MASTERARBEIT | MASTER'S THESIS

Titel | Title

Neural Signatures of Bayesian Perception: Dynamic Predictions in Auditory Tempo Changes

verfasst von | submitted by

Valentin Pellegrini BSc

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

Master of Science (MSc)

Wien | 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
Betreut von | Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge

Table of Contents

1. Introduction

1.1. Perception under Uncertainty

How do we make sense of the world around us? The great mystery of perception has had thinkers occupied for thousands of years. As early as 350 B.C.E, Greek philosopher Aristotle explored how the different senses are used to enable the "soul" to capture the qualities of real-world objects. He immortalized his thoughts in his work "de Anima", which influenced many generations of philosophers and scientists to come (Aristotle, 1986). Fast forward more than two millennia, perception and its underlying mechanisms still raise many questions yet to be answered.

From our subjective point of view, perception may seem like a mere one-to-one representation of our environment, a passive process in which information is captured by our sensory organs and broadcasted into consciousness by our brain. In fact, perception is far from a passive display of our environment but is based on a wide range of different cognitive functions and influenced by a palette of affective (Phelps 2006), motivational (Balcetis & Dunning, 2006), and cognitive factors (Gilbert & Li, 2013). One of the most important factors influencing perception is attention (Carrasco, 2011). By focusing our attention on a particular object or group of objects, our perception of said objects is finer than it is in the periphery (Posner et al., 1980). Imagine reading a magazine while waiting in the doctor's waiting room. While you will have a detailed precept of the letters and pictures in the magazine, it is likely that your percept of a painting hanging on the wall, being in your attentional periphery will be much less detailed (Baldauf & Deubel, 2008), in fact you might not even take notice of it (Mack & Rock, 1998). In 1999, Daniel J. Simons and Christopher F. Chabris demonstrated in a classic study that the effect of attention on what we perceive can even make us "blind" for things that are in plain sight (Simons & Chabris, 1999). This phenomenon is known as inattentional blindness.

Most everyday situations are polluted with background noise (Faisal et al., 2008), for example, the sound of a busy street you are walking along, while trying to have a conversation with a friend. In scenarios like this, you would barely be able to follow the words of your friend if your brain did not apply some sort of filter to the acoustic signals you receive, separating the

signal into the important part (your friend's voice) and background noise that can be ignored (Shinn-Cunningham, 2008). This example demonstrates the issues that would arise if perception was merely a direct representation of our environment.

1.1.1. Perception and Sensory Noise

Even if we could eliminate all external sources that produce background noise, we could not be fully certain that our perception of the world is completely congruent with the actual state of the world. This is due to our sensory organs' inherent fluctuations in accuracy when processing sensory input (Faisal et al., 2008), also referred to as sensory noise. The same sensory input can be perceived differently when presented multiple times under the same exact conditions (Eagleman, 2001; Shen & Ma, 2019). Attentional mechanisms cannot eliminate all uncertainty about the state of the world caused by sensory noise (Cavanagh & Alvarez, 2005). Different sets of sensory input can arise from a single state of the world, while a single set of sensory input could represent different states of the world (Knill & Pouget, 2004). This uncertainty results in perceptual decisions we have to take, regarding the interpretation of sensory input (Gold & Shadlen, 2007). An extensive body of research suggests that this inferential decision process is not only influenced by sensory evidence, that is, the sensory signal the brain receives, but also by prior beliefs about the state of the world (e.g., Bar, 2007; Clark, 2013; Friston, 2005; Griffiths et al., 2010; Hohwy, 2012; Summerfield & De Lange, 2014).

1.2. Bayesian Perception

The Bayesian perception framework posits that the brain generates an internal model of our environment by integrating sensory evidence and prior beliefs, which is constantly updated in response to changes in the environment (Friston, 2005; Knill & Pouget, 2004; Clark, 2013). A growing body of research suggests that the brain follows the rules of Bayesian inference in making perceptual decisions (Summerfield & de Lange, 2014; Tenenbaum et al., 2011). Before exploring the depths of Bayesian perception, it is essential to review the basic concepts underlying Bayesian inference.

1.2.1. Principles of Bayesian Inference

Bayesian inference is named after Reverend Thomas Bayes, who developed a method to estimate and update probabilities based on the integration of current evidence and prior knowledge. In classical frequentist statistics, the observed data is viewed as one of many possible data sets. Contrary to that, Bayesian inference assumes the data to be fixed, but views hypotheses and parameters as probability distributions (Fornacon-Wood et al., 2022). These probability distributions can be updated, once new evidence arises by using Bayes Theorem:

$$p(H \mid E) = \frac{p(E \mid H)p(H)}{p(E)}, \text{ where}$$

$p(H)$ represents the prior probability distribution of hypothesis H without taking any current evidence into account, also dubbed the "prior". $p(E \mid H)$ is the likelihood function of hypothesis H, dubbed the "likelihood". Loosely speaking, it is the conditional probability of observing evidence E, under the assumption of hypothesis H being true. $p(H \mid E)$ is the probability distribution of a Hypothesis H, given the Evidence E, also dubbed the "posterior". $p(E)$ is the probability of observing evidence E and acts as a normalization factor of the numerator (Ma, 2019). Bayesian inference is used to model a wide range of processes in psychology and cognitive neuroscience, as for example eye movements in naturalistic environments (Itti & Baldi, 2009), speech understanding (Goodman & Frank, 2016) or inductive reasoning (Tenenbaum et al., 2006) as well as in other fields like economics (e.g., Cogley & Sargent, 2008) or ecology (Ellison, 2004). As mentioned above, a growing body of research suggests that the human brain also follows the rules of Bayesian inference in interpreting sensory evidence (Friston, 2005; Knill & Pouget, 2004).

1.2.2. Bayesian Inference in Dynamic Environments

Usually, perception is not a single event, happening at a particular point in time, but a continuous process of interpreting and reinterpreting the stream of signals reaching our sensory

organs. As already established, Bayesian perception posits that the brain generates and constantly updates an internal model of the world around us (Friston, 2005). The need for constant updating becomes clear when we consider that in most situations our surroundings are dynamic instead of static. In everyday situations, it is rarely the case that an incoming stream of sensory signals remains stable for a prolonged period of time. Unless you are staring at a wall and ignore the slight head movements produced by your neck muscles, the visual input reaching your retina will most likely be characterized by changes. These changes can be subtle, like the leaves of a tree you are watching swaying in the wind, or they can be obvious, like watching the cars rush by while walking next to a busy street.

Most research on how the brain uses Bayesian inference to update its internal model focuses on visual perception (e.g. Knill & Pouget, 2004, Rao & Ballard, 1999; Summerfield & de Lange, 2014) and multisensory integration (Alais & Burr, 2004; Rohe et al., 2019; Shams & Beierholm, 2010). A smaller, but not insignificant body of literature exists on Bayesian inference in the auditory domain, mostly in the context of sound localization (Barumerli et al., 2020; McLachlan et al., 2021) and interval perception (Burr et al., 2009; Sawai et al., 2012). In the current study, we investigate how the human brain integrates sensory evidence and prior beliefs in the context of auditory tempo changes. Here, tempo changes refer to the changes in inter-stimulus-intervals (ISIs) in a sequence of consecutive sounds. Thus, sound sequences can either be accelerating (i.e., ISIs get shorter) or decelerating (i.e., ISIs get larger). It must be mentioned that the ISIs could certainly also remain the same (i.e., neither accelerating nor decelerating), but we will disregard this option for now. In our experiment, participants had to report whether a sequence of sounds that altered between acceleration and deceleration was either speeding up (acceleration) or slowing down (deceleration) right before the sound sequence abruptly stopped after a random number of sounds. By introducing changepoints (CPs), where the tempo change switched from acceleration to deceleration, or vice versa, at a random point, we were able to model a dynamic environment and force subjects to constantly update their internal model (see Section Task and Procedure). In other words, subjects had to infer the current tempo change of the sequence (acceleration or deceleration) based on sensory evidence (the difference between the last two ISIs) and their prior assumptions. Let us revisit Bayes Theorem and adapt it to the given context:

$$p(T \mid E) = \frac{p(E \mid T)p(T)}{p(E)}$$, where

the prior $p(T)$ represents the current assumption about the state of the sound sequence T without taking the most recent evidence into account. As established before, the prior is a probability distribution over all possible world states, but since in this case there are only two possible states (i.e., acceleration or deceleration), the prior can be summarized as a single value.

The likelihood $p(E \mid T)$ can be interpreted as how well the current evidence E fits the prior assumption T .The evidence E , and therefore the likelihood, is a continuous variable. For example, if the last ISI was much shorter than the second to last, E strongly suggests a preference for acceleration over deceleration. However, if the last ISI is only marginally shorter than the second to last one, E still suggests a preference for acceleration over deceleration, but with less certainty. The posterior $p(T \mid E)$ refers to the current belief about the state of the sound sequence T , after integrating the new evidence and the prior. Similarly to the prior, in this context the posterior is a probability distribution over two possible world states and can therefore be summarized as a single value. In perceptual tasks, the posterior represents the actual contents of perception, that is, the percept (Ma, 2019).

1.3. Bayesian Updating and Surprise

As discussed earlier, our surroundings are typically dynamic, thus new evidence constantly becomes available. In an iterative process, the posterior from a previous step becomes the prior for the next step of inference, which is then again integrated with the latest evidence (Ma, 2019). This process is called Bayesian updating (Knill & Pouget, 2004; Friston, 2005). What happens when the sensory evidence contradicts the prior assumptions about the world's state? Let us assume that our prior assumption about the state of the sound sequence is "acceleration", but the latest evidence suggests that the sequence is decelerating. Again, the strength of the evidence suggesting "deceleration" can vary, depending on how big the change between the last two ISIs was. In the Bayesian perception framework, this discrepancy between

new evidence and prior assumptions is called surprise. Note that in this context, surprise does not refer to the commonly used meaning of the term, for instance in surprise-party, but to this specific concept in Bayesian inference. A key function of Bayesian updating is minimizing surprise (Visalli et al., 2021), and therefore preventing possible cognitive costs (Zénon et al., 2019). In the current context we expect surprise and Bayesian updating to be highly correlated (Visalli, 2021). However, it must be mentioned that factors like reward (McGuire et al., 2014) or the volatility of the environment (Behrens et al., 2007; Nassar et al., 2010) can influence the extent to which surprising events trigger updating of the internal model. The concept of surprise is closely related, if not congruent with the concept of prediction errors (Friston, 2005), commonly used in predictive coding, a framework closely related to Bayesian perception (Clark, 2013). The term prediction error refers to the discrepancy between a predicted state of the world and a perceived state of the world (Den Ouden et al., 2012).

1.3.1. Experimental Manipulation of Surprise

The aforementioned changepoint paradigm is used to experimentally manipulate surprise. When a CP occurs in the sound sequence, that is, the sounds switch from acceleration to deceleration or vice versa, there is a mismatch between the available evidence and the prior assumption. This mismatch (i.e., the surprise) is largest for the sensory input that directly follows the CP, because the prior is only based on information acquired before the CP. The more evidence is accumulated after a CP, the less surprising the incoming sensory signals become (until there is another CP), since the prior is already adapted to the new state of the sound sequence, at least to some extent (Skerritt-Davis & Elhilali, 2018). Thus, the closer a sound follows a CP (Sounds after Changepoint, SAC), the stronger the surprise it elicits. We used the SAC-levels as a proxy for surprise.

1.3.2. Neural Correlates of Surprise

Great efforts have been made in the past to unravel the underlying neural mechanisms of surprise and belief updating in different contexts and sensory domains (e.g., Kolossa et al., 2015; Visalli et al., 2021, 2023). Some studies use functional magnetic resonance imaging (fMRI) to detect brain regions related to surprise and belief updating (Visalli et al., 2019). However, the majority of studies rely on the event-related potential (ERP) technique (Gijsen et al., 2021; Ostwald et al., 2012; Visalli et al., 2021, 2023). The ERP technique aims at identifying specific components in the electroencephalographic (EEG) signal that relate to specific neural functions (Luck, 2014). Most previous studies used Bayesian modeling to dissociate belief updating and surprise and aimed at identifying distinct ERP components related to these functions (e.g. Kolossa et al., 2015; Visalli et al., 2021; Visalli et al., 2023). Specifically, the P2 and P3 components are frequently linked to belief updating and surprise. (Kolossa et al., 2014; Valakos, 2020; Visalli et al., 2021, 2023) The P2 is a positive deflection in the auditory evoked potential, peaking around 150-250 ms after stimulus onset (Luck, 2014). The P3 is a positive deflection peaking around 250-500 ms, or even as late as 850ms after stimulus onset. The P3 can be separated into two subcomponents: an earlier, more frontal P3a and a later, more posterior P3b (Luck, 2014; Polich, 2007). Studies differ regarding which function relates to which (sub)component of the ERPs. Kolossa and colleagues linked belief updating to an anteriorly distributed P3a and surprise to a later and more parietally distributed P3b (Kolossa et al., 2015). In Visalli et al. (2021), surprise was related to a modulation of the P3b, but updating was found to modulate a later and more sustained P3b-like waveform in the context of temporal predictions (Visalli et al., 2021). In a follow-up study, Visalli and colleagues linked surprise to an early fronto-central P3-like modulation, while updating was linked to a later and more posterior P3-like waveform, as well as a centro-parietal P2 modulation, most likely representing an integration process of prior assumptions and likelihood of the observed evidence (Visalli et al., 2023). These different results might be explained to some extent by the fact that the appearance (in terms of peak amplitude and latency) of ERPs typically depends on the specific experimental paradigm (Luck, 2014).

1.4. Research Aims and Hypotheses

Albeit acknowledging that surprise and belief updating may rely on distinct conceptual and neural mechanisms, the current study does not systematically segregate the two concepts. The following two considerations underly this decision: First, within the framework of Bayesian modeling, belief updating and surprise tend to be highly correlated. The more surprising a stimulus is, the stronger is the impact it has on the assumed state of the world. Second, the validity of segregating the two concepts from one another is heavily dependent on the validity of the underlying model. As for now, there is only little relevant literature available examining mechanisms of Bayesian inference in the context of auditory tempo changes. This study aims at providing insights into the behavioral and neural correlates underlying the integration of sensory evidence and prior assumptions, focusing on surprising events (i.e., sensory evidence that does not fit the prior). In the following, when surprise is referred to, it must not be understood as opposed to belief updating, but rather as a term including both concepts.

1.4.1. Hypotheses

To examine if humans follow the principles of Bayesian inference in predicting auditory tempo changes, along with the underlying neural representations, we collected data on a behavioral as well as on an electrophysiological level.

On the behavioral level, we expect participants' accuracy in reporting the current state of the sound sequence (i.e., acceleration or deceleration) to be negatively correlated with the surprise elicited by the last event in the sound sequence (H1a). Furthermore, we hypothesize that this effect is modulated by the evidence level of the last event in the sound sequence (H1b).

On the electrophysiological level, we expect events linked to higher surprise to result in stronger amplitudes of P2 and/or P3 like event-related potentials, compared to less surprising events (H2a). We furthermore expect the modulating effect of surprise on the ERP amplitude to be stronger in high-evidence events, compared to low-evidence events (H2b).

2. Methods

2.1. Subjects

The sample consists of 27 subjects aged 20-33 (M = 25.15, SD = 3.46, 15 female) who were recruited via the Vienna Cognitive Science Hub and received monetary compensation for participating (10€/h). All subjects were right-handed, reported normal hearing and normal or corrected to normal vision. They reported having no history of psychiatric or neurological illness, major concussions, skull fractures or coma. Furthermore, they reported not being under the influence of alcohol or other drugs and not having taken any mood-altering medication. Subjects were naïve to the study's objective and were fluent in either German or English. Before the experiment, subjects received instructions concerning the task both verbally and in written format and provided written informed consent. After the experiment, subjects were debriefed verbally. To participate in the experiment, subjects had to respond correctly in at least 60% (30/50) of the trials in the second of two familiarization blocks. This threshold was implemented to make sure that subjects did not respond randomly. Out of 34 subjects, five were rejected due to this criterion. One subject aborted the experiment due to personal reasons, and a further subject's data set could not be further analyzed due to technical issues regarding the computer hardware. The resulting sample size of 27 is comparable to similar studies (Krishnamurthy et al., 2017; Visalli et al., 2021, 2023).

2.2. Experimental Setup

Subjects were seated in front of a monitor (48 x 27 cm, 1280x1024 pixels with a refresh rate of 60 Hz) in a sound insulated, dark room with a viewing distance of 70 cm. To prevent head movements and to keep the viewing distance constant, their head was fixated via a chinrest. Both their hands were placed on a keyboard. With four fingers on the left hand, they operated four marked keys on the very left of the keyboard, with their right index- and middle finger they operated the up and down arow keys on the right side of the keyboard. The up and down arrow

keys were used to report whether the sequence was speeding up or slowing down. The four marked keys were used to indicate subjects' confidence in their response (*very certain – somewhat certain – somewhat uncertain – very uncertain*). Note that this variable is not included in the present analysis. Throughout the experiment, we monitored brain activity via a 128-electrode EEG (actiCAP with actiCH Brain Products GmbH, Gilching, Germany) at a sampling rate of 1 kHz. Additionally, pupil dilation was monitored via an eye tracker (Eye-Link 1000 Plus; SR Research, Osgoode; Ontario, Canada) with a sampling rate of 1 Khz, however, these data are not part of the present thesis. Sounds were presented binaurally via in-ear headphones (ER-2, Etymotic Research, Inc., Grove Village, Illinois) at 48kHz sampling rate. As auditory stimuli we used 25ms pink noise bursts. The experimental script was run on Windows 10 in MATLAB utilizing the PsychToolbox extensions (Kleiner et al., 2007).

2.3. Task and Procedure

The experiment consisted of two familiarization blocks and four experimental blocks. Each block consisted of 50 trials with a 20 s break after the 25th trial. In between blocks, subjects could take self-paced breaks. After the second familiarization block there was a longer break (~ 1h) during which the EEG was prepared. The whole procedure took three to four hours.

2.3.1. Two-Alternative Forced Choice Task

Subjects initialized each block by pressing the space bar. During the presentation of the auditory stimuli a central fixation dot was displayed on the monitor. Each trial consisted of a sound sequence that altered between acceleration and deceleration. The difference between two consecutive ISIs (i.e., the rate of acceleration/deceleration) was chosen from a bimodal distribution and differed from sound to sound. The bimodal distribution consisted of two normal distributions, one relating to acceleration, one relating to deceleration. The sequences consisted of pseudorandomized numbers of sounds with a minimum of three and a maximum of 45 sounds. After each sound (except the first two) there was a 10% chance of the trial stopping, which resulted in an average number of 12 sounds per trial. The initial state of the sound sequence was counterbalanced between acceleration and deceleration. After each sound, there was a 20%

chance of the sound sequence changing from acceleration to deceleration or vice versa (i.e. a changepoint, CP). At the end of each sequence, subjects had to report whether the last ISI was shorter (acceleration) or longer (deceleration) than the second to last ISI. By pseudo randomizing the number of sounds in each sequence as well as the occurrence of CPs, subjects were encouraged to constantly monitor the current state of the sound sequence (Figure 1). Subjects could adjust their response regarding the state of the sound sequence (speed up or slow down, indicated via up and down arrow keys) until they confirmed their choice by selecting a confidence level with their left hand.

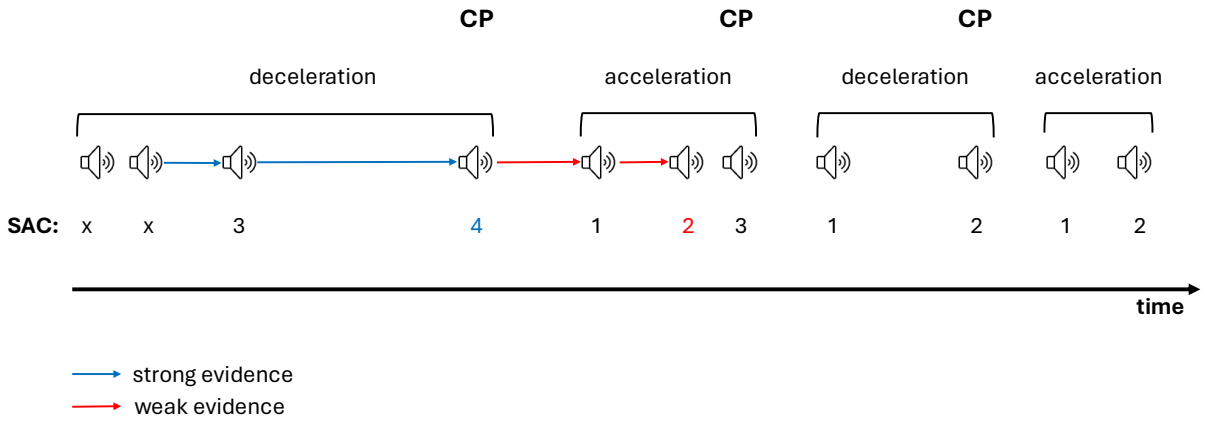

Figure 1: Schematic illustration of an example trial consisting of 11 sounds. The initial state of the sequence is deceleration, which changes to acceleration after the fourth sound, that is, a changepoint (CP) occurs. There is another CP at sound number seven and sound number nine. The subjects' task was to indicate the final state of the sound sequence, which would in this case be acceleration. The numbers under each sound represent the sounds SAC level, that is, their position relative to the latest changepoint (CP). Note that the first two sounds in a trial were not assigned an SAC, nor an evidence level value (see section *Experimental Blocks*) and were not included in the analysis. The blue arrows illustrate the strong evidence condition, where the two consecutive inter stimulus intervals (ISIs) differ strongly from each other (indicated by the length of the arrow). The red arrows illustrate the weak evidence condition, where the two consecutive

ISIs differ only slightly from one another. The rate of the tempo change was therefore lower in the weak evidence condition, compared to the strong evidence condition, which makes it harder to infer the current direction of the tempo change. Each sound was assigned an evidence level, the specific sounds in the figure were only chosen for illustrative reasons.

2.3.2. Familiarization Blocks

In total subjects completed six blocks of 50 trials, with the first two blocks being familiarization blocks. The familiarization blocks were used to individually set the task difficulty and allowed subjects to get used to the experimental paradigm. We aimed at a mean response accuracy of ~80% for the experimental blocks. Task difficulty varied in terms of tempo change rates, that is, how large the differences in ISIs between two consecutive sounds were. At the beginning of the first familiarization block, the difficulty parameter μ was set to 1.3. μ refers to the mean of the normal distributions the rates of tempo changes were chosen from. If the difficulty parameter μ = 1.3, then the second of two consecutive ISIs is on average 1.3 times as large as the first ISI for deceleration (on average 33% larger ISIs), and 1/1.3 times as large as the first ISI for acceleration (on average 23% shorter ISIs). The higher μ, the larger the differences between two consecutive ISIs and therefore the easier the task. At the start, ISI limits were set to 200 to 1333ms. Subjects' accuracy was monitored at each 10^{th} trial. If they responded correctly in more than eight out of the 10 trials, μ was decreased by 0.05, with a lower limit of μ = 1.05. If they responded correctly in less than eight trials, μ was increased by 0.05, with an upper limit of μ = 1.4, and the upper limit of the ISI was increased by 250ms, which enabled larger differences in consecutive ISIs. If they responded correctly in exactly eight trials, μ and ISI limits did not change. The second training block started with the individualized μ value obtained from block one or with μ = 1.3 if the subject's individualized μ from the first block was equal to or higher than 1.3. In the second familiarization block μ could only decrease, which means that task difficulty could only increase. In both familiarization blocks, subjects received feedback after every trial and summarized feedback after each block. If they responded correctly at the end of a trial, a green disc appeared on the screen following the confidence rating, if they responded incorrectly a red disc appeared. At the end of each block, they were informed via the monitor about how many out of the 50 trials they responded correctly to.

2.3.3. Experimental Blocks

Only the four experimental blocks included EEG and eye-tracking measurements. Task difficulty in terms of μ was set to the individualized μ value obtained from the second familiarization block and had a maximum of 1.3. We counterbalanced sounds and trials (in terms of the last sound in a trial) between weak and strong evidence. All sounds that followed an ISI change larger than the mean of the bimodal normal distribution were coded as strong evidence (evidence strongly suggests acceleration/deceleration), those sounds that followed an ISI change lower than the mean ISI change were coded as weak evidence (evidence suggests acceleration/deceleration, but with less certainty, see section *Bayesian Inference in Dynamic Environments*). In the four experimental blocks, subjects did not receive direct feedback after every single trial, but only summarized feedback after each block.

2.4. Variables and Measurements

2.4.1. Behavioral Level

Since subjects only had to report the final state of each sound sequence, on the behavioral level, all variables were measured on a trial-by-trial basis. The dependent variable was response accuracy, that is, whether subjects correctly reported the final state of the sound sequence. The independent variables were 1) the number of sounds between the last changepoint and the end of the trial (sounds after changepoint, SAC level) and 2) the size of the difference between the last ISI and the second to last ISI (evidence level). As established above, we operationalized surprise via the SAC level (see section *Bayesian Updating and Surprise*). In our analysis we included all trials with SAC ≤ 5 (up to five sounds after last CP and before end of trial). This resulted in 10 conditions (two evidence levels × five SAC levels). Trials were counterbalanced with regard to the SAC level, evidence level, length of the last ISI and the total sequence length. Note that on the behavioral level, SAC levels and evidence levels only refer to the last sound of each sequence. On the physiological level, each sound (except the first two) of each sequence was assigned an SAC and evidence level.

2.4.2. Physiological Level

Since physiological measurements in terms of EEG do not rely on the subject's conscious responses, we were able to analyze all sounds of each sound sequence. The first two sounds within each trial were excluded from the analysis since it needs at least three sounds for a sound sequence to be accelerating or decelerating. First, we separated all sounds into two evidence level conditions (weak and strong evidence) according to the relative size of the difference between the two previous ISIs and five SAC level conditions (SAC 1-5) according to their relative position to the previous CP. The sounds before the first CP were assigned to the SAC level conditions according to their relative position to the start of the trial, except for the first two sounds, which were disregarded. Since effect sizes in ERP studies are typically rather small (Luck, 2005), we aimed at comparing the more extreme conditions to one another. For this reason, we pooled SAC 4 and SAC 5 epochs into a low-surprise condition, which we then compared to SAC 1, that was the high-surprise condition. Further, since after each sound there was a 20% chance for a CP, there are more sounds in the lower SAC conditions, however, SAC 4 and SAC 5 combined included roughly the same number of sounds as SAC 1. SAC > 5 sounds were discarded from the analysis, because the number of sounds per condition would have been too small. This resulted in two evidence conditions (weak and strong) × two surprise (low and high) conditions. Our dependent variables were peak amplitudes in the P2, P3a and P3b ERP components.

2.5. Electroencephalography (EEG)

2.5.1. EEG Data Acquisition

The EEG preparation took place after subjects finished the familiarization blocks. We aimed at an impedance of < 25 kΩ for each electrode. During the task, subjects were instructed to move as little as possible, besides using their hands to respond. Also, they were asked not to blink during the presentation of the sounds.

2.5.2. EEG Preprocessing

All preprocessing steps were conducted in MATLAB 9.11.0 (Version R2021b; The MathWorks, Natick, MA) and EEGLAB 2023.0 (Delorme & Makeig, 2004). First, we down sampled the continuous EEG signal to 125 Hz, applied a high-pass filter with 0.25Hz cut-off frequency to eliminate slow drifts (e.g., resulting from sweat) and removed line noise using the 'nt_zapline' function from the NoiseTools toolbox (de Cheveigné, 2019). The data was then visually inspected and noisy channels and segments were manually removed. Removed channels were replaced by using spherical channel interpolation. Channel FCz was used as a reference electrode and later added back to the dataset. The data were re-referenced to an average reference. For that, the mean signal across all electrodes was used which was then subtracted from each individual electrode. Then independent component analysis (ICA) was applied to remove ocular artifacts such as blinks or eye movements. We obtained the ICA weights using the Picard algorithm with principal component analysis (PCA) to control for data rank deficiency caused by channel interpolation. The data was epoched into segments of –500 to 800ms around stimulus onset. Baseline correction was conducted using the interval from –100 to 0ms before stimulus onset. In the last step, we excluded the first and second sound in each trial from the analysis. We aimed at analyzing the EEG data from 0 to 500ms from stimulus onset. To avoid interferences in the neural response with the following sound from overlapping epochs, only sounds where the previous, as well as the following ISI were at least 500ms were included. Epochs were binned according to the factors surprise (high and low) and evidence level (weak and strong). To ensure each condition had an equal number of epochs, a random set of epochs, to match the condition with the smallest number of epochs, was subsampled. Both surprise conditions and both evidence conditions were adjusted to contain 4.859 and 7.345 epochs, respectively. To analyze the interaction between surprise and evidence level, we additionally subdivided the data in a 2 × 2 way, resulting in the following subsamples: high surprise / strong evidence, high surprise / weak evidence, low surprise / strong evidence, low surprise / weak evidence.

2.6. Analyses

2.6.1. Behavioral Data

To analyze the behavioral data, we conducted repeated measures analysis of variance (rANOVA) in MATLAB 9.11.0 (Version R2021b; The Math Works, Natick, MA). Factors were *evidence level* (weak and strong) and *SAC level* (1-5). Where appropriate, we then performed post-hoc t-tests and applied Dunn-Sidak correction for familywise errors.

2.6.2. EEG Data

All EEG data analyses were conducted using the Fieldtrip toolbox (Oostenveld et al., 2011). The aim was to analyze whether there were significant differences in peak amplitude between conditions regarding the P2, P3a, and P3b. For each of the three components, we predefined which channels we would use for the analysis based on previous studies. Former studies hint toward the auditory P2 peak being generated by different anatomical sources but is often observed in fronto-central electrode sites (For a review, see Crowley & Colrain, 2003). Thus, channels Cz, FCz, and Fz were chosen for the P2 analysis. The P3a is known to have a fronto-centrally distributed scalp topography (Kolossa et al., 2015), thus, channels FCz and Cz were used for the P3a analysis. The P3b component is typically measured over parietal regions (Kolossa et al., 2012, 2015), so parietal electrode Pz was chosen for the P3b analysis.

Based on previous literature, we searched for the peak of the P2 component within the time window of 150 to 250 ms post stimulus onset (Corwley & Colrain, 2004). The peak of the P3a component has been found to occur between 300 and 400 ms post stimulus onset, so we chose this as the time window of interest for our analysis. The P3b peak has been found to occur after the P3a peak, between 300 to 600 ms after stimulus onset (Fjell et al., 2009). Therefore, the P3b peak was defined as the second peak between 300 and 500 ms post stimulus onset. Note, that 500 instead of 600 ms stems from our epoch only containing the time span from -100 to 500 ms around stimulus onset. For each of the three components, the maximum in EEG amplitude was calculated within the respective time window and electrodes. Around this peak, a 40 ms time window was defined. Within these 40 ms time windows, the average amplitudes were calculated for each condition. The main effects of surprise and evidence level were analyzed by comparing the means in the high surprise condition to the means in the low surprise condition, and the

means in the strong evidence condition to the means in the weak evidence condition, respectively, via dependent samples t-tests. The interaction between surprise and evidence level was analyzed via a 2×2 rANVOA, with the factors surprise (high / low) and evidence level (weak / strong).

3. Results

3.1. Effects of surprise and evidence level on response accuracy

Trials were separated into ten conditions, resulting from five possible SAC levels (SAC 1-5) and two possible evidence-levels (weak evidence, strong evidence). A repeated measures ANOVA showed significant main effects for SAC ($F(1, 4) = 57.071$, $p < .001$) and evidence level ($F(1, 4) = 36.564$, $p < .001$), as well as a significant interaction between SAC and evidence-levels ($F(1, 4) = 3.5519$, $p < .01$). Post-hoc *t*-tests (Dunn-Sidak corrected) revealed significant differences in response accuracy between all combinations of SAC-levels, except between SAC 4 and 5, regardless of evidence-level (weak evidence: $t(26) = 2.62$, $p = .14$; strong evidence: $t(26) = 1.94$, $p = .48$). The main effect of evidence-level was significant in all SAC levels, again except for SAC 4 and 5 (SAC 4: $t(26) = 1.69$, $p = .10$; SAC 5: $t(26) = 1.43$, $p = .20$). Surprise (in terms of SAC level) as well as evidence level modulated subjects' ability to correctly report the final state of the sound sequence, but only up until a certain amount of evidence was accumulated (Figure 2). The more surprising the last sound of the sequence was, the poorer was subjects' ability to correctly report the final state of the sound sequence. This effect was even stronger for the weak evidence condition. The less surprising the sounds became (i.e., higher SAC levels), the smaller was the mismatch between subjects' assumptions and the actual sensory evidence they received, and therefore the better was their ability to correctly report the final state of the sequence. The interaction between surprise and evidence level suggests, that the effect of evidence level declines with rising SAC levels.

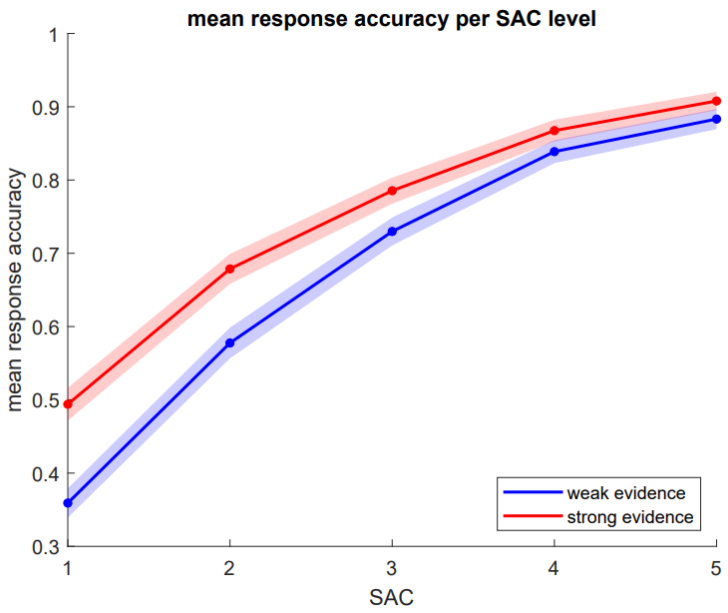

Figure 2: Mean response accuracy for sound-after-changepoint (SAC) levels 1-5 for both weak (blue line) and strong (red line) evidence conditions. The lighter patches represent the standard error.

3.2. EEG Results

3.2.1. ERP Waveform

The grand average ERP measured at channel Cz is shown in Figure 3. The ERP waveform is in line with a typical early auditory brainstem response (ABR; Luck, 2005) directly after stimulus onset, followed by a relatively high peak at around 170ms after stimulus onset, most likely representing the P2 component. The P2 is followed by a positive potential around 330ms after stimulus onset, which is in line with the typical latency of the P3a. After around 400ms post stimulus onset starts a lower-amplitude and more prolonged positive deflection, likely representing the P3b component. There are also negative peaks at around 100ms, 280ms and 390ms, with the first two possibly reflecting the N1 and N2 components (Luck, 2005). Since our hypotheses focus only on positive components, these peaks are not included in the analysis. The peak latencies for the P2, P3a and P3b were 168, 328, and 496 ms post stimulus onset, respectively. The resulting time windows used for the analysis were 148-188 ms, 308 – 348 ms, and 460 to 500 ms. The time window for the P3b had to be manually set, because the peak was too close to the onset of the next stimulus, and therefore to the end of the epoch.

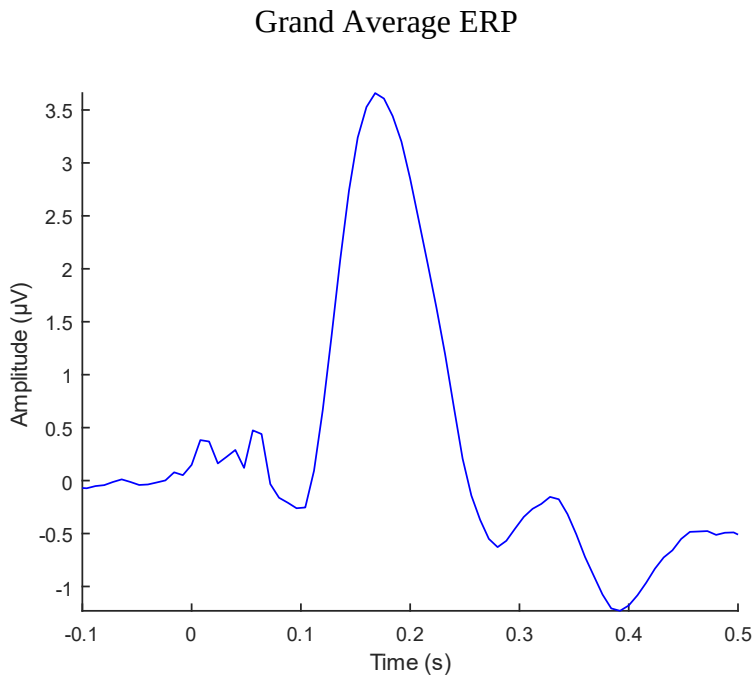

Figure 3: Grand average event-related potential at electrode Cz, averaged over all sounds and subjects.

3.2.2. Surprise Increases P2 and P3a Peak Amplitude

Dependent samples t-test between the high and low surprise conditions revealed significant differences in peak amplitude for all three components. Regarding the P2, highly surprising sounds evoked stronger peak amplitudes compared to less surprising sounds ($t(26) = 2.33$, $p = .027$). Figure 4A shows the ERP measured at the channels we used to analyze the P2. Figure 4B shows the spatial distribution of the difference between the high and low surprise condition averaged over the P2 time window. The effect was strongest in frontal and fronto-central electrodes.

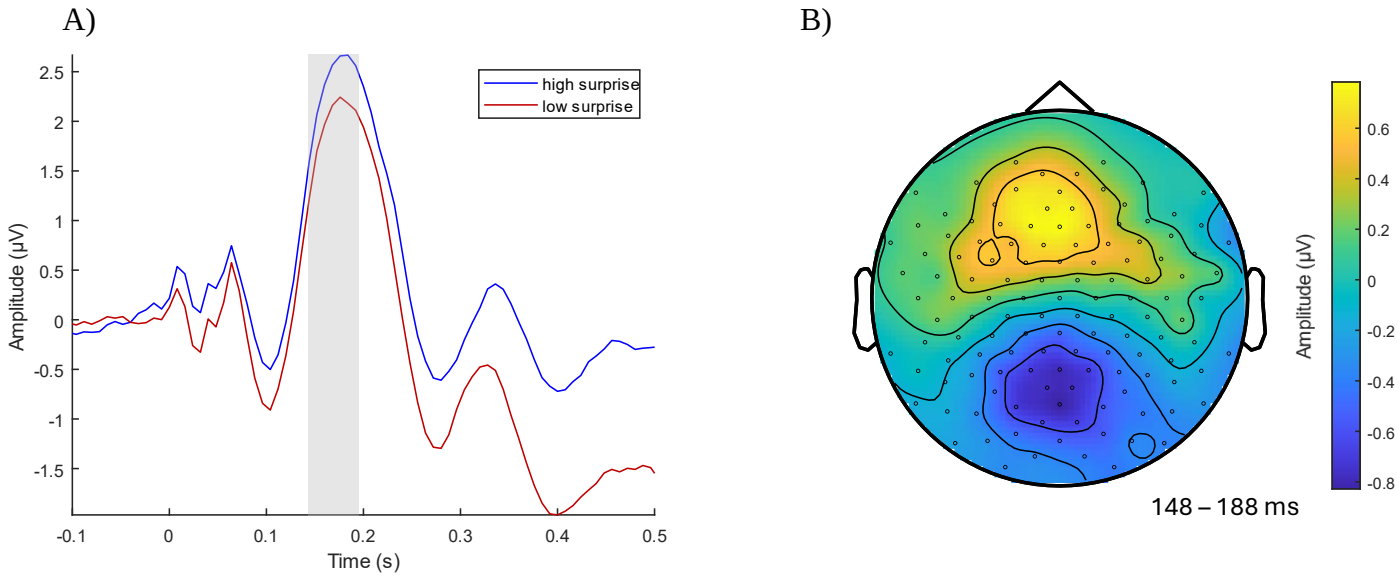

Figure 4: A) Event-related potential measured at Fz, Cz, and FCz for both surprise conditions. The grey area represents the P2 time window. B) Spatial distribution of the difference between high and low surprise averaged over the P2 time window.

Also, for the P3a, highly surprising sounds evoked stronger peak amplitudes compared to less surprising sounds ($t(26) = 2.66$, $p = .013$). Figure 5A shows the ERP measured at the channels we used to analyze the P3a. Figure 5B shows the spatial distribution of the difference between the high and low surprise condition averaged over the P3a time window and reveals the typical fronto-central spatial distribution of the P3a.

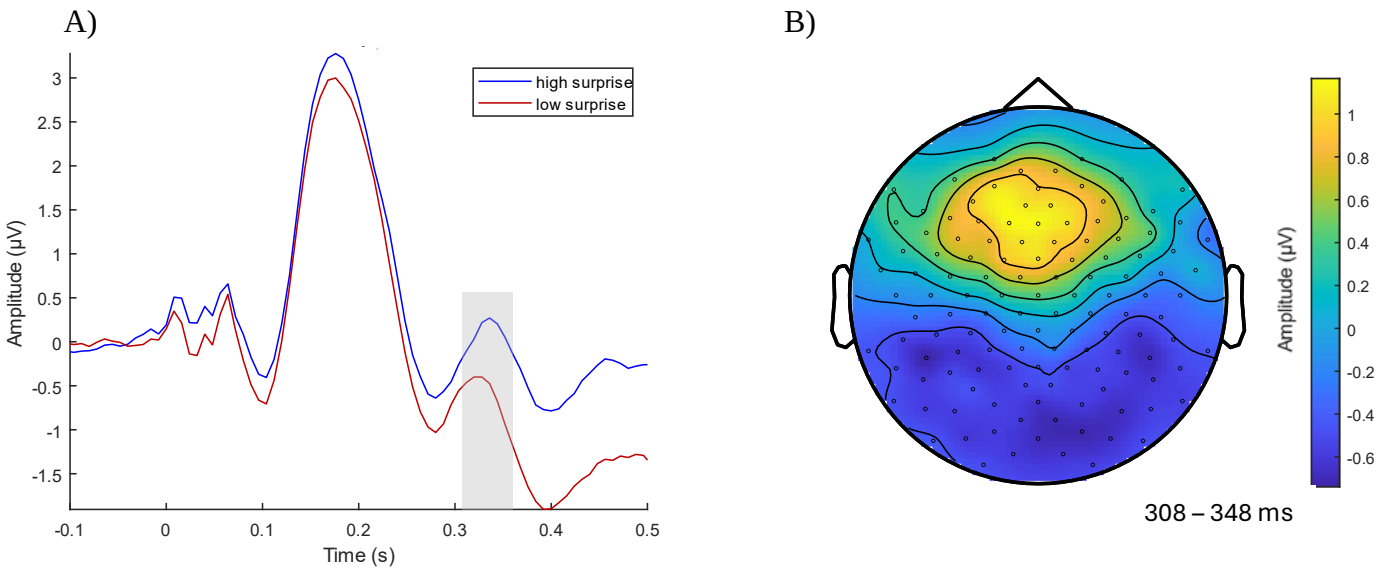

Figure 5: A) ERP measured at Cz and FCz for both surprise conditions. The grey area represents the P3a time window. B) Spatial distribution of the difference between high and low surprise averaged over the P3a time window.

Contrary to the hypotheses regarding the P3b, highly surprising sounds evoked significantly weaker peak amplitudes compared to less surprising sound ($t(26) = -3.54$, $p = .013$). Figure 6A shows the ERP measured at the channel Pz, which was used to analyse the P3b. The P3b seems to peak right at the end of the epoch. Figure 6B shows the spatial distribution of the difference between both surprise conditions. Figure 6C shows the spatial distribution of the signal during the P3b time window in the grand average ERP, indicating that a typical posterior P3b positivity is present, eventhough its peak amplitude is not increased by highly surprising events.

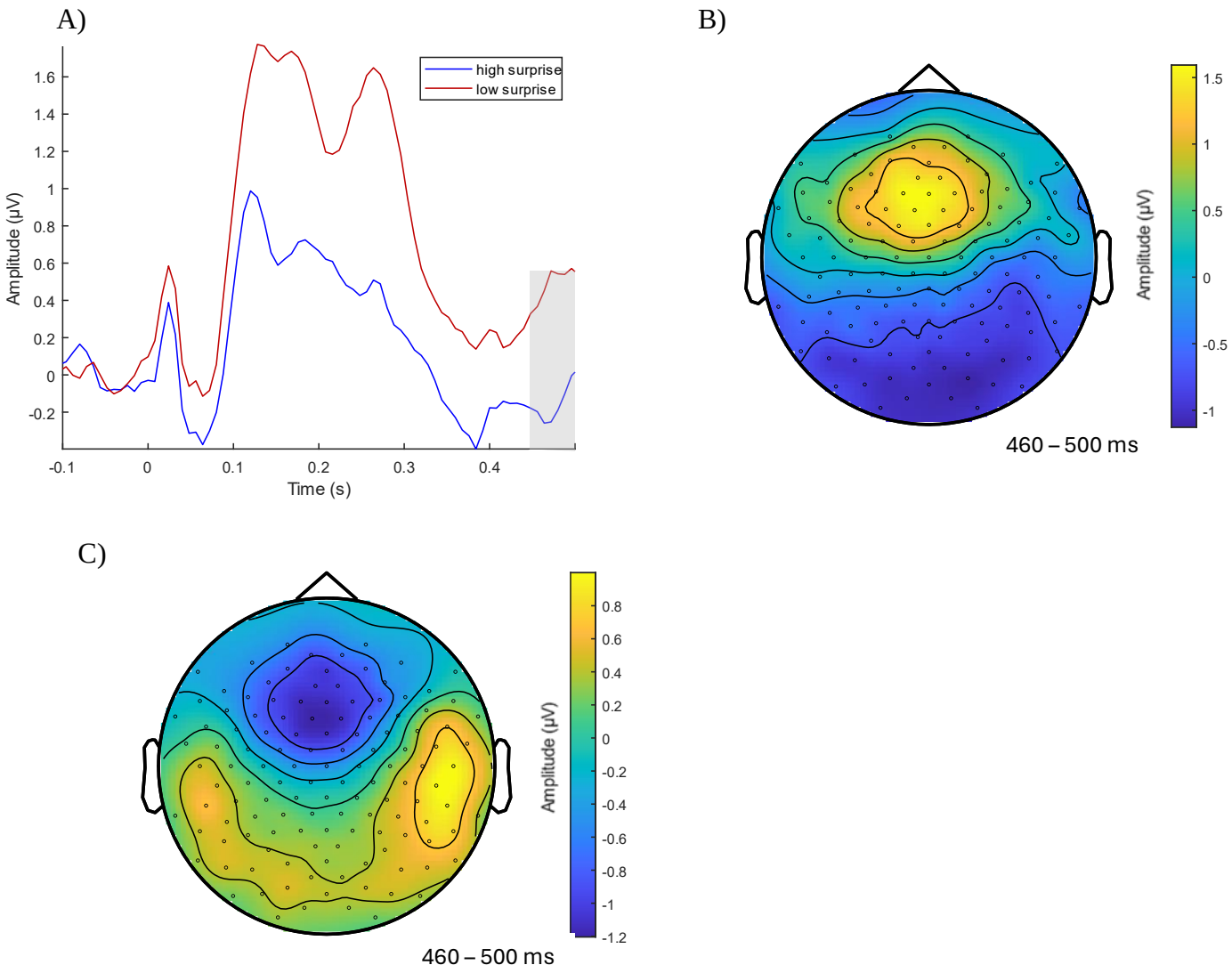

Figure 6: A) ERP measured at Pz for both surprise conditions. The grey area represents the P3b time window. B) Spatial distribution of the difference between high and low surprise averaged over P3b time window. Figure C) Scalp distribution of the P3b in the grand average ERP.

3.2.3. No Neural Signature of Evidence Levels

The ERP measured at fronto-central electrodes Fz, Cz, FCz (Figure 7A) shows slight differences in peak amplitude for the P2 and the P3a component, suggesting slightly higher peaks for the strong evidence condition, yet these differences did not turn out significant in the analysis (P2: $t(26) = -0.593$, $p = .56$; P3a: $t(26) = 0.715$, $p = .48$). There was also no significant effect of evidence level on the peak amplitude in the P3b time window at parietal electrode Pz (Figure 7B, $t(26) = -0.440$, $p = .66$).

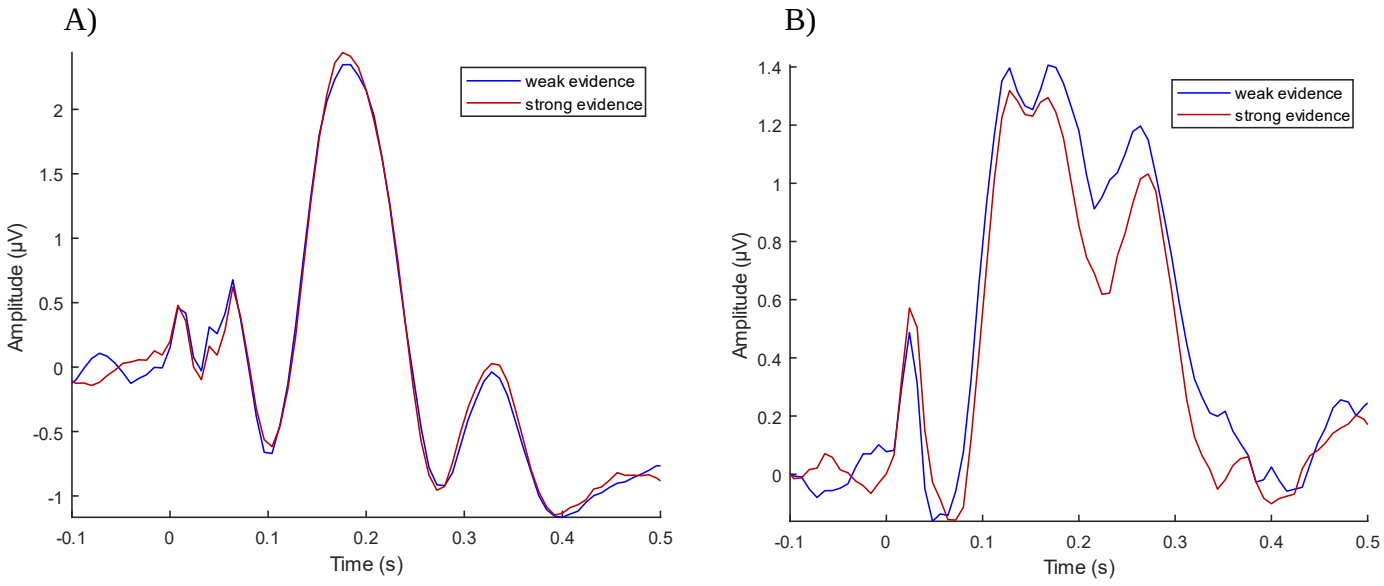

Figure 7: A) ERP for weak and strong evidence measured at fronto-central electrodes Fz, Cz and FCz. B) ERP for weak and strong evidence measured at parietal electrode Pz.

3.2.4. No Modulating Effect of Evidence Level Found on the Effect of Surprise on Peak Amplitudes

Repeated-measures ANOVA with the factors surprise and evidence level revealed no significant interaction in any of the three components (P2: $F(1, 24) = 1.01$, $p = .32$; P3a: $F(1, 24) = 2.21$, $p = .15$; P3b: $F(1, 24) = 0.62$, $p = .44$).

4. Discussion

Bayesian perception posits that the brain constantly generates and updates a model of our surroundings, and that perception results from the integration of incoming sensory evidence and prior expectations. The current study aimed at shedding light on the underlying behavioral and neural mechanisms of Bayesian perception in the context of auditory tempo changes. We linked surprise to impaired task performance and revealed a difference in peak amplitude in the P2, P3a and P3b ERP components between highly and little surprising sounds.

4.1. Different Understandings of Surprise

Within the literature on Bayesian perception, there appears to be an issue concerning the definition of surprise. Modirshanechi et al. (2022) provide an overview, in which they identify 18 different (mathematical) definitions of surprise and classify them in four different categories, based on the specific quantities they measure: prediction surprise, change-point detection surprise, confidence-corrected surprise, and information gain surprise. This should illustrate the variety of definitions used. Despite the differences in their mathematical formulations and specific meanings, in their core almost all of these definitions measure the discrepancy between what is predicted and what actually occurs (Modirshanechi et al., 2022). The variety of surprise definitions surely has a positive side, but can lead to confusion, especially when the specific definition of surprise is not clearly stated in the literature. The current study focusses on the core principle of surprise in terms of discrepancy between prior and sensory input. By implementing SAC levels, we managed to modulate the prior in a very simple way. Regarding the EEG analyses, we simplified our approach even further by contrasting high versus low surprise conditions, instead of single SAC levels to one another. This way, we analyzed the neural correlates of surprise in the context of auditory tempo changes on a basic level and provided a groundwork which future studies can build on.

4.2. Surprise Adversely Affects Task Performance

Subjects' ability to correctly identify the final state of a sound sequence alternating between acceleration and deceleration was strongly impaired, if the last sound of the sequence was highly surprising. The effect surprise had on task performance was modulated by the evidence level, that is, how strongly two consecutive ISIs differed from each other. If the sequence ended with a strong evidence – high surprise sound, subjects responded at chance level. If the last sound was weak evidence – high surprise, subjects responded incorrectly in almost two thirds of the trials. In Bayesian terms, this could be interpreted, as follows: In 50% of the strong evidence – high surprise and 65% of the weak evidence – high surprise sounds, the sensory evidence did not suffice to change the prior accordingly. The resulting posterior and thus the percept (Ma, 2019) therefore did not correctly reflect the final state of the sound sequence. The higher the SAC level, that is, the more evidence available and the stronger the prior, the better subjects performed in the task. This effect went up to SAC level 4 (i.e., four sounds safter the last CP). There was no difference in task accuracy between SAC 4 and SAC 5. Also, the effect of evidence level on task accuracy was only significant up until SAC 4. It seems that evidence levels only play a role up until a certain prior strength, and that the effect of prior strength on task accuracy plateaus at a certain threshold point, beyond which additional evidence does not further enhance task accuracy. Visalli et al. (2021) found comparable effects of surprise on task performance in the visual domain. They implemented a variable foreperiod paradigm, where subjects had to respond as quickly as possible to the appearance of a visual target stimulus. The less subjects were able to predict the onset of the target, that is, the more surprising it was, the longer it took them to respond to the target. This effect is explained by the notion that surprising events require more cognitive effort to encode, therefore increasing the reaction times (Visalli et al., 2021). Other studies exploring surprise in Bayesian perception mostly focus exclusively on physiological measurements, which makes it uneasy to contextualize the current behavioural results. Future studies should also focus on the effects of surprise on a behavioural level, for example in terms of error rates or reaction times.

4.3. Auditory P3 Event-Related Potential Reflects Surprise in Temporal Tasks

The P3 group (involving the P3a and P3b) seems to play a vital role in understanding the neural mechanisms underlying Bayesian perception (Visalli 2023). The results showed that surprising events triggered a positive modulation of peak amplitudes in the P3a subcomponent of the P3 group. We did not find increased peak amplitudes in the component most likely relating to the P3b (Figure 6C) at the typical posterior electrode sites. In fact, our results suggest, that surprising events triggered significantly lower P3b peak amplitudes compared to less surprising events. This effect is most likely an artifact produced by a rather long lasting P3a, overlapping with the P3b (see section *P3 Topography*). Given the current data, it is impossible to properly interpret the link between surprise and the P3b. Thus focusing on the P3a, our results are in line with a growing body of research suggesting that the P3 represents processing of surprising events and possible Bayesian belief updating (Bennett et al., 2015; Kolossa et al., 2014; Visalli., 2021; 2023). Furthermore, the current findings are in line with the notion that the relationship between surprise and the P3 holds in the temporal domain (Visalli et al., 2023), while expanding this finding from the visual to the auditory sense.

4.4. Different P3 Subcomponents Might Represent Different Mechanisms

As already discussed, (see section *Research Aims and Hypotheses*) the aim of the current study was not to computationally disentangle surprise and belief updating and link them to specific subcomponents, but to provide general insights into how events that contradict our prior assumptions are processed on a neural level. While there is no consensus about which subcomponent of the P3 group reflects which of the two concepts, there are at least some studies that found a very similar P3a as in the current study related to surprise, and a later, more posterior P3b supposedly linked to belief updating (Kolossa et al., 2014; Visalli et al., 2023). In the current paradigm, we did not separate surprise and belief updating, and are therefore unable to draw any relevant conclusions regarding this issue.

4.5. P3 Topography

The P3a subcomponent is mostly detected in fronto-entral electrodes (Katayama & Polich, 1999). This is also reflected in the scalp distribution of the difference between high and low surprise around the P3a (Figure 4A), where the strongest difference can be observed in frontal and fronto-central electrode sites. The scalp distribution of the difference between high and low surprise in the P3b time window is only slightly more posterior compared to the P3a. It seems like this topography does in fact not represent a P3b, but still captures the positivity linked to the P3a. The P3a is generally linked to frontal electrode sites and is thought to be generated in frontal areas. Especially the anterior cingulate cortex seems to play a role in P3a generation. The P3b is thought to be generated by parietal and temporal cortical structures and is typically observed in more parietal electrode sites (Polich, 2007). The spatial distribution of the P3b is influenced by a range of factors, such as the specific experimental paradigm, stimulus characteristics and task demand. Specific experimental contexts (e.g., that subjects did not have to directly respond to most stimuli) could be one explanation for the anterior distribution of the P3b difference that was observed (Gaeta et al., 2003). A more plausible explanation could be a strong overlap between the P3a and the P3b (Hartbauer et al., 2001; Polich, 2007). It is possible, that the effect on fronto-central electrodes observed around the P3a time window is still rather strong during the P3b time window and overshadows a possible effect of the P3b, in fact even reversing it due to the dipolar nature of neural fields. Figure 6C shows the scalp distribution of the EEG grand average during the P3b time window. Contrary to the plot of the high and low surprise difference wave (Figure 6B), this fits the typical posterior scalp distribution of the P3b component to at least some extent. The P3a has been linked to represent unexpected or salient stimuli, while the P3b was found to be linked to task relevance (for a review, see Polich, 2007). It can be argued that highly surprising sounds are both salient, since they indicate a change in the state of the sound sequence as well as relevant for the task, since subjects had to update their internal model of the sequence. Based on the scalp topography depicted in Figure 6C, it appears that the current paradigm did elicit a P3b like component, yet given the current data, it is impossible to tell if this is in fact the case, and if yes, how the P3b peak amplitude is linked to surprise. I is also possible that the posterior positivity depicted in Figure 6C does not represent a P3b, but results from an anteriorly distributed negativity. While highly surprising sounds might be relevant for the task due to model updating, subjects did not have to respond immediately to

the majority of stimuli, as they are required to in other ERP paradigms that elicited a P3b (e.g., Visalli et al., 2021, 2023). Future research should investigate, if surprise in auditory temporal tasks is in fact more closely related to P3a modulations than to P3b modulations, or if the current results regarding (absence of) the P3b are based on artifacts resulting from the specific experimental and analytical context.

4.6. P2 Might Reflect Comparison Mechanisms

In line with the hypothesis, we found a significant modulation of peak amplitude in the P2 that was linked to surprising events. Visalli and colleagues (2023) computationally separated surprise from belief updating. While they linked the P3 to surprise in the context of visual temporal predictions, they reported an association between the P2 and belief updating. From a theoretical point of view, these results are rather surprising. Belief updating often occurs as a result of surprising events, so it seems unlikely that updating is neurally encoded earlier than surprise. Again, the current study does not allow for a separate analysis of surprise and belief updating, so this question is yet to be answered. Nevertheless, it is still possible that the findings from Visalli and colleagues (2023) also hold here. The functional meaning of the P2 component is still relatively unclear (Luck, 2005). It is possible that the P2 represents a signature for comparing mechanisms between information stored in memory and sensory input (Freunberger et al., 2007). Subjects constantly had to compare the latest ISI to the previous ISI to deduce if the sequence was speeding up or slowing down. Nevertheless, this was the case for all sounds and not only for highly surprising sounds, so it is unlikely that this can explain higher peaks in the high surprise condition. The increased P2 peak amplitude for surprising sounds might also index the recruitment of additional cognitive resources needed for sounds that strongly deviate from the subjects' prior assumptions (Visalli et al., 2021). Studies on time perception also examined the P2 component. It seems that the P2 reflects the temporal distance between a learned standard interval and a longer comparison interval. This might play a role in the current paradigm, especially for sounds following long ISIs, but does not explain the link between the P2 and surprise. Visalli and colleagues (2023) argue that the P2 reflects the neural computation of the posterior. Given the fact, that highly surprising sounds deviate strongly from the prior, and therefore the computation of the posterior might be cognitively more costly, the interpretation of

Visalli and colleagues (2023) could theoretically explain the relationship between surprise and P2 amplitude. Future research should focus on the specific role of the P2 in auditory Bayesian perception.

4.7. No Physiological Signatures of Evidence Level

Contrary to hypothesis H2b, evidence level did not significantly modulate the effect surprise had on the peak amplitudes in any of the peaks we analyzed. Furthermore, there was no main effect of evidence level in the physiological data. On the behavioral level, evidence level significantly modulated participants task accuracy and significantly interacted with SAC level. Especially in the lower SAC levels, these effects were quite strong. Therefore, it is surprising that the EEG data did not reveal any sort of neurophysiological signature of evidence level. This also raises questions about the notion that belief updating might be reflected in the P2 or P3 component (Kolossa et al., 2015; Visalli., 2021; 2023). Theoretically, evidence level and belief updating should be correlated. The evidence level represents how clearly the sound sequence is accelerating or decelerating (see section *Variables and Measurements*). Note that strong evidence does not directly correspond to higher updating. For instance, if the sound sequence has already been accelerating for 6 sounds, a seventh, high-evidence sound will not trigger a strong update of the prior, since the prior is already very strong. In the high surprise condition on the other hand, strong evidence sounds should trigger stronger updates of the prior, given the likelihood of the new input not fitting the current prior is larger (see Section Bayesian Inference in Dynamic Environments). Keep in mind that this is mere theoretical speculation but let us assume that high evidence is linked to higher updating in the high surprise condition. There is no consensus in the literature on which of the P2, P3a or P3b components reflect updating, but several studies hint that there is a link between belief updating and one of these components (Kolossa et al., 2015; Visalli et al., 2021; 2023). Following our hypothetical association between evidence level and belief updating, then evidence level should also be represented in one of the components, which is not the case in the current findings. In the past, only few studies have focused on experimentally manipulating the likelihood (which can loosely be translated to the evidence level) and analyzing its underlying implications and neural representations. Goodwin et al. (2023), for example, found that greater uncertainty in the likelihood of observed sensory

information negatively affects subjects' ability to correctly infer the current state of their environment. This is in line with the current findings, that suggest an adverse effect of uncertainty regarding the likelihood (i.e., weak evidence) on subjects' ability to correctly report the final state of the sound sequences. Future research should address this issue more directly.

4.8. Limitations

In the current study, we deliberately avoided a computational separation of surprise and belief updating, and rather focused on the broader neural processing mechanisms underlying auditory Bayesian perception. We provide evidence that the foundational concepts of Bayesian perception hold in the context of auditory tempo changes. Future research could apply computational and/or experimental methods to disentangle the concepts of surprise and belief updating in order to find out about their specific underlying mechanisms and neural representations.

In our experiment, we modelled two possible states of the world: acceleration or deceleration. This adequately fits the highly specific context of auditory tempo changes. In fact, there is one more possible state, that is, the absence of tempo change. Follow up studies could implement the possibility that two consecutive ISIs do not differ at all in addition to acceleration and deceleration. Needless to say, our acoustic environment can resemble countless numbers of different states in everyday life. Future research should aim at implementing more naturalistic settings when examining Bayesian perception.

Furthermore, higher P3 amplitudes in highly surprising sounds might (to some extent) be explained by involuntary shifts in attention towards the task in response to CPs. Since change itself can be inherently salient (Liesefeld et al., 2017), it is possible that subjects responded with some form of mental resource allocation to CPs, which has been linked to stronger P3 amplitudes in the past (Luck, 2005).

5. Conclusion

The current study supports the hypothesis, that humans follow the principles of Bayesian inference in perceiving auditory tempo changes. Our behavioral results support the notion, that after a change in the environment, subjects accumulate evidence in order to adapt their internal model of the world accordingly. This process is influenced by how surprising new sensory input is, that is, the discrepancy between their prior assumptions about the world and the sensory signal they receive, as well as by how strongly the new sensory input suggests one state of the world or another. We further linked the processing of highly surprising sounds to increased P2 and P3a peak amplitudes. Contrary to our hypotheses, we did not find a relevant link between surprise and P3b peak amplitude, nor did we discover any neural signatures of evidence level. The study provides basic insights into auditory Bayesian perception and should encourage future research to implement increasingly complex and naturalistic paradigms to further investigate these ERP components and their underlying neural mechanisms.

6. Acknowledgements

Data collection, analysis and the writing of this thesis happened within the Dynamates project at the Acoustics Research Institute (ARI) of the Austrian Academy of Sciences (Young Independent Researchers Group Dynamates, ZK66). I received funding from the Austrian Science Fund (FWF). I want to express my thanks to my co-supervisor, Dr. Ulrich Pomper, for the guidance and especially the trust in allowing me to participate in this project, to Burcu Bayram, for spending countless hours answering my silly questions and making me familiar with the methodology, to Dr. David Meijer for explaining the specifics of Bayesian inference over and over again, to my supervisor Prof. Dr. Ulrich Ansorge for teaching me to always strive for improvement, and to Dr. Robert Baumgartner for always having an open ear when big decisions had to be made. Furthermore, I want to thank my desk mate Roman Fleischmann and my internship supervisor Tobias Greif for the fruitful discussions. Last but not least, I want to thank everybody at ARI for the warm welcome to the institute.

7. Literature

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

Aristotle. (1986). De anima (J.A. Smith, Trans.). Princeton University press. (Original published circa 350 BCE)

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology, 91*(4), 612–625. https://doi.org/10.1037/0022-3514.91.4.612

Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research, 184*(3), 411–425. https://doi.org/10.1007/s00221-007-1114-x

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Barumerli, R., Majdak, P., Baumgartner, R., Geronazzo, M., & Avanzini, F. (2020). Evaluation of a human sound localization model based on Bayesian inference. 5 pages. https://doi.org/10.48465/FA.2020.0704

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bennett, D., Murawski, C., & Bode, S. (2015). Single-trial event-related potential correlates of belief updating. *Eneuro, 2*(5), ENEURO.0076-15.2015.

https://doi.org/10.1523/ENEURO.0076-15.2015

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*(1), 49–57. https://doi.org/10.1007/s00221-009-1933-z

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Cavanagh, P., & Alvarez, G. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349–354. https://doi.org/10.1016/j.tics.2005.05.009

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

Cogley, T., & Sargent, T. J. (2008). Anticipated utility and rational expectations as approximations of Bayesian decision making. *International Economic Review, 49*(1), 185–221. https://doi.org/10.1111/j.1468-2354.2008.00477.x

Crowley, K. E., Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

de Cheveigné, A. (2019). ZapLine: a simple and effective method to remove power line artifacts. *NeuroImage, 207*, 116356.

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

Den Ouden, H. E. M., Kok, P., & De Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience*, *2*(12), 920–926. https://doi.org/10.1038/35104092

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, *7*(6), 509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Fjell, A. M., Rosquist, H., & Walhovd, K. B. (2009). Instability in the latency of P3a/P3b brain potentials and cognitive function in aging. *Neurobiology of Aging*, *30*(12), 2065–2079. https://doi.org/10.1016/j.neurobiolaging.2008.01.015

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

Freunberger, R., Klimesch, W., Doppelmayr, M., & Höller, Y. (2007). Visual P2 component is related to theta phase-locking. *Neuroscience Letters*, *426*(3), 181–186. https://doi.org/10.1016/j.neulet.2007.08.062

Friston, K. (2005). A theory of cortical responses. *Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gaeta, H., Friedman, D., & Hunt, G. (2003). Stimulus characteristics and task category dissociate the anterior and posterior aspects of the novelty P3.

Psychophysiology, *40*(2), 198–208. https://doi.org/10.1111/1469-8986.00022

Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D., & Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. *PLOS Computational Biology*, *17*(2). https://doi.org/10.1371/journal.pcbi.1008068

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Goodwin, I., Hester, R., & Garrido, M. I. (2023). Temporal stability of Bayesian belief updating in perceptual decision-making. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02306-y

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Hartbauer, M., Hutter-Paier, B., Skofitsch, G., & Windisch, M. (2001). Antiapoptotic effects of the peptidergic drug Cerebrolysin on primary cultures of embryonic chick cortical neurons. *Journal of Neural Transmission*, *108*(4), 459–473. https://doi.org/10.1007/s007020170067

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing

Brain. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00096

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

Katayama, J., & Polich, J. (1999). Auditory and visual P300 topography from a 3 stimulus paradigm. *Clinical Neurophysiology*, *110*(3), 463–468. https://doi.org/10.1016/S1388-2457(98)00035-2

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

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

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), 0107. https://doi.org/10.1038/s41562-017-0107

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

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron, 84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

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

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological aata. *Computational Intelligence and Neuroscience*, 1–9. https://doi.org/10.1155/2011/156869

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage, 62*(1), 177–188. https://doi.org/10.1016/j.neuroimage.2012.04.050

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science, 17*(4), 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x

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

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology, 109*(2), 160.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Rock, I, M., A. (1998). Inattentional Blindness. MIT Press.

Rohe, T., Ehlis, A.-C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications, 10*(1), 1907. https://doi.org/10.1038/s41467-019-09664-2

Sawai, K., Sato, Y., & Aihara, K. (2012). Auditory time-interval perception as causal inference on sound sources. *Frontiers in Psychology, 3.* https://doi.org/10.3389/fpsyg.2012.00524

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences, 14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Shen, S., & Ma, W. J. (2019). Variable precision in visual perception. *Psychological Review, 126*(1), 89–132. https://doi.org/10.1037/rev0000128

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception, 28*(9), 1059-1074.

Skerritt-Davis, B., & Elhilali, M. (2018). Detecting change in stochastic sound sequences. *PLOS Computational Biology, 14*(5). https://doi.org/10.1371/journal.pcbi.1006162

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), 745–756. https://doi.org/10.1038/nrn3838

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009

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

Valakos, D., d'Avossa, G., Mylonas, D., Butler, J., Klein, C., & Smyrnis, N. (2020). P300 response modulation reflects breaches of non-probabilistic expectations. *Scientific Reports, 10*(1), 10254. https://doi.org/10.1038/s41598-020-67275-0

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

Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2023). P3-like signatures of temporal predictions: A computational EEG study. *Experimental Brain Research, 241*(7), 1919–1930. https://doi.org/10.1007/s00221-023-06656-z

Visalli, A., Capizzi, M., Ambrosini, E., Mazzonetto, I., & Vallesi, A. (2019). Bayesian modeling of temporal expectations in the human brain. *NeuroImage*, *202*, 116097. https://doi.org/10.1016/j.neuroimage.2019.116097

Zénon, A., Solopchuk, O., & Pezzulo, G. (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia*, *123*, 5–18. https://doi.org/10.1016/j.neuropsychologia.2018.09.013

8. List of Figures

9. List of Abbreviations

EEG	Electroencephalography
ERP	Event-related potential
CP	Change point
SAC	Sounds after change point
ISI	Inter-stimulus interval
rANOVA	Repeated-measures analysis of variance

10. Appendix

10.1. Abstract

The Bayesian perception framework posits, that the brain constantly generates and updates a model of our surroundings, and that perception results from the integration of bottom-up sensory evidence and top-down expectations reflecting prior knowledge. The present study investigates whether humans follow the principles of Bayesian inference in the case of auditory temporal estimations. To that end, we presented 27 subjects with sound sequences of unpredictable length that alternated between acceleration and deceleration at random change points (CPs), while high-density electroencephalography (EEG) was being recorded. At the end of each trial, subjects had to report the final state of the sound sequence. The closer sounds followed a CP, the bigger was the discrepancy between subjects' prior beliefs and the sensory input. Response accuracy significantly increased with the number of sounds after the last CP. This effect was negatively modulated by the evidence level, defined as the relative size of the difference between the last two consecutive inter-stimulus intervals (ISIs). In general, higher evidence levels were associated with higher response accuracy. Event-related potential (ERP) analysis revealed stronger P2 and P3a peak amplitudes for sounds that directly followed a CP, compared to the 4^{th} and 5^{th} sound after a CP. No neural representations of evidence level were found.

10.2. German Abstract

Das Rahmenmodell der Bayesianischen Wahrnehmung geht davon aus, dass unser Gehirn kontinuierlich ein Modell unserer Umwelt erstellt und ständig aktualisiert, und dass Wahrnehmung aus der Integration von äußeren Sinneseindrücken und Erwartungen, basierend auf bestehendem Vorwissen, resultiert. Die vorliegende Studie untersucht, ob Menschen den Prinzipien der Bayesianischen Inferenz folgen, im Kontext von auditiven, zeitlichen Prädiktionen. Wir präsentierten 27 Versuchspersonen Tonsequenzen von unvorhersehbarer Länge, die zufällig zwischen Beschleunigung und Verlangsamung wechselten, während ein hochauflösendes Elektroenzephalogramm (EEG) aufgezeichnet wurde. Nach jedem Durchgang mussten die Versuchspersonen den finalen Zustand der Tonsequenz rückmelden. Je unmittelbarer die Töne einem Wendepunkt (WP) folgten, desto größer war die Diskrepanz zwischen den Prädiktionen der Versuchspersonen und dem sensorischen Signal. Die Genauigkeit der Antworten war stieg signifikant, mit der Anzahl an Tönen nach dem letzten WP. Dieser Effekt wurde durch das Evidenzniveau verstärkt, was definiert war als die relative Größe des Unterschieds zwischen den letzten beiden aufeinanderfolgenden Interstimulusintervallen (ISIs). Stärkere Evidenzniveau waren mit höherer Antwortgenauigkeit assoziiert. Die Analyse der Ereigniskorrelierten Potentiale (EKPs) ergab stärkere P2 und P3a Spitzenamplituden bei Tönen, die direkt auf einen WP folgten im Vergleich zu Tönen, die an vierter bzw. fünfter Stelle nach einem WP auftraten. Es zeigten sich keine neuronalen Repräsentationen des Evidenzniveaus.