruel & Kjaer 2260 Investigator sound level meter).

For stimulus spatialization, we acoustically recorded individual head-related transfer functions (HRTFs) per participant. When auditory signals travel through space they are shaped by their interaction with the individual body before they reach the ear canal, influencing timing, level, and spectral properties of the original sound for each ear differently (Li & Peissig, 2020). Human listeners learn and utilize the resulting characteristics (interaural time and level differences, monaural spectral characteristics) for sound localization. HRTFs quantify the change induced to an auditory signal by its travel from the source location to a person's ear canals (Li & Peissig, 2020; Plack, 2013). An HRTF is therefore highly unique for a person's anatomy (most importantly pinna geometry, but also size and shape of head and torso, etc.) and contains all information for sound localization for a specific point in 3D space. For simulating spatial audio, the process can be reversed by artificially applying the resulting characteristics quantified by a person's individual HRTF to an audio signal via filtering (Li & Peissig, 2020).

We placed participants in a semi-anechoic chamber (T60 = 50 ms) inside a spherical 91 loudspeaker array (E301, KEF, radius 1.2 m) with their head at its center. We inserted small microphones into their ear canals (KE4-211-2, Sennheiser) and presented them with 91 exponential sweeps (one per loudspeaker) from 20 Hz to 20 kHz within 6 seconds, multiplexed across directions. The acoustically recorded HRTFs were cleaned from acoustic influences of the chamber and its equipment during postprocessing by equalization with the transfer functions of the equipped room (measured previously at the center of the array with no human present) and shortening the HRTFs impulse responses to 5 ms. The exact multiplexing and windowing followed a previously proposed procedure by Majdak et al. (2007). The HRTFs were further spatially upsampled to obtain a 1° resolution in azimuth by following a previously proposed procedure by McLachlan et al. (2023). The resulting set of HRTFs was used to individually spatialize the stimuli. Due to equipment failure, we could not measure HRTFs for 8 of 26 participants and used generic HRTFs for them instead (Braren & Fels, 2020).

B. 3.3. Timing of sequences

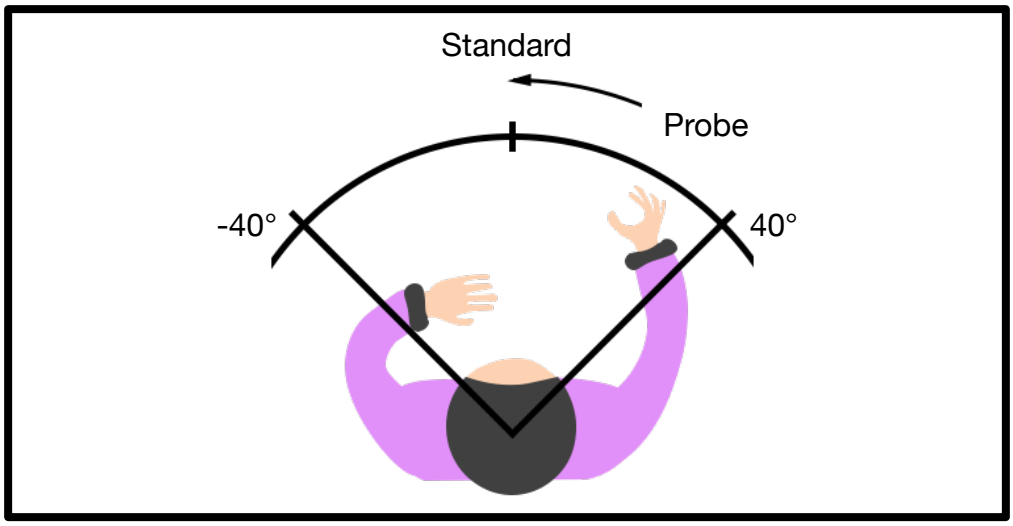
Each trial started with a variable silent lead-in period randomized between 750-1000 ms, followed by the first stimulus (50 ms), a silent interstimulus interval of 450 ms (stimulus onset asynchrony = 500 ms) and the next stimulus. This was repeated until the last stimulus, which was followed up by 1450 ms lead-out period, before the response prompt appeared.

B. 3.4. Familiarization task

The familiarization task consisted of 50 trials, done in one block. A trial consisted of one motion presented in two stimuli: a “standard” sound and a “probe” sound. The standard sound was randomly presented as starting- or endpoint of the motion and always at 0°, therefore the probe was always presented at a peripheral position in space that is proportional to the velocity (see Figure 6). The task started easy with a velocity of ±20°/SOA, successively decreasing its absolute value by 1°/SOA for ten consecutive trials. This was followed by 40 trials, randomized between ±20°/SOA and ±1°/SOA. The motion’s direction was pseudo-randomized.

Figure 6

Single-motion trial

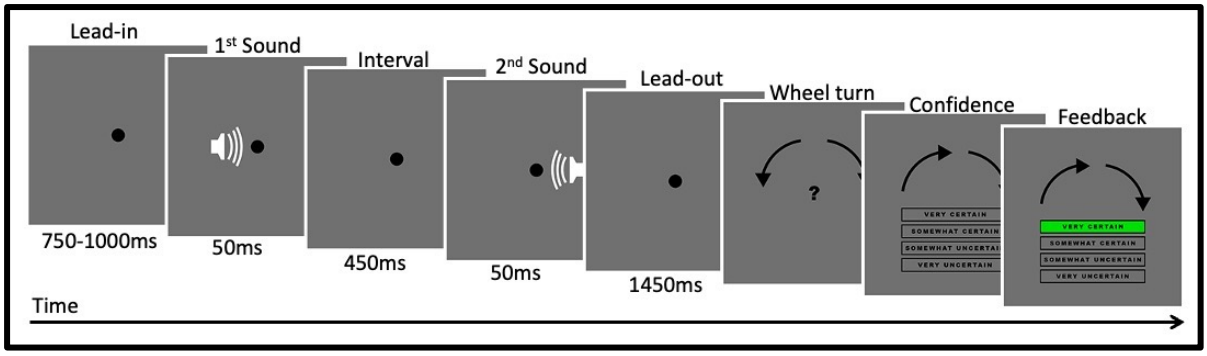


Note. Exemplary illustration of a regular trial in the familiarization task or minimal audible angle (MAA) task, or a single-motion trial in the last direction discrimination (LDD) task. In the familiarization and MAA task the stimuli were termed “probe” and “standard”. Probe was either the first or the second stimulus and on either peripheral side. Standard was always presented at 0°. The distance between, illustrated by the arrow, is proportional to the velocity.

After the responses on direction and confidence the participants were provided with color-coded feedback. The indicated confidence briefly (500 ms) lit up green for a correct and red for an incorrect directional answer (see Figure 7).

Figure 7

Familiarization task and minimal audible angle task



Note. One trial in the familiarization task and minimal audible angle (MAA) task. A random lead-in period of 750-1000 ms was followed by a first stimulus (50 ms). After a silent interval of 450 ms (SOA = 500 ms) the second stimulus (50 ms) was presented at the distance travelled during one SOA under the current velocity, relative to the first stimulus. After a fixed lead-out period of 1450 ms the participant was asked via prompt to indicate the direction (clockwise, counterclockwise). Subsequently, the participant was asked to indicate their confidence on a vertical Likert scale and was then provided with color-coded feedback (correct → green, incorrect → red). The fixation dot stayed on screen from the lead-in period until the end of the lead-out period.

B. 3.5. Minimal audible angle task

The MAA task consisted of 3 blocks with 100 trials each. Just like in the familiarization task, each trial consisted of one motion, comprised of standard and probe. The standard was set at fixed locations that changed through blocks: 0° in block 1, ±20° in block 2, and ±40° in block 3. In blocks 2 and 3 the standard was always presented on the same side for a quarter of trials (i.e. 25 consecutive trials), changing sign after the first and third quarter, but only with a 50% chance after the second quarter. During small breaks of 30 seconds separating these quarters the side of presentation for the next quarter was announced to the participant. After their responses the participants received the same feedback as in the familiarization task.

The probe stimulus was offset relative to the standard stimulus by the direction and the distance defined by the velocity. The order of standard and probe as well as the direction were pseudo-randomized. The velocities (and therefore difficulties) of the trials were adaptively set by the Psi-marginal algorithm (Kontsevich & Tyler, 1999; Prins, 2012, 2013; using code from https://github.com/lacerbi/psybayes). We obtained estimates of the minimal distance between two sounds that participants need to correctly infer the direction.

We call these estimates MAAs. Three MAAs are obtained per participant, at 0°, ±20° and ±40° azimuth, with both hemispheres considered equal.

If the participants exceeded 3.3° at 0° azimuth ($MAA_{0°}$), or 10° at either ±20° or ±40° azimuth, they were offered to repeat the block, once for each eccentricity. If the MAA still exceeded the threshold after one repetition, the participant was excluded from the experiment. The MAAs were used for the generation of sequences (see *B. 5.1. The generative function*) to set, among other things, the level of difficulty for the LDD task. For MAAs larger than these thresholds we could not generate suitable sequences without exceeding the bounds of space (±40°). Outside the bounds of space MAAs can increase rapidly (Stevenson-Hoare et al., 2022), hence the spatial limits.

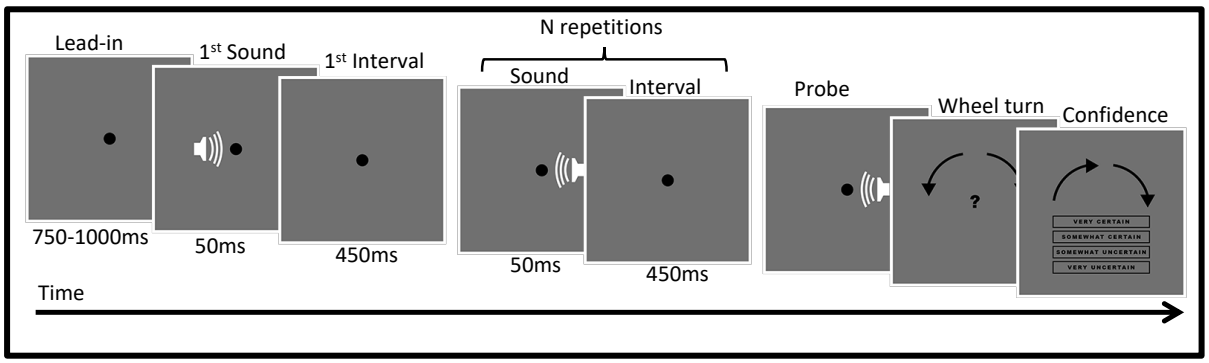
B. 3.6. Last direction discrimination task

The LDD task consisted of 4 blocks with 50 trials each. Each trial consisted of a sequence of consecutive motions that were pseudo-randomized according to a set of rules (see *B 5.1. The generative function*). The sequence started with a stimulus at a random location within the bounds of the space (±40° for all sounds of the trial) and was followed by one stimulus per motion in the sequence, presented at distance proportional to the velocity relative to the previous stimulus. The number of stimuli in a sequence was therefore the number of motions plus one. Each velocity was sampled independently from a normal distribution. The motion's direction switched at the CPs with a hazard rate of $H = .2$. Each motion could be the last of the sequence with a chance of 10%, therefore attention had to be paid throughout. The sequence could be as short as one motion (two stimuli). One sequence could contain multiple CPs, therefore occurences of CPs could not be anticipated by the length of the sequence.

The participant's task remained the same as in the previous tasks: indicating whether the last motion's direction was clockwise or counterclockwise. Crucially, the difference to the previous tasks was the length of sequence: So far each "sequence" only had the length of a single motion, hence all stimuli of the trial defined the last motion's direction. Now sequences could have lengths from 1 up to 50 motions, which all had the same chance of being the last one, with only the final two stimuli defining the last motion's direction. The participants indicated their directional answer and confidence response in the same fashion as in all tasks but received no feedback (see Figure 8).

Figure 8

Last direction discrimination task



Note. One trial in the last direction discrimination (LDD) task. Longer sequences were possible, compared to a trial in the familiarization or multiple audible angle (MAA) task. The answering procedure stayed essentially the same, but participants no longer received feedback.

B. 4. Relevant measures and variables

From the LDD task we extracted three relevant variables: (1) behavioral responses, whether the directional judgement per trial was correct or incorrect. This is further expressed in accuracy over specified subsets of trials. (2) Momentary estimates of surprisal, a latent variable extracted per motion from a Bayesian model fitted to the behavioral responses (see *B. 5. Model*). (3) Cluster amplitude, a quantification of the event-related potential in the EEG (see *B. 6. Electroencephalography*) per motion. A further relevant variable is sound after change point (SAC) level, an enumeration of the consecutive stimuli resetting to 1 after every CP.

B. 5. Model

For this experiment we used a generative function for the generation of sequences and, building on that, a Bayesian change point (BCP) model inferring motion directions for the extraction of latent variables, developed by Dr. David Meijer at the ÖAW Acoustics Research Institute, Vienna. A detailed description of the model can be found in the appendix.

B. 5.1. The generative function

The generative function is best understood as a sampling process in five steps. (1) First, a starting location $x_{t=0}$ (t = timepoint) was sampled from a uniform distribution within the spatial bounds (±40°). All following stimulus locations x_t were set by a sequence of velocities v_t starting from $x_{t=0}$, therefore defined as $x_t = x_{t-1} + v_t$.

(2) In a second step, a first direction $d_t \in \{-1, +1\}$ was sampled at random, with -1 indicating a clockwise direction and +1 indicating a counterclockwise direction.

(3) The third step sampled CPs from a binomial distribution with a success probability of the hazard rate, $H = .2$. A CP occurred when $cp_t = 1$, therefore on average in one out of five motions, and it changed the sign of the direction d_t. For reasons related to the analysis every first velocity, $v_{t=1}$, was defined to have followed a CP.

(4) The fourth step sampled the single velocities from a normal distribution for every timepoint $t \geq 1$. In this step the task's difficulty was adjusted for every participant individually by utilizing their personal MAAs. The MAA_{0° was used to define the normal distribution's variance ($\sigma^2 = {MAA_{0^\circ}}^2$) and mean ($\mu = 3 \times MAA_{0^\circ}$). The mean's sign was set by multiplication with the direction d_t. Consecutive motions were likely to have the same direction, but not the same velocity.

(5) In a fifth and last step the end of the sequence was defined by sampling from a binomial distribution with a 10% success probability. Every motion had a 10% chance of being the last one, forcing participants to attend to every stimulus equally.

B. 5.2. Controlling of sequences

To ensure fair distributions of trial lengths, hazard rates, directions, locations, velocities and SAC levels the sequences were controlled for these parameters after generation. In a brute-force procedure sequences were generated for entire blocks of 50 trials at a time until the entire set of sequences met the following criteria:

(1) Sequence lengths were separated into two categories, with several *bins* each: 20 "short" trials with separate bins for lengths of 1, 2, 3, 4, and 5 motions and 30 "long" trials with separate bins for lengths of 6-7, 8-10, 11-14, 15-20, and 21-44 motions. There were exactly four trials in each of the five "short" bins and exactly six trials in each of the five "long" bins.

(2) Hazard rates for long trials were kept between $H = 0.1$ and $H = 0.3$ for each trial to stay close to the intended hazard rate for CPs of $H = 0.2$. Short trials were kept unidirectional, therefore did not contain any CPs.

(3) SAC levels for short trials are equal to their length in motions, as they were unidirectional. Therefore, SAC levels 1-5 all occurred equally often in short trials. In long

trials the SAC levels 1-5 occurred once in each of the “long” bins, therefore also equally often.

(4) The last velocity is seen as a proxy for the reliability of the likelihood function for the task-relevant direction. Velocities were grouped for direction and relative speed in an equal partition of the generative normal distribution for velocities and distributed fairly across SAC levels and trial lengths. Each SAC level in short trials contained two clockwise and two counterclockwise directions, each split in one relatively fast and one relatively slow velocity. Each SAC level in long trials contained three clockwise and three counterclockwise directions, each split in one relatively fast, one relatively slow and one medium velocity.

(5) The lateral space (±40°) was divided in five equal regions. Each region contained the last stimulus of a trial equally often, separately for long and short trials. Each region contained one of both directions and one of both relative speed groupings (fast, slow) for short trials and at least one of both directions and one of all three relative speed groupings (fast, medium, slow) for long trials.

B. 5.3. The Bayesian directions change point model

Participants were briefed superficially about the concept of sequences of motions, but without mention of the specific hazard rate or any experimental parameters. They only had access to the result of the generative function, the sequence of locations of the stimuli x_t. Accounting for sensory noise, these locations are corrupted, leaving the participants only with internal estimates λ_t of the true locations.

According to the ideal Bayesian observer theory (Ma, 2019; Manning et al., 2023) we assume that participants established an internal estimate of the hazard rate and the velocity distribution $N(d_t \times \mu, \sigma^2)$ during the first LDD block in session 1 (Knill & Pouget, 2004; Nassar et al., 2012). It is assumed that participants are able to use those estimates to reverse the generative function for perceptual inference by generating a likelihood ratio for whether this subjective estimate is the result of a motion with a clockwise or counterclockwise direction. This likelihood ratio is henceforth multiplied with a prior ratio to determine the posterior ratio, the perceptual inference for the current direction. At the beginning of the sequence the prior is simply a uniform distribution since there should not be any existing expectations about direction whatsoever. For every additional step in the sequence, the previous posterior is adjusted for the possible occurrence of a CP and then

used as the next prior, quantifying the accumulation (and decay) of information. The last posterior ratio of the sequence determines the ideal observer's judgement whether the last motion's direction is clockwise or counterclockwise.

The participant's sensory noise is not accessible to us and had to be estimated. We applied a Gaussian jitter, linearly increasing with increasing azimuth, to the true sound locations, to simulate the internal noise-corrupted estimates of the true sound locations. Based on different simulated internal estimates and the behavioral answer, 1000 possible perceived sequences of motions were simulated per trial. The model was fitted across these simulations to compute the likelihood, prior and posterior for every point in the sequence and therefore the probability of a clockwise or counterclockwise final response.

B. 5.4. Computation of surprisal

For every simulated sequence, surprisal can be computed based on likelihood ratio and prior ratio for every motion in the sequence. Across all simulations we averaged these surprisal values to obtain a single value, *expected surprisal,* per motion. Equivalent to comparable literature (Kolossa et al., 2015; Sedley et al., 2016; Visalli et al., 2021), we defined surprisal as the Shannon information, the negative log probability of a sensory event (perceived direction) given the prior. Surprisal (S) at a certain time point (t) is therefore computed as follows (Eq. 2), based on the current prior (P), likelihood (L) and direction (clockwise = $clock$, counterclockwise = cnt):

$$S_t = -log_2(P_{cnt,t} * L_{cnt,t} + P_{clock,t} * L_{clock,t}) \quad \text{(Equation 2)}$$

An in-depth explanation of the BCP model and the computation of surprisal by Dr. David Meijer can be found in the appendix.

B. 6. Electroencephalography

B. 6.1. Preprocessing

The preprocessing of the EEG data was entirely done in MATLAB 9.11.0 (Version 2021; The MathWorks, Natick, MA) and EEGLAB 2023.0 (Delorme & Makeig, 2004). We recorded 128 scalp channels. The location of FCz was used for the reference electrode. Recorded trigger markers in the data were re-aligned with time points of stimulus presentation to compensate for different hardware delays of audio signal and triggers. During stimulus presentation we simultaneously sent three audio signals, two going to the participant as a stereo stimulus, one to be recorded with the EEG as a simple square

wave, marking the stimulus duration. The first step synchronized the trigger markers to the exact stimulus onset, marked by the third audio signal. The data, initially recorded at a sampling rate of 1000 Hz, was then downsampled to 250 Hz. We removed line noise utilizing *nt_zapline* from ZapLine Plus (Klug & Kloosterman, 2022) and high-pass filtered it with a cut-off at 0.25 Hz, using EEGLAB's default filter, a linear (zero-phase) non-causal FIR filter. No low-pass filtering was applied.

The data was visually inspected and noisy stimuli and channels were removed manually. An average of 1.3 channels per participant were removed, with a maximum of 5 channels in 3 participants. An average of 5.9% of sounds per participant were removed, with a maximum of 26.3% in one participant. The participant's data set still includes over 1700 sounds and was therefore not excluded. The rejected channels were interpolated and FCz was recovered while re-referencing the data as average reference. The data was then epoched from 0-500 ms after stimulus onset. Subsequently, principal component analysis (PCA) and independent component analysis (ICA) were applied to a copy of the data that had undergone the same preprocessing steps as the original data, but with a cut-off of 1 Hz instead of 0.25 Hz for the high-pass filter. The PCA was applied due to rank deficiency, as the 128 electrodes do not contain independent signals after interpolation and recovery.

The ICA-weights retrieved from the copy were applied to the original data set and horizontal eye movements as well as blink- and heart-related components were rejected if classified as such with *ICLabel* (Pion-Tonachini et al., 2019) with an accuracy greater than 80%. In a last step, the data was baseline-corrected over the whole epoch, from 0 – 500 ms after stimulus onset.

B. 6.2. Cluster-based permutation

Instead of pre-defining an area of interest in the spatiotemporal (channels, time) structure of the EEG-signal, I used a cluster-based permutation approach, utilizing the *Fieldtrip* toolbox (Oostenveld et al., 2011). This data-driven approach is designed to identify clusters in the spatiotemporal structure that differ significantly across certain conditions, in this case SAC levels.

Comparisons between conditions can be made for every point in space and time, driving the number of total comparisons up into the thousands. This is a classical multiple comparisons problem which makes it impossible to control for the family-wise error rate

(FWER). The cluster-based permutation approach works around this problem (Pernet et al., 2015), by performing only one final test per cluster.

First, a repeated measures ANOVA was conducted over all 5 SAC levels for every point in the spatiotemporal EEG-structure (Epoch: 0 – 500 ms, 128 electrodes). The result is a spatiotemporal structure of the same dimensions containing F-values for each point in the structure. All significant F-values ($\alpha = 0.05$) were selected, clustered, and summed per cluster. A cluster was defined as connected sets, with a minimum of 3 spatially adjacent electrodes and a minimum of 2 temporally adjacent samples.

The *Monte Carlo* method (Kroese & Rubinstein, 2012) was used to assess the significance of the observed summed cluster-values by comparing them against a permutated reference distribution of summed cluster-values simulating no effect between conditions (1000 permutations from collapsed SAC levels, $\alpha = .05$). Bonferroni correction was applied to account for multiple significant clusters.

Significant clusters provide a fixed set of positions as a spatiotemporal matrix, marking differences between the SAC levels in the EEG-data. Note that while this method is somewhat exploratory towards the exact spatiotemporal location of effect, it was only conducted over SAC levels and did not make use of any outcome measures from the BCP model (i.e. surprisal). The result merely defines the channels and time frames (per channel) that were then kept constant over all sounds, trials and participants for further statistical analysis.

B. 6.3. Cluster amplitude

The spatiotemporal matrix of a significant cluster allows to extract data points from the same positions in time and space from each epoch. The extracted values of the epoch were averaged over both dimensions (channel, time). The outcome variable of the EEG-data is therefore the average amplitude of the cluster in microvolts, per sound in every trial, henceforth simply referred to as *cluster amplitude*.

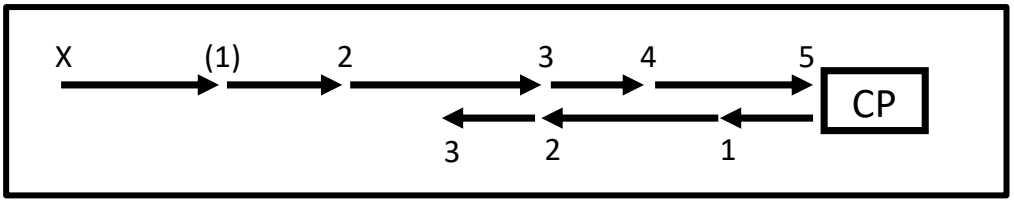
B. 7. Statistical analyses

Statistical analyses were conducted in MATLAB (Version 2022; The MathWorks, Natick, MA). The first sound of each trial was ignored for all analyses. The second sound of each trial was included for analyses regarding surprisal and cluster amplitude, but not for analyses regarding SAC level. Since CPs refer to changes in direction and the second

sound only completes the first motion, the third sound is the earliest possible CP; therefore, the SAC levels start in every trial at the third sound with a SAC level of 2 (see Figure 9).

Figure 9

Exemplary sequence with one change point and sound after change point levels



Note. First stimulus is ignored, from the second stimulus ongoing sound after change point (SAC) levels are numerated as if the first sound was a change point (CP), hence SAC level 1 for the second stimulus. A trial is not limited to one CP.

I ran two analyses of variances (ANOVA) to test whether the accuracies or surprisal differ significantly over SAC levels. For each participant I calculated the accuracy for the SAC levels 1-5 by dividing the number of correct trials by the total number of trials, per SAC level. To test whether accuracies differ over SAC levels I ran a one-way ANOVA on the accuracies for SAC levels and participants, using a significance threshold of $p < .05$. For surprisal I calculated the mean surprisal for each SAC level 1-5, for each participant. To test whether surprisal differs over SAC levels I ran a one-way ANOVA on the surprisal means for SAC levels and participants, using a significance threshold of $p < .05$. For cluster amplitudes I did not run any further statistical analysis, as cluster amplitudes are defined by their difference over SAC levels, implemented in the cluster-based permutation.

To assess the main research question, I performed linear regressions for surprisal on cluster amplitude. For each participant I z-scored both surprisal and cluster amplitude, regressed surprisal against cluster amplitude and calculated the regression's slope and intercept. To test if surprisal significantly predicts the cluster score I performed a one-sample t-test with a significance threshold of $p < .05$ on the individual slopes.

C. RESULTS

C. 1. Effects of SAC level

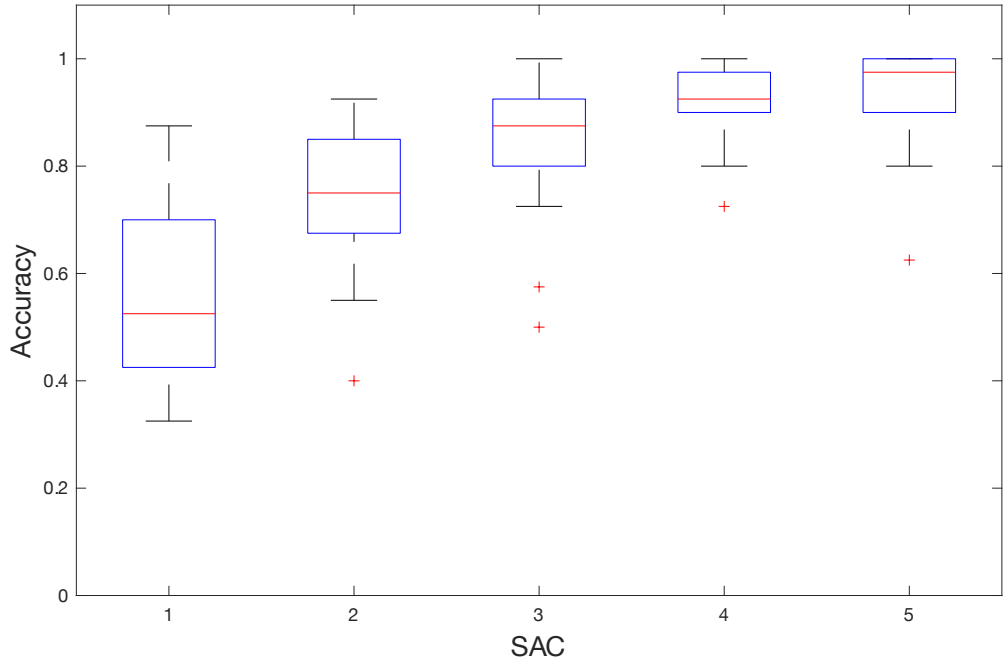
Under the Bayesian hypothesis we expected information in the prior to accumulate until occurrence of a CP. SAC level 1 represents the first stimulus after a CP, this marks a situation in which the present prior just got rendered irrelevant and points in the wrong direction, making it detrimental to solving the task. SAC level 1 marks a break upon from which a new prior is built up with the increase of SAC level.

C. 1.1. Accuracy

As an indicator of Bayesian inference, we expected accuracy to increase with accumulated information, hence SAC level. A one-way ANOVA over the SAC levels 1-5 is significant ($F_{4,125}$ = 50.72, p < .001, η^2= .62) and Figure 10 shows individual mean accuracies increase as a function of SAC level. Accuracy is defined as the portion of correct answers of all trials per participant and SAC level, therefore ranges from 0 to 1, with 0.5 being chance level.

Figure 10

Accuracy per sound after change point level



Note. Accuracies per participant increase as a function of sound after change point (SAC) level. Rounded to two decimals the mean (*M*) accuracies and standard errors of mean (*SEM*) are: SAC 1, *M* = .55 (*SEM* = .03), SAC 2: *M* = .74 (*SEM* = .02), SAC 3: *M* = .85 (*SEM* = .02), SAC 4, *M* = .92 (*SEM* = .01), SAC 5, *M* = .94 (*SEM* = .02).

Participants seemed to be biased towards the hemisphere the last direction was presented in. I quantified the probability of an incorrect answer to be congruent with the hemisphere of presentation per participant (answer “clockwise” when “counterclockwise” would have been correct for presentation in right hemisphere, ergo “clockwise” being seen as congruent to “right”, and vice versa). The probabilities range from .45 to .98 (M = .77, SEM = .03), with 20 participants exceeding .70.

To test whether participants answered according to their perceived direction like instructed, or rather just answered according to the hemisphere the last stimulus was presented in, I calculated a measure of overall congruency per participant and compared it to their overall accuracy. Congruency quantifies a possible answering pattern according to hemisphere presentation. It can be understood as an “accuracy”, if the instruction was misunderstood to indicate in which hemisphere the last sound was presented in, as opposed to which direction the last motion had. For congruency a trial with a last sound in the right hemisphere is counted as correct, if indicated with a clockwise turn, and vice versa. Both congruency and accuracy are calculated over all trials, not separating by SAC levels.

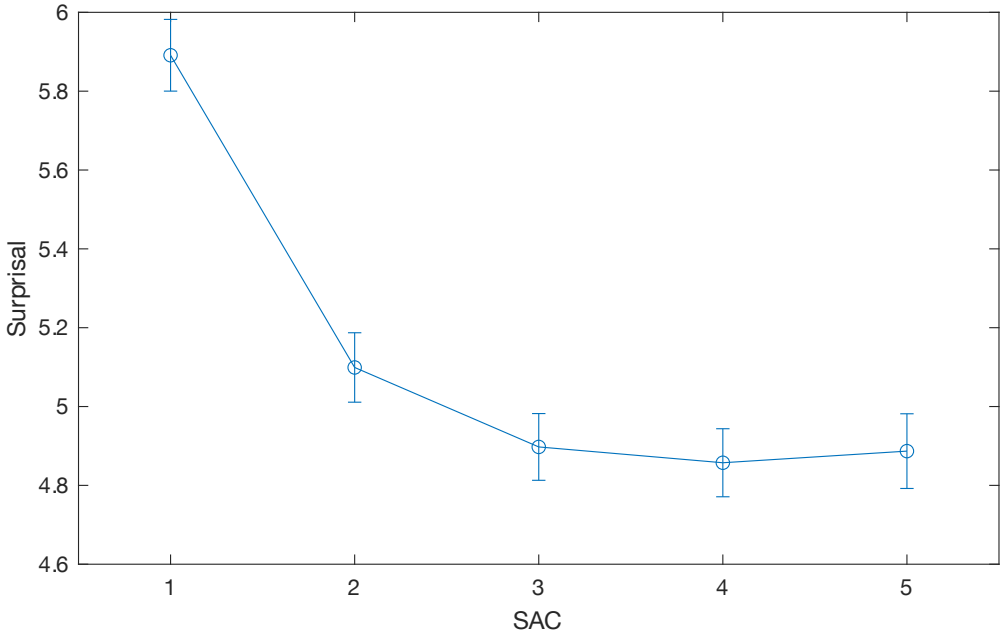
Mean congruency is .61 while mean accuracy is .80. A two-sample t-test across participants revealed a significant difference (t = 9.49, p < .001, df = 25). I conclude that participants seem to have understood and followed instructions but tended to answer according to hemisphere when the task got too difficult, hence the bias in the incorrect answers.

C. 1.2. Surprisal

As an indicator of Bayesian inference, we expected surprisal to decrease with accumulated information, hence SAC level. A one-way ANOVA over individual surprisal means and SAC levels 1-5 is significant ($F_{4,125}$ = 24.22, p < .001) and Figure 11 shows individual mean surprisal to decrease as a function of SAC level.

Figure 11

Surprisal as a function of SAC



Note. The graph shows means (M) of individual surprisal means per SAC level and the standard error of means (SEM). All values are rounded to two decimals. SAC1: *M* = 5.89 (*SEM* = 0.09), SAC 2: *M* = 5.10 (*SEM* = 0.09), SAC 3: *M* = 4.90 (*SEM* = 0.08), SAC 4: *M* = 4.86 (*SEM* = 0.09), SAC 5: *M* = 4.89 (*SEM* = 0.09).

C. 2. Clusters

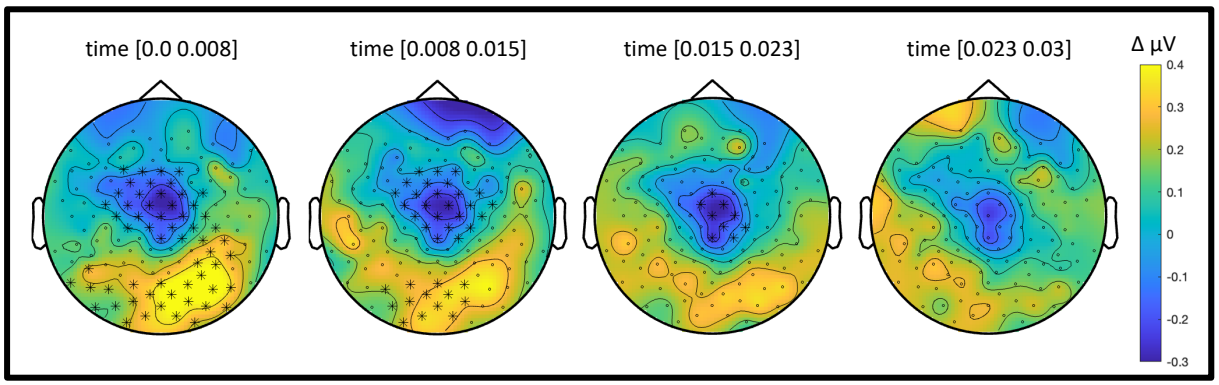
The cluster-based permutation found two significant clusters that differ significantly over SAC levels. One cluster ranges from 0-20 ms, henceforth simply referred to as early cluster, and one cluster ranges from 370-440 ms, henceforth referred to as P3b cluster.

C. 2.1. Early cluster

The early cluster (p = 0.017) appeared in the time period from 0-0.02 ms in 57 electrodes (Fz, F3, FC1, C3, P7, O1, Oz, O2, P4, CP6, CP2, Cz, C4, FC2, F1, FC3, C1, PO7, PO3, POz, PO4, PO8, P6, CPz, CP4, C2, FC4, F2, FFC1h, FCC3h, CCP1h, P9, PPO9h, PO9, O9, OI1h, OI2h, O10, PO10, PPO10h, TPP8h, CCP6h, CCP2h, FCC4h, FFC2h, FFC3h, FCC1h, FCC5h, CCP3h, POO9h, POO1, POO2, PPO6h, POO10h, CPP6h, CCP4h, FCC6h, FCC2h, FCz). This characterizes the maximum extent of only a part of the cluster, as the cluster starts before 0ms and only ends inside the epoch (0 – 500 ms). Since the cluster is incomplete and ends long before the earliest onset points for the P300 it will not be taken further into account. A scalp map showing the cluster and its evolution through time as well as activity differences between SAC levels 1 and 2 can be seen in Figure 12.

Figure 12

Scalp maps of the early cluster



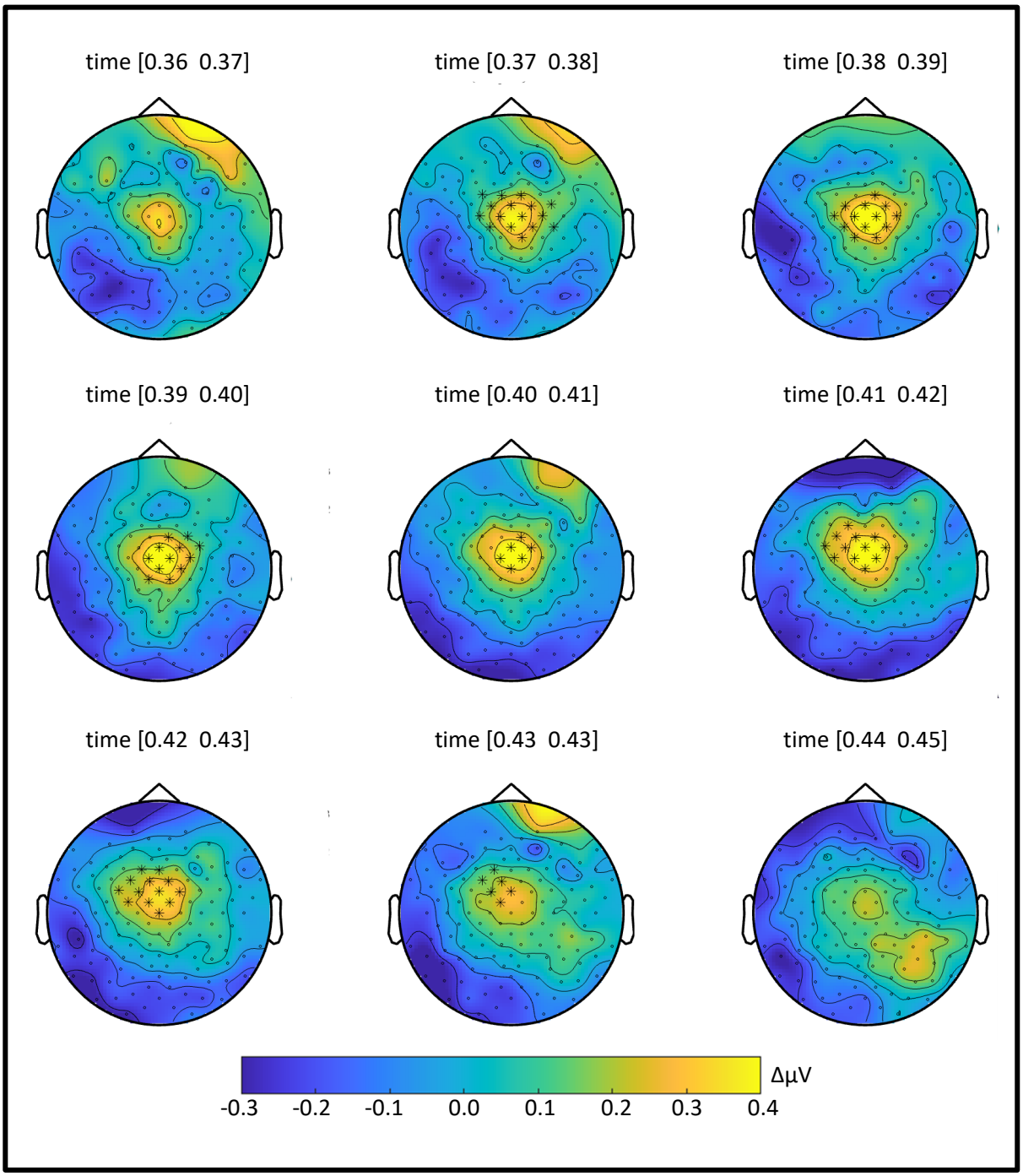
Note. Colormap visualizes activation for sound after change point (SAC) level 1 as a difference from SAC level 2, highlighted channels visualize significant channels (SAC levels 1-5).

C. 2.2. P3b cluster

The P3b cluster (p = 0.036) ranges from 370-440 ms over 21 central electrodes (Fz, FC1, Cz, C4, FC2, F1, FC3, C1, C2, FC4, FFC1h, FCC3h, CCP1h, CCP2h, FCC4, FFC2h, FFC3h, FCC1h, FCC2h, FFC4h, FCz), although this only characterizes the maximum extent of the cluster. Some of the named electrodes are only significant for short periods of time while others are significant for longer periods, some even extending over its entirety. For example, Cz, FC1, FC2, FFC2h, FCC1h, FCC2h and FCz are each significant for >50% of the cluster's timeframe (17 temporal samples of 4 ms), while Fz, C4 and FFC4h only turn significant for the duration of one temporal sample. Note that FCz, the only electrode significant over the entirety of the cluster, is not actually physically measured, but digitally restored (see *B. 6. 1. Preprocessing*). For quantification and statistical analyses, the cluster is used in its full extent. Due to the timing and spatial properties of the cluster it fits perfectly into common definitions of the P3b (Luck, 2005; Polich, 2007), hence named P3b cluster. Scalp map showing the cluster and its evolution through time as well as activity differences between SAC levels 1 and 2 can be seen in Figure 13.

Figure 13

P3b cluster as scalp maps over time



Note. Scalp maps visualizing the difference between sound after change point (SAC) level 1 and SAC level 2, highlighted channels visualize significant channels (calculated over SAC levels 1-5).

The cluster mask used to quantify the P3b cluster amplitude as a single value per sound can be seen in Figure 14. It holds the significant spatiotemporal positions as a matrix of binary values for temporal samples and channels, in the same format as the EEG data, reduced to only the electrodes contributing to the cluster.

Figure 14
P3b cluster mask

Elec. / *Ms*	*372*	*376*	*380*	*384*	*388*	*392*	*396*	*400*	*404*	*408*	*412*	*416*	*420*	*424*	*428*	*432*	*436*
Fz	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
FCz	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FC1	0	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	1
Cz	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0
C4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FC2	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0
F1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
FC3	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
C1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
C2	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
FC4	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
FFC1h	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
FCC3h	0	1	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
CCP1h	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
CCP2h	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
FCC4h	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
FFC2h	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	0	0
FFC3h	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
FCC1h	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
FCC2h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
FFC4h	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

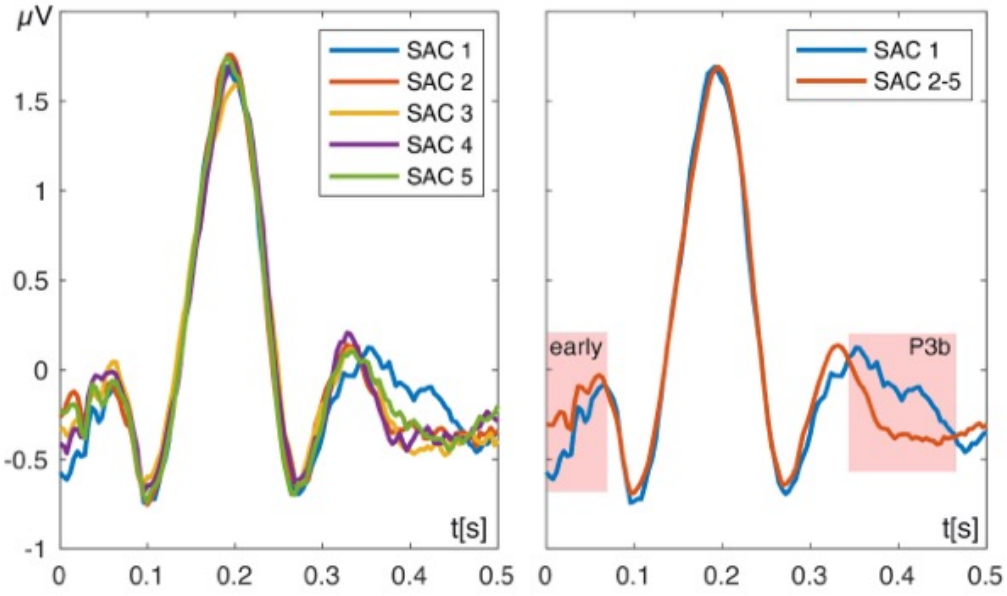
Note. The Pb3 cluster translated into a mask in the same spatiotemporal format as the EEG-data. The EEG-amplitude (in microvolts) is multiplied with the cell-value, therefore erased in grey cells and kept in white cells. The cluster amplitude is a result of applying this mask to a single epoch and averaging across both dimensions.

C. 3. Cluster event-related potentials

Figure 15 shows the ERPs plotted as a grand average separately per SAC level, showing the difference clusters described earlier. Comparing SAC levels 1 and 2, which show the expected largest differences across accuracy and surprisal so far, we see a later and longer sustained peak in SAC level 1. The ERP waveform shows both the early and the late cluster in their respective time periods.

Figure 15

Cluster ERPs for SAC levels 1-5 and 1-2

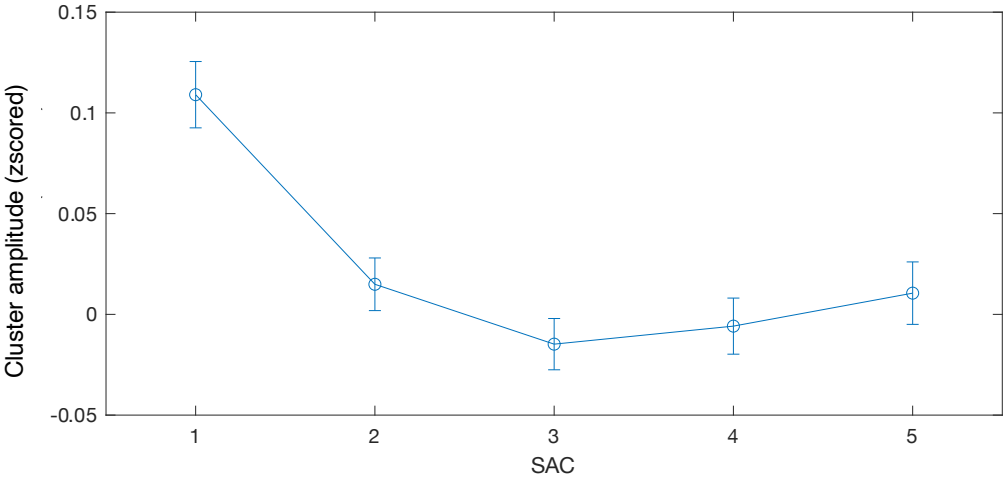


Note. Event-related potentials (ERPs) are shown as a grand average per sound after change point (SAC) level (left) and for SAC level 1 and 2-5 respectively (right) averaging over participants and electrodes significant in the Pb3 cluster: Fz, FC1, Cz, C4, FC2, F1, FC3, C1, C2, FC4, FFC1h, FCC3h, CCP1h, CCP2h, FCC4, FFC2h, FFC3h, FCC1h, FCC2h, FFC4h and FCz.

The cluster amplitude decreases as a function of SAC (see Figure 16). As expected under the Bayesian hypothesis, the biggest difference is seen from SAC level 1 to 2. Note the similarity to surprisal as a function of SAC level, as seen in Figure 11.

Figure 16

P3b Cluster amplitude as a function of SAC level



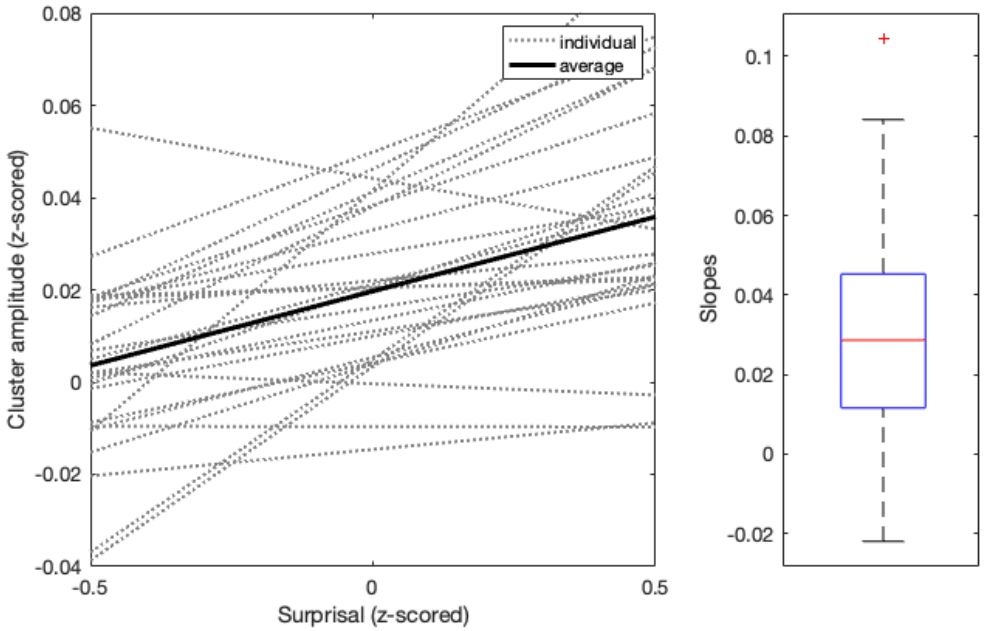
Note. Cluster amplitudes of the P3b cluster decrease as a function of sound after change point (SAC) level. The graph shows the mean (*M*) of individual means of z-scored cluster amplitudes per SAC level and the standard error of means (*SEM*). All values are rounded to two decimals: SAC 1: *M* = 0.11 (*SEM* = 0.02), SAC 2: *M* = 0.02 (*SEM* = 0.01), SAC 3: *M* = -0.02 (*SEM* = 0.01), SAC 4: *M* = -0.01 (*SEM* = 0.01), SAC 5: *M* = 0.01 (*SEM* = 0.02).

C. 4. Relationship between surprisal and cluster amplitude

The main research question is whether the P300 functions as a neural correlate for surprisal in an auditory motion perception task. Given the respective relations of surprisal and cluster score with SAC level I expected a clear link between surprisal and cluster score. I performed linear regressions for normalized surprisal on normalized cluster score for every participant across all motions and extracted the regression coefficient and intercept for every participant. Histograms and scatter plots of the residuals suggest a linear relation in each participant. A one-sample t-test on the regression coefficients revealed a significant difference from zero (t = 5.63, p < .001, df = 25). The mean regression coefficient is .03 (*SEM* = .01) with a 95% CI [.02, .04]. The mean predictive power of surprisal on single-trial P3b cluster amplitude is R^2 = .002.

Figure 17

Regressions and slopes



Note. Individual and average regression lines (left) and regression slopes (right).

D. DISCUSSION AND LIMITATIONS

The present data confirms all three of our hypotheses and is coherent with the general idea of Bayesian inference, as well as the specific idea of Bayesian perceptual adaptation as formulated in the Bayesian CP model by David Meijer (see appendix). The data shows a clear influence of postulated momentarily established directional priors on participants' accuracy. The extracted surprisal reflects this utilization of priors, visible as high surprisal directly following a CP and the following decrease in surprisal with accumulated evidence for the new direction of motion. The EEG data further suggests a neurophysiological marker of Bayesian perceptual adaptation in form of P3b ERP amplitude variability.

The results in this thesis can only be seen as indications in favor of Bayesian perceptual adaptation, as it represents one possible model explaining the data. As previously stated, alternative models explaining perceptive processes exist but have not been included in the present thesis. Any conclusion about Bayesian perceptual adaptation in contrast to other possible frameworks of perception would need a direct comparison between the respective models, which would have exceeded the possibilities of this thesis by far.

D. 1. Accuracy and surprisal reflect momentarily established priors

Both accuracy and surprisal show the expected dependency on SAC level coherent with the idea of Bayesian inference.

D. 1.1. Accuracy depends on occurrence of change points

Participants were expected to show the influence of priors in their accuracy when judging the perceived last motion in a sequence of motions. Specifically, the expectation was an increase in accuracy when priors are more informative and a decrease directly following a CP, when the prior is detrimental. This can be clearly seen in relation to SAC level (see Figure 10).

We defined SAC level, an enumeration of the occurred stimuli since the last CP, as a simple stand-in for priors. An exemplary SAC level of 5 refers to the fifth stimulus since the last CP, indicating a prior with four consecutive preceding motions in the same direction, plus one last motion, which is the current likelihood. An exemplary SAC level of 1 indicates one single motion since the last CP, which is the current likelihood, with all

information stored in the current prior still being built upon motions in the opposite direction, making it detrimental to solving the task.

This definition does not capture the nature of priors in its full extent, as it ignores the amount of accumulated evidence *leading up* to the last CP, especially relevant in SAC level 1. A postulated prior before an SAC 1 stimulus could theoretically be built upon any number of motions in the opposite direction, depending on the second to last CP. This is not to be seen as a flaw: SAC level is a relatively simple and clear variable that allows for analyses with few assumptions, independently from our specific Bayesian model. The analyses were done before any fitting process and can be interpreted within a general understanding of Bayesian inference. I view it as a strength of this thesis to include a model-free analysis to complement the later model-based analysis.

In Figure 10 it is clearly visible that participants judgements on direction were not independent judgements of the current motion's direction but depended strongly on previous motions and the occurrence of CPs. As to be expected within Bayesian perceptual inference, accuracy increased with SAC levels, indicating usage of a prior to reduce uncertainty and increase perceptual acuity. SAC level 1 represents a special case, where the prior is detrimental and integration with the current sensory input results in failure due to changes in the environment. This is reflected in the sudden drop in SAC level 1 to an accuracy close to chance level.

D. 1.2. Surprisal depends on occurrence of change points

Surprisal was predicted to reflect momentarily established priors on direction, when judging the perceived last motion in a sequence of motions. Specifically, we expected surprisal to be lower for higher SAC level and vice versa. This relationship can be clearly seen in Figure 11, when shown as a function of SAC level.

Under Bayesian inference we expected priors to build up over SAC levels by successively decreasing uncertainty with every new piece of evidence, or stimulus in the same direction (de Lange et al., 2018; Knill & Pouget, 2004; Ma, 2019). A stronger prior that represents the true state of the world better means less new information is transported in new (but coherent) evidence and the prediction about further (coherent) incoming evidence is more accurate. Hence, surprisal is expected to be larger in early SAC levels compared with later ones. SAC level 1 marks the case where the current prior is detrimental, which leads to the highest surprisal.

D. 2. Accuracy and surprisal possibly reflect Bayesian perceptual adaptation

When assessed as a function of SAC level, both accuracy and surprisal show quick recoveries from the effects caused by the occurrence of a CP, indicative of Bayesian perceptual adaptation.

D. 2.1. Accuracy quickly increases after occurrence of a change point

A closer look on Figure 10 reveals the sharpest increase of accuracy to be between SAC levels 1 and 2, right after occurrence of a CP. Generally, the increase in accuracy is visibly strong at the beginning and slows down with every additional motion in the same direction.

This is coherent with the idea of Bayesian perceptual adaptation as to be expected from an ideal Bayesian observer accounting for possible changes in the environment. A simpler observer operating under the assumption of a static environment would be assumed to stoicly integrate the incoming sensory evidence, the likelihood, with the current beliefs about the state of the world, the prior (de Lange et al., 2018; Ma, 2019; Yon & Frith, 2021). Evidence of a sudden change in the state of the world, like a CP, would not register as evidence of sudden change, but rather as unusually corrupted evidence of the earlier state of the world. Without accounting for the possibility of a CP, the observer is not capable of sudden adaptation of the prior. I would expect the now obsolete prior to have a longer lasting negative influence on perceptual acuity under a changed environment. Expressed in accuracy this would suggest the relationship with SAC level to be closer to linear. The overall relationship observed in the data, especially the sudden and sharp increase between SAC levels 1 and 2, seems to be indicative of sudden adaptation to change.

Although this interpretation is speculative and has not been tested for within this thesis, it is coherent with the present data. A future thesis could investigate this theory closer by directly comparing different Bayesian and near-Bayesian models with focus on adaptive processes in the face of changing environments.

D. 2.2. Surprisal quickly decreases after occurrence of a change point

A closer look on Figure 11 reveals the sharpest decrease of surprisal to be between SAC levels 1 and 2. Generally, the decrease is strong at the beginning and slows down for SAC levels 2 and 3, then stalls for levels 3-5.

Analog to accuracy, this is also coherent with the idea of Bayesian perceptual adaptation as to be expected from an ideal Bayesian observer accounting for possible changes in the environment. A simpler observer operating under the assumption of a static environment and incapable of quick adaptive processes would be expected to be impacted longer by a detrimental prior, only vanishing after integration of several new pieces of sensory evidence in the new direction. I would therefore expect a more linear decrease of surprisal over SAC levels. I interpret the sharp decrease in surprisal from SAC level 1 to 2 as a sign of quick adaptation to the changed environment. Analog to accuracy, this interpretation is coherent with the present data, but it is also speculative and has not been explicitly tested for within this thesis.

D. 3. Alternative explanations for accuracy

A possible alternative explanation for the relationship between SAC level and accuracy might lie in so-called *congruency effects*. Congruency effects can influence accuracy, if certain properties of the stimulus are also present in some equivalent form in the behavioral answer expected from the participant (Schmidt, 2016). It leads to higher accuracies for answers that are congruent to the stimulus, compared to incongruent ones. In the present task the behavioral answer always matched the perceived motion (clockwise rotation for a clockwise motion and vice versa), making all answers congruent and eliminating a congruency effect in the classical sense.

Nevertheless, a certain conflict in direction between perception and answer stays present. Participants could be preparing for a behavioral answer in a certain direction during several consecutive motions in the same direction. According to *ideomotor theory* (Thomaschke et al., 2018) action can prime perception. Motorvisual priming experiments have shown actions, but also the mere planning of actions, to affect perception, if stimuli and action share essential features (Thomaschke, 2012). It is possible that the planned motor response of a directional answer primed participants in the incongruent direction, influencing their perception of the presented direction right after a CP.

Another effect to be mentioned is the so-called *sequential effect*, which describes a bias in answers towards reinforcing a local pattern in stimulus history (Cho et al., 2002; Yu & Cohen, 2008). After several same-direction motions in a row participants would be expected to avoid violating a perceived pattern, hence answering with the same direction. This could explain the sudden decrease in accuracy after SAC occurrence of a CP. This

effect is stronger, the longer the pattern lasted, similar to what we would expect under the Bayesian hypothesis. The effect's characteristics integrate well into the idea of Bayesian perceptual adaptation and do not pose a real issue to the interpretation of our results. Yu and Cohen (2008), show that the sequential effect might possibly be rooted in Bayesian perception.

D. 4. P3b amplitude as a neural signature for Bayesian perceptual adaptation

The EEG data collected during the experiment corroborate the behavioral results and the idea of Bayesian perceptual adaptation, suggesting P3b amplitude variability to serve as a neural signature for Bayesian perceptual adaptation.

D. 4.1. Amplitude differences over SAC levels showed P3b variability

Cluster-based permutation analysis revealed differences over SAC levels, most prominently a central cluster around the Cz and FCz electrode from 372 – 436 ms after stimulus onset. Due to its temporal and spatial boundaries I termed it P3b cluster. This somewhat exploratory analysis was done to find and pinpoint the exact spatiotemporal location of differences between SAC levels, and it confirmed the expected connection to the P3b (Visalli et al., 2021).

Sadly, there is no consensus on the neural functionality of the P3b yet, despite it being one of the most heavily researched ERP components. A well-replicable key factor of the P3b is its inverse correlation with stimulus probability. This refers to both overall probability and (very importantly in the context of SAC levels) local probability, the stimulus occurrence compared to number of preceding stimuli, since last target stimulus (Luck, 2005).

This correlation is at the heart of our analysis and coherent with my interpretation of the acquired data under our Bayesian model, specifically the use of priors. With priors encapsulating belief about the environment and predictions about incoming stimuli, SAC levels tell us which stimulus should be expected to be perceived as particularly (locally) improbable: the one directly after a CP, so SAC level 1.

This is touching on the very definition of surprisal (Baldi & Itti, 2010; Sedley et al., 2016), which has already been shown to be highly dependent on SAC level (see Figure 11) and is known to be reflected in the P3b (Visalli et al., 2021). We see this connection between probability and P3b cluster amplitude, combined with previous knowledge and

assumptions about the connections of priors and surprisal with SAC level, clearly reflected in Figure 16.

Interestingly, P3b cluster amplitude is slightly increasing again from SAC level 3, after its initial decrease from level 1 to 3. While I do not have an explanation for this, I would like to point out that the differences between levels 2 – 5 are very small overall and the main difference appears to be between level 1 and the subsequent levels. This is also visible in the ERPs (see Figure 15) that show a later and longer sustained P3b peak in SAC level 1, compared to the other levels.

Analog to the reasoning for accuracy and surprisal, I interpret the sharp drop, and the later and sustained peak from level 1 to 2 as an indicator for Bayesian perceptual adaptation. The overall similarity of P3b characteristics (latency, sustain, cluster amplitude) between SAC levels 2 to 5 point toward a quicker adaptation to change after SAC level 1 than to be expected under a simpler Bayesian observer, although this interpretation is again somewhat speculative and not explicitly tested for.

To conclude, the measured P3b ERP hitherto fits perfectly within the previous results and within the idea of Bayesian perceptual adaptation and adds a neurophysiological component to the so far purely behavioral analysis.

D. 4.2. Surprisal predicted P3b amplitude variability

As predicted, momentary estimates of surprisal correlate with amplitude differences in the P3b ERP. The hypothesis can be confirmed based on the presented data and analysis (see Figure 17). P3b cluster amplitudes, retrieved from a P3b cluster calculated over amplitude differences in SAC levels 1 to 5, show larger amplitudes to be significantly correlated with larger surprisal values, although the predictive power is very low. This is coherent with existing literature describing P3b reflecting surprisal (Visalli et al., 2021) and the P3b's sensitivity to stimulus probability (Luck, 2005). Surprisal as a measure of the (reliability weighted) prediction error and the information content of a sensory input is at its core defined by said inputs (im-)probability.

I interpret this as an indicator that participants utilize priors and adapt to change as predicted by the BCP model, and henceforth also perceive the probability of incoming stimuli as predicted by the model. This perceived probability, which is quantified as surprise, then translates into the described P3b amplitude variability due to its sensitivity to stimulus probability and surprisal, partly explaining its variance.

As a neurophysiological measure, P3b cluster amplitude is naturally subject to substantial variance and noise, and there is to this date no broadly supported theory of the P3b's function. The present data suggests that, because of its well-established connection to probability and surprisal, the P3b can be understood as a neural signature for Bayesian perceptual adaptation during auditory motion discrimination.

D. 5. Alternative explanations for P3b variability

It has been proposed that P3b amplitudes reflect resource allocation, increasing for difficult tasks in comparison to less difficult tasks (Isreal et al., 1980; Luck, 2005; Polich, 1987). While our participants only performed one task, I cannot rule out effects of resource allocation completely, due to stimulus saliency and task difficulty.

Sudden change is generally perceived as salient (Huang & Elhilali, 2017; Liesefeld et al., 2017). It is possible that CPs, purely by their nature of being change, are in themselves perceived as salient, which could be responsible for increased resource allocation. In the visual domain the P300 has shown to be modulated by stimulus saliency (Teixeira et al., 2010). We have designed the task in a way that every stimulus had the same chance of being the last and therefore most task-relevant one, which we expected to lead to high and constant attention in alert subjects. In theory, this should have minimized attentional fluctuations and therefore effects of resource allocation. In practice, psychoacoustic experiments are repetitive and tiring and are likely to take a toll on participants alertness despite frequent breaks, so I cannot rule out possible effects of resource allocation. Nonetheless, saliency should not pose a significant issue, as CPs are generally hard to perceive (accuracy close to chance level at SAC 1, see Figure 10), which would be a prerequisite for effects caused by saliency.

Regarding task difficulty, it is very complicated to assess how it would affect the present data. Difficult tasks are known to increase resource allocation, while uncertainty about a stimulus decreases it (Luck, 2005). Thinking about the present task as a series of small tasks, the most difficult one would be the directional judgement at SAC level 1, with an accuracy close to chance level. The same task also represents the maximum uncertainty about the motions' direction. The present data would indicate the effects of difficulty to outweigh the effects of uncertainty, which could result in the presented P3b amplitude variability, but this reasoning leads down an exceptionally speculative path, as it is a post-hoc explanation that could have justified any other possible result just as well.

It is furthermore important to stress that effects of task difficulty usually occur in dual-task experiments, where resources need to be split between two simultaneous tasks (Chipunza & Mandeya, 2005; Isreal et al., 1980; Wickens et al., 1979), which is not the case in the presented task. Additionally, the difficulty is not fluctuating over tasks or blocks, but over presented stimuli, which gives participants no possibility of adapting to any overall "task difficulty". The P3b amplitude would rather need to be a reflection of the perceived stimulus difficulty, *after* stimulus presentation, as there is no way of anticipating the difficulty of upcoming stimuli. Existing research has shown the P300 to be linked to anticipated task difficulty (Ullsperger et al., 1987).

Lastly, if task difficulty played a role, it would still integrate well into the idea of Bayesian inference. Looking at each stimulus or each directional perception independently, their difficulty should rely only on their respective velocity. Bayesian inference offers an explanation to why SAC levels have different difficulties.

D. 6. Conclusion

The results presented in this study suggest auditory motion discrimination to show hallmarks of Bayesian inference on a behavioral and neurophysiological level. We employed a change point paradigm to examine auditory motion perception under the premise of a volatile world. The present data suggests participants accounted for possible changes in the environment and were able to adapt quickly to it. EEG data collected during the experiment suggests P3b amplitude variability to serve as a neural signature for Bayesian perceptual adaptation, during auditory motion discrimination.

E. ACKNOWLEDGMENTS

This study was conducted (and this thesis was written) within the Dynamates project of the Acoustics Research Institute, Austrian Academy of Sciences (Young Independent Researcher Group Dynamates, ZK66). Within this project I received funding from the Austrian Science Fund (FWF). I would like to thank my co-supervisor Dr. Robert Baumgartner for his help and support! I would further like to thank Burcu Bayram, Dr. David Meijer, Dr. Roberto Barumerli, Tobias Greif and Dr. Karolina Ignatiadis who have all been involved either in the present study or in neighboring projects and have all been a huge help.

F. REFERENCES

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Braren, H. S., & Fels, J. (2020). *A High-Resolution Head-Related Transfer Function Data Set and 3D-Scan of KEMAR* [dataset]. RWTH Aachen University. https://doi.org/10.18154/RWTH-2020-11307

Chapman, R. M., & Bragdon, H. R. (1964). Evoked Responses to Numerical and Non-Numerical Visual Stimuli while Problem Solving. *Nature*, *203*(4950), Article 4950. https://doi.org/10.1038/2031155a0

Chipunza, C., & Mandeya, A. (2005). Dual-Task Processing: Effects of Task Difficulty and Stimulus Similarity on Dual-Task Performance. *South African Journal of Psychology*, *35*(4), 684–702. https://doi.org/10.1177/008124630503500405

Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., & Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(4), 283–299. https://doi.org/10.3758/CABN.2.4.283

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

Cole, S. R., Edwards, J. K., & Greenland, S. (2021). Surprise! *American Journal of Epidemiology*, *190*(2), 191–193. https://doi.org/10.1093/aje/kwaa136

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019). Hallucinations and Strong Priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris*, *98*(1), 249–258. https://doi.org/10.1016/j.jphysparis.2004.03.011

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

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

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

Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O'Connor, J. P. B., & Price, G. J. (2022). Understanding the Differences Between Bayesian and Frequentist Statistics. *International Journal of Radiation Oncology*Biology*Physics*, *112*(5), 1076–1082. https://doi.org/10.1016/j.ijrobp.2021.12.011

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176. https://doi.org/10.1121/1.4979055

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and Tracking Difficulty: Evidence For Multiple Resources in Dual-Task Performance. *Psychophysiology*, *17*(3), 259–273. https://doi.org/10.1111/j.1469-8986.1980.tb00146.x

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Kammers, M. P. M., de Vignemont, F., Verhagen, L., & Dijkerman, H. C. (2009). The rubber hand illusion in action. *Neuropsychologia*, *47*(1), 204–211. https://doi.org/10.1016/j.neuropsychologia.2008.07.028

Kersten, D., & Mamassian, P. (2009). Ideal Observer Theory. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 89–95). Academic Press. https://doi.org/10.1016/B978-008045046-9.01435-2

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1-16. https//doi.org/10.1068/v070821

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, *43*(9), 2743–2758. https://doi.org/10.1002/hbm.25832

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

Kolossa, A., Fingscheidt, T., Wessel, K., & Kopp, B. (2013). A Model-Based Approach to Trial-By-Trial P300 Amplitude Fluctuations. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00359

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of Bayesian inference. *NeuroImage*, *106*, 222–237. https://doi.org/10.1016/j.neuroimage.2014.11.007

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737. https://doi.org/10.1016/S0042-6989(98)00285-5

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 6. https://doi.org/10.1038/s41562-017-0107

Kroese, D. P., & Rubinstein, R. Y. (2012). Monte Carlo methods. *WIREs Computational Statistics*, *4*(1), 48–58. https://doi.org/10.1002/wics.194

Li, S., & Peissig, J. (2020). Measurement of Head-Related Transfer Functions: A Review. *Applied Sciences*, *10*(14), Article 14. https://doi.org/10.3390/app10145014

Liesefeld, H. R., Liesefeld, A. M., Müller, H. J., & Rangelov, D. (2017). Saliency maps for finding changes in visual scenes? *Attention, Perception, & Psychophysics*, *79*(7), 2190–2201. https://doi.org/10.3758/s13414-017-1383-9

Luck, S. (2005). *An Introduction to the Event-Related Potential Technique,* Bradford Books

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

Maij, F., Seegelke, C., Medendorp, W. P., & Heed, T. (2020). External location of touch is constructed post-hoc based on limb choice. *eLife*, *9*, e57804. https://doi.org/10.7554/eLife.57804

Majdak, P., Balázs, P., & Laback, B. (2007). Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. *Journal of The Audio Engineering Society*. http://www.aes.org/e-lib/browse.cfm?elib=14190

Manning, T. S., Naecker, B. N., McLean, I. R., Rokers, B., Pillow, J. W., & Cooper, E. A. (2023). A General Framework for Inferring Bayesian Ideal Observer Models from Psychophysical Data. *eNeuro*, *10*(1). https://doi.org/10.1523/ENEURO.0144-22.2022

McLachlan, G., Majdak, P., Reijniers, J., Mihocic, M., & Peremans, H. (2023). Dynamic spectral cues do not affect human sound localization during small head movements. *Frontiers in Neuroscience*, *17*. https://doi.org/10.3389/fnins.2023.1027827

McLachlan, G., Majdak, P., Reijniers, J., & Peremans, H. (2021). Towards modelling active sound localisation based on Bayesian inference in a static environment. *Acta Acustica*, *5*, 45. https://doi.org/10.1051/aacus/2021039

Modirshanechi, A., Becker, S., Brea, J., & Gerstner, W. (2023). Surprise and novelty in the brain. *Current Opinion in Neurobiology*, *82*, 102758. https://doi.org/10.1016/j.conb.2023.102758

Modirshanechi, A., Brea, J., & Gerstner, W. (2022). A taxonomy of surprise definitions. *Journal of Mathematical Psychology*, *110*, 102712. https://doi.org/10.1016/j.jmp.2022.102712

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), Article 7. https://doi.org/10.1038/nn.3130

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, *31*(7), 1221–1236. https://doi.org/10.1016/0042-6989(91)90047-9

Parr, T., Benrimoh, D. A., Vincent, P., & Friston, K. J. (2018). Precision and False Perceptual Inference. *Frontiers in Integrative Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnint.2018.00039

Pernet, C. R., Latinus, M., Nichols, T. E., & Rousselet, G. A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods*, *250*, 85–93. https://doi.org/10.1016/j.jneumeth.2014.08.003

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

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Plack, C. J. (2013). *The Sense of Hearing* (3rd ed.). Routledge. https://doi.org/10.4324/9781315208145

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 311–320. https://doi.org/10.1016/0168-5597(87)90052-9

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Prins, N. (2012). The Adaptive Psi Method and the Lapse Rate. *Journal of Vision*, *12*(9), 322. https://doi.org/10.1167/12.9.322

Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, *13*(7), 3. https://doi.org/10.1167/13.7.3

Rahnev, D. (2019). The Bayesian brain: What is it and do humans have it? *Behavioral and Brain Sciences*, *42*, e238. https://doi.org/10.1017/S0140525X19001377

Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, *350*(6320), Article 6320. https://doi.org/10.1038/350699a0

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence. *Topics in Cognitive Science*, *11*(1), 50–74. https://doi.org/10.1111/tops.12292

Schmidt, J. R. (2016). Context-Specific Proportion Congruency Effects: An Episodic Learning Account and Computational Model. *Frontiers in Psychology*, *7*, 1806. https://doi.org/10.3389/fpsyg.2016.01806

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

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stevenson-Hoare, J. O., Freeman, T. C. A., & Culling, J. F. (2022). The pinna enhances angular discrimination in the frontal hemifield. *The Journal of the Acoustical Society of America*, *152*(4), 2140–2149. https://doi.org/10.1121/10.0014599

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Teixeira, M., Cas[TEL]o-Branco, M., Nascimento, S., & Almeida, V. (2010). The P300 signal is monotonically modulated by target saliency level irrespective of the visual feature domain. *Acta Ophthalmologica*, *88*(s246), 0–0. https://doi.org/10.1111/j.1755-3768.2010.4325.x

Thomaschke, R. (2012). Investigating Ideomotor Cognition with Motorvisual Priming Paradigms: Key Findings, Methodological Challenges, and Future Directions. *Frontiers in Psychology*, *3*, 519. https://doi.org/10.3389/fpsyg.2012.00519

Thomaschke, R., Miall, R. C., Rueß, M., Mehta, P. R., & Hopkins, B. (2018). Visuomotor and motorvisual priming with different types of set-level congruency: Evidence in support of ideomotor theory, and the planning and control model (PCM). *Psychological Research*, *82*(6), 1073–1090. https://doi.org/10.1007/s00426-017-0885-3

Ullsperger, P., Neumann, U., Gille, H.-G., & Pietschmann, M. (1987). P300 and anticipated task difficulty. *International Journal of Psychophysiology*, *5*(2), 145–149. https://doi.org/10.1016/0167-8760(87)90018-3

Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2021). Electroencephalographic correlates of temporal Bayesian belief updating and surprise. *NeuroImage*, *231*, 117867. https://doi.org/10.1016/j.neuroimage.2021.117867

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743. https://doi.org/10.3758/BF03196329

Wickens, C. D., Tsang, P. S., & Benel, R. A. (1979). The Allocation of Attentional Resources in a Dynamic Environment. *Proceedings of the Human Factors Society Annual Meeting*, *23*(1), 527–531. https://doi.org/10.1177/107118137902300113 1

Yang, S., Bill, J., Drugowitsch, J., & Gershman, S. J. (2021). Human visual motion perception shows hallmarks of Bayesian structural inference. *Scientific Reports*, *11*(1), 3714. https://doi.org/10.1038/s41598-021-82175-7

Yon, D. (2021). Prediction and Learning: Understanding Uncertainty. *Current Biology*, *31*(1), R23–R25. https://doi.org/10.1016/j.cub.2020.10.052

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., & Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, *21*, 1873–1880.

G. LIST OF FIGURES

H. LIST OF ABBREVIATIONS

Abbrevitation	Definition
BCP model	Bayesian change point model
CP	Change point
EEG	Electroencephalography
ERP	Event-related potential
HRTF	Head-related transfer function
LDD	Last direction discrimination
MAA	Minimal audible angle
SAC	Sound after change point
SEM	Standard error of mean
SOA	Stimulus onset asynchrony

I. APPENDIX

I. 1. Abstract

Auditory perception is subject to sensory noise and rapidly changing environments. To deal with ambiguous input, the auditory system needs to find the correct balance between flexibility and robustness. Bayesian inference determines the statistically optimal solution. The present thesis investigated on a behavioral and neurophysiological level whether auditory motion perception employs Bayesian inference. While monitored via high-density EEG, 26 young adults indicated the final direction of auditory motion sequences with random length and change points (CPs). Participants' accuracy changed significantly based on the occurrence of CPs with a sharp decrease directly following a CP and a steady increase with additional motions in the same direction, revealing a strong influence of momentarily established priors on auditory perception. Cluster-based permutation analysis of the EEG data revealed a centrally distributed P3b component showing later and longer sustained activity for motions directly following a CP. Momentary estimates of surprisal were estimated by a Bayesian CP model and significantly predicted the cluster amplitudes on a single-sound level. These findings suggest auditory motion perception to continuously adapt to unpredictable changes as the Bayesian observer would do.

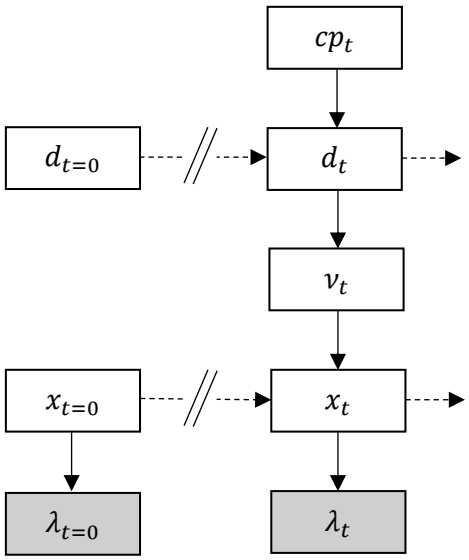
I. 2. Deutsches Abstract

Hörwahrnehmung unterliegt sensorischem Rauschen und einer sich plötzlich verändernden Umwelt. Um mehrdeutige sensorische Reize zu verarbeiten, benötigt das auditive System eine Balance aus Flexibilität und Robustheit. Bayes'sche Inferenz bestimmt die statistisch optimale Lösung. Die vorliegende Arbeit untersucht, ob auditive Bewegungswahrnehmung Bayes'schen Prinzipien folgt, auf Verhaltens-, sowie auf neurophysiologischer Ebene. Unter Aufzeichnung eines Elektroencephalograms (EEG) wurden 26 jungen Erwachsenen Sequenzen von auditiven Bewegungen mit zufälliger Länge und direktionalen Wendepunkten (WP) präsentiert, wobei die Teilnehmer*innen die Richtung der letzten wahrgenommenen Bewegung angaben. Die Genauigkeit der Teilnehmer*innen änderte sich signifikant abhängig von WP, mit sehr niedriger Genauigkeit direkt nach einem WP und einem stetigen Anstieg über multiple Bewegungen ohne Richtungsänderung. Dies zeigt

einen starken Einfluss temporär etablierter *Prior* auf auditive Bewegungswahrnehmung. Clusterbasierte Permutationsanalyse der EEG-Daten zeigt im zentralen Ereigniskorrelierten Potential P3b spätere und länger anhaltende Aktivität für auditive Bewegungen direkt nach einem WP. Aus einem Bayes'schen model wurden Schätzwerte des reliabilitäts-gewichteten Vorhersagefehlers (*surprisal*) extrahiert, welche P3b Clusteramplituden für einzelne auditive Stimuli signifikant vorhersagen. Die Ergebnisse legen nahe, dass sich die Wahrnehmung akustischer Bewegungen, wie ein idealer Bayes'scher Beobachter, kontinuierlich an unvorhersehbare Veränderungen der Umwelt anpasst.

I. 3. Bayesian CP model by Dr. David Meijer

Generation of sensory signals



$$cp_{t=1} = 1$$

$$cp_{t>1} \sim B(1, H)$$

$$d_{t=0} \sim 2B(1, 0.5) - 1$$

$$d_t = \begin{cases} d_{t-1} & if\ cp_t = 0 \\ -d_{t-1} & if\ cp_t = 1 \end{cases}$$

$$v_t \sim N\left(v_t; d_t \mu_{exp}, {\sigma^2}_{exp}\right)$$

$$x_{t=0} \sim U(-40°, 40°)$$

$$x_t = x_{t-1} + v_t$$

$$\lambda_t \sim N\left(\lambda_t; x_t, \left(\sigma_{sens,0°} + \gamma_\sigma |x_t|\right)^2\right)$$

We sampled the first sound location $x_{t=0}$ from a bounded uniform distribution between -40° and +40°. Subsequent sound locations x_t were determined by velocity v_t according to: $x_t = x_{t-1} + v_t$. At each timepoint $t \geq 1$, velocity v_t was sampled at random from a normal distribution with variance ${\sigma^2}_{exp}$ and mean equal to $d_t * \mu_{exp}$. Here, $d_t \in \{-1, +1\}$ indicates the direction of the current movement direction, whereas σ_{exp} and μ_{exp} determine the task difficulty. [We set them individually as $\mu_{exp} = 3 * MAA_{0°}$ and $\sigma_{exp} = 1 * MAA_{0°}$, where $MAA_{0°}$ is the minimum audible angle at 0° azimuth that was measured for every subject before starting the main task.] Movement direction d_t was sampled at random at the beginning of each trial, but it would change suddenly during the sound sequence whenever a changepoint occurred: i.e. when $cp_t = 1$. Changepoints were sampled at random with a fixed hazard rate: $H = 1/5$. For analyses purposes, we defined every trial's first velocity, $v_{t=1}$, as having followed from a changepoint; i.e. $cp_{t=1} = 1$.

Observers only have access to noise-corrupted internal estimates λ_t of the true sound locations x_t. We assume normally distributed sensory noise with standard deviation $\sigma_{sens}(x_t)$, which may linearly increase with the absolute stimulus location: $\sigma_{sens}(x_t) = \sigma_{sens,0°} + \gamma_\sigma |x_t|$, with $\gamma_\sigma \geq 0$.

Prior knowledge and likelihood functions

Participants were superficially informed by the experimenter about the concepts of changepoints and experimental noise, without mentioning specific values for hazard rate H or experimental parameters μ_{exp} and σ_{exp}. Nevertheless, we assume that our observers learned about the velocity distribution, $N(v_t; \pm\mu_{exp}, {\sigma^2}_{exp})$, during the main task practice block. We also assume that they were able to form a stable, but not necessarily accurate estimate of the constant hazard rate H.

Moreover, we assume that observers have access to the instantaneous uncertainty of their own internal estimates, $\sigma_{sens}(\lambda_t) = \sigma_{sens}(x_t) = \sigma_{sens,0^\circ} + \gamma_\sigma|x_t|$. Hence, their likelihood function for sound location x_t is normally distributed and centred on internal estimate λ_t:

$$p(\lambda_t|x_t) = N(x_t;\ \lambda_t, {\sigma^2}_{sens}(\lambda_t)) \qquad \text{(Eq. 1)}$$

Combining equation 1 and $v_t = x_t - x_{t-1}$, it follows that the likelihood function for velocity v_t must also be normally distributed, with mean equal to the difference of the last two consecutive internal estimates, and variance equal to the sum of the variance for each stimulus location:

$$p(\lambda_t - \lambda_{t-1}|v_t) = N(v_t;\ \lambda_t - \lambda_{t-1}, {\sigma^2}_{sens}(\lambda_t) + {\sigma^2}_{sens}(\lambda_{t-1})) \qquad \text{(Eq. 2)}$$

Furthermore, since velocity v_t was sampled from a normal distribution according to the generative process, $v_t \sim N(v_t; d_t\mu_{exp}, {\sigma^2}_{exp})$, we can compute the likelihood of observing $\lambda_t - \lambda_{t-1}$ conditional on either direction as:

$$p(\lambda_t - \lambda_{t-1}|d_t) = N\left(\lambda_t - \lambda_{t-1}; d_t\mu_{exp}, {\sigma^2}_{exp} + {\sigma^2}_{sens}(\lambda_t) + {\sigma^2}_{sens}(\lambda_{t-1})\right) \qquad \text{(Eq. 3)}$$

In other words, the likelihood for a left-/rightward direction d_t can be computed by evaluating the probability density of $N\left(\pm\mu_{exp}, {\sigma^2}_{exp} + {\sigma^2}_{sens}(\lambda_t) + {\sigma^2}_{sens}(\lambda_{t-1})\right)$ at the point $\lambda_t - \lambda_{t-1}$.

Bayesian directions Change Point (BdCP) model

An observer's task is to infer the direction of the sound movement at the end of the sequence. The simplest estimate of d_t is given by the likelihood ratio:

$$LikeRatio_{d_t} = \frac{p(\lambda_t - \lambda_{t-1}|d_t = 1)}{p(\lambda_t - \lambda_{t-1}|d_t = -1)} = \frac{LikeL_t}{LikeR_t} \qquad \text{(Eq. 4)}$$

If this ratio is larger than one, then a leftward movement is more likely. Vice versa, a $LR_{d_t} < 1$ indicates that a rightward movement is more likely. This is called the maximum likelihood method. However, such a strategy that bases its direction responses merely on the relative likelihood for either direction would give many erroneous responses if there were much sensory noise. I.e. when ${\sigma^2}_{sens}(\lambda_t) + {\sigma^2}_{sens}(\lambda_{t-1})$ is large it is likely that the instantaneous internal estimate of the current velocity, $\lambda_t - \lambda_{t-1}$, is wildly inaccurate and a change of sign may frequently occur.

Instead, Bayesian theory postulates that ideal observers can minimize their perceptual errors by making use of all available sensory information ($\lambda_{1:t}$) and their knowledge of the generative process for these sensory signals, including learned parameters μ_{exp}, σ_{exp}, and H. In particular, one can make use of the fact that changepoints only occur sometimes. This means that a belief about the previous movement direction is also useful information about the current movement direction. According to the Bayesian belief updating framework, such pre-existing beliefs are used to construct

a “prior”. After hearing a new sound, the prior information is then integrated with the information from the latest likelihood function to obtain a “posterior”.

At the start of a sequence, each direction is equally likely (50%). The prior ratio thus equals 1:

$$PriorRatio_{d_{t=1}} = \frac{p(d_{t=1}=1)}{p(d_{t=1}=-1)} = \frac{PriorL_{t=1}}{PriorR_{t=1}} = 1 \qquad \text{(Eq. 5)}$$

After hearing a new sound, the likelihood ratio is computed according to equation 4. Subsequently, prior and likelihood are integrated by multiplication to form a posterior ratio:

$$PostRatio_{d_t} = \frac{p(d_t=1|\lambda_{0:t})}{p(d_t=-1|\lambda_{0:t})} = \frac{PostL_t}{PostR_t} = \frac{c^{-1}*PriorL_t*LikeL_t}{c^{-1}*PriorR_t*LikeR_t} = PriorRatio_{d_t} * LikeRatio_{d_t} \qquad \text{(Eq. 6)}$$

Notice that the normalization constant from Bayes’ rule, $c = PriorL_t * LikeL_t + PriorR_t * LikeR_t$, drops out and does not need to be computed.

To form the next prior ratio, the posterior ratio is updated to account for the possibility of a changepoint before the presentation of the next sound:

$$PriorRatio_{d_{t>1}} = \frac{p(d_t=1|\lambda_{0:(t-1)})}{p(d_t=-1|\lambda_{0:(t-1)})} = \frac{PriorL_t}{PriorR_t} = \frac{(1-H)PostL_{t-1}+HPostR_{t-1}}{(1-H)PostR_{t-1}+HPostL_{t-1}} \qquad \text{(Eq. 7)}$$

So, a Bayesian ideal observer model iteratively updates its beliefs by following equations 4, 6 and 7. After the last sound has played, the final decision is made based on the posterior ratio (equation 6).

Latent variables of the BdCP model

From the above-described BdCP model we extract a few useful variables that give insight into the belief updating mechanism at each time-point during the sound sequences.

We compute the posterior probability of a changepoint as:

$$p(cp_t|\lambda_{0:t}) = \frac{H*(PostR_{t-1}*LikeL_t+PostL_{t-1}*LikeR_t)}{H*(PostR_{t-1}*LikeL_t+PostL_{t-1}*LikeR_t) + (1-H)*(PostL_{t-1}*LikeL_t+PostR_{t-1}*LikeR_t)} \qquad \text{(Eq. 8)}$$

As a measure of belief uncertainty, we compute the entropy for prior and posterior as:

$$PriorEntropy = -PriorL_t * log_2(PriorL_t) - PriorR_t * log_2(PriorR_t) \qquad \text{(Eq. 9)}$$

$$PostEntropy = -PostL_t * log_2(PostL_t) - PostR_t * log_2(PostR_t) \qquad \text{(Eq. 10)}$$

We also compute the information theoretic measure of surprisal as:

$$Surprisal = -log_2(PriorL_t * LikeL_t + PriorR_t * LikeR_t) \qquad \text{(Eq. 11)}$$

Finally, we define ‘information gain’ as the Jensen Shannon distance from prior to posterior:

$$InfoGain = \sqrt{0.5 * [KL(P,M) + KL(Q,M)]} \qquad \text{(Eq. 12)}$$

Here, $P = \{PriorL_t, PriorR_t\}$ and $Q = \{PostL_t, PostR_t\}$,
$M = \{0.5 * [PriorL_t + PostL_t], 0.5 * [PriorR_t + PostR_t]\}$

and $KL(X,Y)$ represents the Kullback Leibler divergence: $KL(X,Y) = \sum \left[X_i * log_2\left(\frac{X_i}{Y_i}\right)\right]$.