Expectation suppression across sensory modalities: a MEG investigation

Doctoral thesis by
Sanjeev Nara

Supervised by
Nicola Molinaro and **Mathieu Bourguignon**

Donostia, Spain 2022

Sanjeev Nara

Basque Center on Cognition, Brain and Language
Paseo Mikeletegi 69
Donostia-San Sebastián, Spain
December 2022

Expectation suppression across sensory modalities: a MEG investigation

Doctoral thesis by
Sanjeev Nara

Supervised by
Nicola Molinaro and **Mathieu Bourguignon**

Donostia, Spain 2022

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
NAZIOARTEKO BIKAINTASUN CAMPUSA
CAMPUS DE EXCELENCIA INTERNACIONAL
bcbl
BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

This research was supported by a Predoctoral grant (Ayudas para contratos para la formación de doctores reference number: BES-2016-077560, project number PSI2015-65694-P) from the Spanish Ministry of Economy and Competitiveness (MINECO). The research was also supported by the European Molecular Biology Organization (EMBO) short term grant (number 8020).

ACKNOWLEDGMENT

The research work presented in the thesis work is carried out at the "Basque Center on Cognition Brain and Language" (BCBL), under Dr. Nicola Molinaro, Scientist, BCBL, Spain, and Dr. Mathieu Bourguignon, ULB, Brussels, Belgium.

I want to express my sincere gratitude to Dr. Molinaro and Dr. Bourguignon for their continuous support, patience, motivation, and delivering immense knowledge during the whole course of my Ph.D. Their guidance helped me in all the research and writing of this thesis. It was an absolute pleasure to be under their supervision and learn from them. I never realized when the mentor-mentee relationship turned into an invaluable friendship. The support I have received from Dr. Molinaro and Dr. Bourguignon cannot be expressed on such a small piece of paper; hopefully, I need to write another thesis on their generosity, knowledge, motivation, and support.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Ana Belén Chica Martínez (University of Granada, Spain), Dr. Svetlana Pinet (BCBL/UPV, Spain) and Prof. Girijesh Prasad (Ulster University, UK) for their insightful comments and encouragement, which incited me to widen my research from various perspectives.

Apart from my supervisors, I am also thankful to Dr. Radoslaw Martin Cichy, FUB, Berlin for hosting me as a visiting student in his lab. I would like to thank my group mates Dr. Mikel Lizarazu, Dr. Craig Richter (Guruji), and Dr. Anastasia Klimovich-Gray. They helped me grow in different data analysis techniques and bore endless questions. Apart from all the members of Brain Rhythms and Cognition group: Brhyco's group (former Proactive Group), I am

thankful to my friends Dr. Amit Jaiswal, Scientist, MEGIN Finland, and Dr. Dheeraj Rathee, Scientist, Essex University, UK, for always motivating me.

I want to acknowledge the IT department of BCBL (Jose, Borja, Guti, and team); special thanks to the admin team (Vanessa, Eider, Ana, Maider, and Larraitz) as well the labs' team (Larraitz, Oihana, Manex, David and the rest of the team). All these guys work as the pillars of BCBL who, in their capacity, contribute to making the BCBL working environment fun and easy-going.

Last but not least, I would like to thank my parents, Capt. Randhir Singh and Somwati Devi who always encouraged me to pursue my dreams. My father's hard work in his career always motivates me to work hard and be like him in my life. I am thankful to my brother Pankaj Nara, who took care of all my responsibilities towards the family in my absence at my home. This thesis could not have been written without my wife, Neha Malik, who always supported me in every aspect of life. I want to apologize to my newborn babies Paarth Nara and Prisha Nara, as I take some time from the time-period I am supposed to spend with them.

Dedicated to my family

Resumen:

Se ha propuesto al procesamiento predictivo como un mecanismo cognitivo fundamental que explica cómo el cerebro interactúa con el entorno externo a través de las modalidades sensoriales. La idea principal subyacente a las teorías sobre el procesamiento predictivo es que el cerebro humano desarrolla un modelo generativo (es decir, una representación interna) del entorno que le rodea, el cual se utiliza constantemente para generar inferencias *top-down* sobre la causa externa de la energía que impacta en nuestros sentidos. Este modelo interno se actualiza constantemente mediante la comparación entre la información predicha y la actual. Este proceso se basa en la interacción bidireccional entre las regiones de procesamiento de bajo nivel y alto nivel que tienden a segregarse jerárquicamente en la corteza humana (como la corteza visual). Las regiones de orden superior desarrollan predicciones sobre la información externa, mientras que las regiones de orden inferior aportan información sobre la información externa. La comparación entre estas dos fuentes de información (top-down de las regiones de orden superior y bottom-up de las regiones de orden inferior) se utiliza para determinar la identidad de las estimulaciones externas. La diferencia entre la información actual y la predicción se denomina error de predicción, una señal que se utiliza para actualizar el modelo generativo interno del entorno que nos rodea.

En las últimas dos décadas, gran parte de la investigación se ha centrado en entender cómo el cerebro humano genera expectación sobre las respuestas sensoriales entrantes y cómo gestiona la información inesperada o impredecible. En la literatura sobre procesamiento predictivo se evidencia que el cerebro humano suprime las respuestas neurales a estímulos predecibles/esperados (denominado como efecto de supresión de expectativa), y aumenta las

respuestas neurales para los estímulos impredecibles/no esperados. Gran parte de la investigación se ha centrado en cómo el cerebro humano procesa los estímulos sensoriales para entender el contexto/identidad («¿*qué* es?») de la información. Esta podría ser tanto predecible como impredecible, dependiendo de la información previa sobre el estímulo. Cabe destacar que, en situaciones reales, los objetos del entorno que nos rodea son «temporalmente dinámicos». Mientras que ciertos estímulos son temporalmente regulares y, por tanto, predecibles, otros podrían ser altamente impredecibles en el tiempo. Esta propiedad se define como predictibilidad temporal (es decir, «¿*cuándo* está ocurriendo?»). El cerebro procesa la información externa sobre el contenido, «*qué*», y el momento justo o *timing*, «*cuándo*», de los estímulos del entorno para entender el mundo a su alrededor. En esta tesis he empleado la técnica de magnetoencefalografía (MEG) para entender cómo las propiedades predecibles *qué* y *cuándo* de un estímulo afectan a la supresión de la expectativa en las modalidades sensoriales (es decir, visual y auditiva). La MEG nos permite registrar la actividad del cerebro humano con una alta resolución temporal y una buena resolución espacial, lo que la convierte en la modalidad ideal para investigar la actividad cerebral, que es espaciotemporal por naturaleza.

Se han utilizado diversos diseños experimentales, como el potencial de disparidad, los paradigmas de coincidencia y desajuste o los paradigmas de supresión de repetición y expectativa, para entender la percepción visual y auditiva. Algunas de las limitaciones de estos experimentos fueron que: a) utilizaban información contextual reducida con una señal o *cue* y un objetivo o *target*, b) el procesamiento de los estímulos predichos/esperados se vio afectado por la tarea de los participantes, mezclando así la predicción y la atención y c) los estímulos predecibles (o *predicciones*) se compararon con los estímulos impredecibles/violaciones (*errores de predicción*), pero estas condiciones se han asociado con procesos neurocognitivos muy diferentes.

Teniendo en cuenta estas limitaciones, diseñé un paradigma experimental que presenta una serie de señales (*cues*) que podrían tanto predecir las propiedades/características del objetivo (*target*) o no. Asimismo, introduje una manipulación temporal en el experimento, que podía predecir o no el momento justo del estímulo entrante (es decir, momento predecible o *predictable timing*) o no (momento aleatorio/fluctuante o *random/jittered timing*). Para resolver la controvertida relación entre la atención (inducida por la tarea del experimento) y las expectativas, se mantuvo una relación ortogonal entre la tarea del participante y la característica de predictibilidad; por ejemplo, los participantes reportan una característica del estímulo y la predicción se relaciona con otra característica. Además, la comparación se diseñó entre predicciones vs. no predicciones, de forma que no había violación.

En general, he observado que las áreas visual y auditiva procesan las señales sensoriales de distinta forma. El área visual es más sensible al aspecto *qué* de las señales sensoriales que el *cuándo*. El área visual muestra sensibilidad a la incertidumbre temporal (*cuándo*) solo cuando el contenido (*qué*) es predecible. Si no existe predictibilidad en cuanto al contenido, la incertidumbre temporal no ejerce ninguna influencia. El área auditiva, por otro lado, es igual de sensible a los aspectos *qué* y *cuándo* del procesamiento sensorial. En contraste con el área visual, el área auditiva muestra sensibilidad a la incertidumbre temporal incluso en ausencia de la predictibilidad del *qué*.

Más allá de la interacción entre *cuándo* y *qué*, observé que el área auditiva muestra, en general, un efecto de supresión de expectativa más robusto al estímulo predecible inminente, en comparación con el área visual. Por otra parte, la tarea (ortogonal a la manipulación de predictibilidad) eliminó por completo los efectos de predictibilidad en el área visual, al contrario que en el caso del área auditiva.

Esta tesis contiene seis capítulos principales. El primer capítulo es una introducción al marco del procesamiento predictivo, donde se abordan las nociones de incertidumbre temporal y supresión de la expectativa. Asimismo, este capítulo incluye los objetivos e hipótesis de la presente tesis, así como un resumen básico sobre la neuroanatomía de los sistemas visual y auditivo para situar mejor al lector con respecto al tema.

El segundo capítulo es una introducción general a los métodos que he utilizado para esta tesis. En este capítulo he incluido detalles sobre la modalidad de magnetoencefalografía (MEG), su instrumentación y las diversas técnicas de preprocesamiento de datos y análisis empleados en esta tesis. Este capítulo se centra más en los aspectos teóricos y explica tanto los principios generales como las matemáticas subyacentes a todas las técnicas.

El tercer capítulo contiene los detalles sobre el primer experimento centrado en la supresión de la expectativa en el área visual. Incluye una descripción detallada del diseño experimental, detalles sobre los métodos empleados para el análisis de datos, resultados y debate. En este capítulo he incluido evidencias que respaldan que el área visual es más sensible a las expectativas basadas en el contenido (*qué*). Además, la información en las respuestas neurales específicas del estímulo (rastreado vía modelos de descodificación) aumenta en la medida que se construyen las expectativas sobre el estímulo.

El cuarto capítulo contiene los detalles sobre el segundo experimento, basado en la supresión de la expectativa en el área auditiva. Este capítulo sigue la misma estructura que el capítulo anterior, proporcionando detalles sobre el diseño experimental, los métodos empleados para el análisis de datos, resultados y debate. Aquí he explicado que el área auditiva es igual de

sensible tanto a las expectativas basadas en el contenido (*qué*) como a las temporales (*cuándo*). Asimismo, aporto pruebas que respaldan que la supresión de la expectativa deriva en una mejor percepción, pero no aumenta la información en las respuestas neurales específicas del estímulo.

El quinto capítulo consiste en un debate general sobre nuestros resultados, limitaciones que abren vías futuras para este trabajo. En este capítulo he explicado cómo mis resultados contribuyen a la literatura existente sobre el procesamiento predictivo y sugiero incluir el «*timing*» como un factor esencial a la hora de investigar el modelo de procesamiento predictivo en el cerebro humano. También concluyo que las modalidades sensoriales gestionan de distinta forma las expectativas contextuales y la predictibilidad temporal. Esto sugiere que, a la hora de investigar el procesamiento predictivo en el cerebro humano, deberían considerarse las diferencias específicas de cada modalidad, ya que el mecanismo predictivo en funcionamiento en un área no debe generalizarse necesariamente también a otras áreas.

En el último capítulo indico los artículos científicos en publicaciones de prestigio, conferencias y charlas relacionadas con el presente trabajo.

Abstract:

Predictive processing has been proposed as a fundamental cognitive mechanism that accounts for how the brain interacts with the external environment through sensory modalities. The core idea behind the predictive processing theories is that the human brain develops a generative model (i.e., an internal representation) of the surrounding environment that is used to constantly generate top-down inferences about the external cause of the energy impacting our senses. This internal model is constantly updated through the comparison between predicted and actual inputs. This process is based on the bidirectional interaction between low-level and high-level processing regions that tend to be hierarchically segregated in the human cortex (such as the visual cortex). Higher order regions develop predictions about the external input, while lower order regions bring information about the external input. The comparison between these two sources of information (top-down from higher order regions and bottom-up from lower- order regions) is used to determine the identity of the external stimulations. The difference between both the actual input and the prediction is termed as *prediction error*, a signals that is used to update the internal generative model of the environment around us

In the last two decades, a lot of research focus has been to understand how the human brain generates expectation about the incoming sensory responses and how it deals with surprise or unpredictable input. It is evident in predictive processing literature that the human brain suppresses the neural responses to predictable/expected stimuli (termed as expectation suppression effect), and enhances the neural responses for unpredictable/unexpected stimuli.

A lot of research has focused on investigating how the human brain processes the sensory stimuli to understand the context/identity ("*what* is it?") of the input. This could be either predictable or unpredictable depending upon the prior information about the stimuli. Importantly, in real life situations objects in the surrounding environment are "temporally dynamic". While certain stimuli are temporally regular and hence predictable, others could be highly unpredictable in time. We define this property as temporal predictability (i.e., "*when* is it happening?"). The brain processes the external information about the content "*what*" and timing "*when*" of environmental stimuli to understand the world around it. In this thesis, I used Magnetoencephalography (MEG) to understand how *what* and *when* predictable properties of a stimulus affect the expectation suppression across sensory modalities (i.e., visual and auditory). MEG allows us to record the human brain activity with a high temporal resolution and a good spatial resolution, making it an ideal modality to investigate human brain activity which is spatio-temporal in nature.

Several experimental designs such as mismatch negativity, match-mismatch paradigms, repetition and expectation suppression paradigms have been used to understand the visual and auditory perception. Few of the shortcomings of these experiments were a) they used reduced contextual information with one cue and one target b) the processing of the predicted/expected stimuli were affected by the task in which participants were involved, thus mixing prediction and attention c) predictable stimuli (or *predictions*) were compared to the unpredictable/violation stimuli (*prediction errors*), but these conditions have been associated with largely different neurocognitive processes.

Considering these shortcomings, I designed an experimental paradigm which presents a series of cues that could either predict the properties/features of the target or not. A temporal

manipulation was also introduced in the experiment, which could either predict the timing of incoming stimulus (i.e., predictable timing) or not (random/jittered timing). To deal with much debated relationship between attention (induced by task involved in the experiment) and expectations, an orthogonal relationship was maintained between the participant's task and the predictable feature, i.e., the participants report one feature of stimuli and the prediction is related to another feature. Also, the comparison was designed between *predictions* vs *no predictions,* such that no violation was involved.

Overall, I observed that visual and auditory domains process the sensory signals differently. Visual domain is more sensitive to the *what* aspect of the sensory signals than *when.* The visual domain shows sensitivity to temporal uncertainty (*when*) only when the content (*what*) is predictable. If there is no predictability in terms of content, temporal uncertainty does not exert any influence. The auditory domain, on the other hand, is equally sensitive to *what* and *when* aspects of sensory processing. In contrast to visual domain, auditory domain shows sensitivity to temporal uncertainty even in absence of the predictability of *what.*

Beyond the interaction between *when* and *what*, I observed that the auditory domain shows an overall stronger expectation suppression effect to the upcoming predictable stimulus compared to the visual domain. On a separate note, the task (orthogonal to the predictability manipulation) completely washed out the predictability effects in the visual domain, while it did not in the auditory domain.

This thesis mainly comprises six chapters. The first chapter provides an introduction to the predictive processing framework, addressing the notions of temporal uncertainty and expectation suppression. This chapter includes the goals and hypothesis proposed in this thesis.

Basic overview of neuroanatomy of visual and auditory systems has also been included to better situate the readers on the topic.

The second chapter is a general introduction to the methods I have used in this thesis. In this chapter I have provided details about the magnetoencephalography (MEG) imaging modality, its instrumentation and different data preprocessing and analysis techniques used in this thesis. This chapter is more theoretical and explains the general principles and mathematics behind all the techniques.

The third chapter contains the details about the first experiment focused on expectation suppression in the visual domain. This includes detailed description of the experimental design, details about the methods used for data analysis, results and discussion. In this chapter, I have provided evidence in support that the visual domain is more sensitive to the content-based expectations (*what*). Also, the information in stimulus-specific neural responses (traced via decoding models) increases as the expectations about the stimulus are built.

The fourth chapter contains the details about the second experiment focused on expectation suppression in the auditory domain. This follows the same structure of the previous chapter providing details about experimental design, methods used for data analysis, results and discussion. In this chapter, I have reported that auditory domain is equally sensitive to both content-based (*what*) as well as temporal (*when*) expectations. I also provide evidence that expectation suppression leads to a better perception, but does not increase the information in stimulus-specific neural responses.

The fifth chapter consists of a general discussion about our findings, limitations that provide future direction to this work. In this chapter, I have explained how my results contribute to the existing literature on predictive processing and suggest to include '*timing*' as an essential factor while investigating predictive processing model in human brain. I also conclude that the sensory modalities deal differently with the contextual expectations and temporal predictability. This suggests that while investigating predictive processing in the human brain, the modality specific differences should be considered, since the predictive mechanism at work in one domain should not necessarily be generalised to other domains as well.

In the last chapter I report the scientific publications in prestigious journals and conferences along with the invited talks for this work.

CONTENTS:

Chapter 1: Introduction

1.1 Predictive Processing, Temporal Uncertainty and Expectation Suppression

शरीरेन्द्रिय्सत्त्वात्म्संयोगो धारि जीवितम् ।

नित्यगश्च्चानुबन्धश्च्च पर्यायौरायुरुच्यते

- *The Body combined with sense organs, mind, and soul, becomes life.*

The Sanskrit sloka mentioned above is taken from Charak Samhita, chapter 1 sloka 42, which is believed to have arisen around 400-200 BC. This underscores the importance of *'sense organs'* and '*mind'* in life. Since then, research on sensory systems and the brain/mind has been a key attraction. In 335 BC, the Greek philosopher Aristotle thought the brain was simply a radiator that kept the all-important heart from overheating. Around 170 BC, Roman physician Galen suggested that four fluid-filled cavities (ventricles) of the brain were the seat of bodily functions, complex thought, and determined personality. This was one of the foremost suggestions that the brain was where our personality, memory, and thinking reside. There were several hints about the brain which kept on coming into the picture for centuries. Still, the first detailed map of the nervous system was created in the 16th century by Belgian anatomist Andreas Vesalius. He argued against the ventricles as the site of brain functions. This is one of the research outputs we mostly accept in modern neuroscience. In 1791, Italian Luigi Galvani showed that electricity applied to nerves could make muscles contract. This was the first suggestion that electrical impulses were important in the nervous system. In the 1860-70s, physicians Paul Broca and Carl Wernicke showed that the brain had dedicated parts to process different components of speech. By that time, microscopes were already developed, and researchers started looking deeper into the brain by taking advantage of new staining methods.

Spanish neuroanatomist Santiago Ramón y Cajal and Italian scientist Camillo Golgi were awarded the Nobel Prize in 1906 for identifying that nerve cells (neurons) are the building blocks of the brain and showing there are many different types of such nerve cells in the brain.

Meanwhile, brain waves were first described in humans by Hans Berger at the end of the 1920s. Their dependence on behavioral state (e.g., wakefulness, sleep), sensory and cognitive processing has since been further explored by a large body of research. Charles Sherrington and Edgar Adrian won the Nobel Prize in 1932 for proposing the concept of synapses (junctions between neurons). Alan Hodgkin, Andrew Huxley, and John Eccles won a Nobel Prize in 1963 by showing how neurons communicate via electrical and chemical signaling. Since 1963, neuroscience research exploded with the rapid advancement of technology and collaborations between several domains like physics, genetics, medicine, engineering, etc. All these factors lead to several advances in the field of neuroscience. The last few decades are also called "decades of the brain." Many collaborative initiatives between different labs have emerged and led to facing multiple uncertainties about brain functioning. The quest to understand the human brain has evolved, and several new theories of brain functioning have been proposed in the last few years. It is out of the focus of this thesis to explain all the theories about brain processing. I will focus on one theory that has attracted wide acceptance in the neuroscience community, i.e., ***Predictive Processing***.

Predictive Processing

Predictive processing has been proposed as a fundamental cognitive mechanism that accounts for how the brain interacts with the external environment through the different sensory modalities (Clark, 2013; Friston, 2005; Mumford, 1992). Predictive processing was initially proposed as Linear Predictive Coding (LPC) (Makhoul, 1975), which was developed for signal processing in communications technology; by that time, this was not intended to be

implemented as a model of brain function. From an engineering perspective, LPC assumes that reconstruction of a signal is informative and is worth preserving, while the residual error generated from the prediction and actual signal is uninformative and can be discarded. The complementary explanation of this concept can be: the predictable component of a signal can be removed to reduce the signal amplitude and allow efficient transmission. This predictive coding model was employed to illustrate the retina's function (Srinivasan et al., 1982). Srinivasan and colleagues proposed that the predicted local intensity value can be calculated from intensity values measured at nearby locations and those measured at preceding times. Later, Rao and Ballard (1999) proposed a predictive coding model applied to cortical functioning, which assumed that there is a hierarchy in cortical processing (similar to (Mumford, 1992)) and hypothesized that cortical feedback connections act to suppress the information which is predicted by higher-order brain areas (Rao and Ballard, 1999). This was one of the most influential models that inspired many researchers to pursue further studies on predictive coding. However, Rao and Ballard's algorithm required neurons to perform both positive and negative firing rates, which is biologically not possible, so this emerged as one of the shortcomings of this algorithm. Although the algorithm was re-implemented with non-negative firing rates (Ballard and Jehee, 2012), still the model was considered quite complex in explaining cell biology. Another variant of Predictive coding (i.e., the PC/BC-DIM, which reformulated the Rao and Ballard model using Biased computation theories of cortical function), was implemented using the Divisive Input Modulation (DIM) method (Spratling et al., 2009): this theory computes the residual error by *division* rather than *subtraction*. Nevertheless, also for this case how cortical cells implement the division process remains unclear. To address these issues, Friston proposed the principle of Free Energy Minimization (Friston, 2005), which was similar to Rao and Ballard's algorithm in terms of hierarchy, but unlike any previous theory, the variable in the free energy principle did not represent the

absolute values of signal, but they represent the statistics of these signals. The free energy model reconstructs the probability distribution from which the samples are believed to come, and it estimates a posterior probability density. A particular, simplified version of the free energy minimization scheme has been identified as "Predictive Coding" (Friston, 2009). The core idea behind all the predictive coding theories (Spratling, 2017) is that the brain develops a generative model (i.e., an internal representation) of the surrounding environment that is used to constantly generate top-down inferences about the external cause of the energy impacting our senses. The model is updated continuously through the comparison between predicted and actual inputs. This process is based on the bidirectional interaction between low-level and high-level processing brain regions that tend to be hierarchically organized in the human cortex (such as the visual cortex). Higher-order brain regions develop predictions about the external input, while lower-order brain regions bring information about the external input. The comparison between two sources of information (top-down from higher-order brain regions and bottom-up from lower-order brain regions) is used to determine the identity of the external stimulations. The difference between the actual input and the prediction is called *prediction error*, a signal used to update the internal generative model of the environment around us (Keller and Mrsic-Flogel, 2018). Two main concepts deserve attention in this domain, the concept of "*prediction*," as a source of internal expectation about the external stimuli, and the concept of "*prediction error*" that reflects the discrepancy between the expectation and the actual stimulus (Friston, 2005; Rao and Ballard, 1999). Prediction errors are valuable sources of information since they update the internal generative model, and consequently, they support learning.

Predictive models contrast with more classical feedforward models, which propose only a bottom-up flow of information causing perceptual experience. In a non-predictive account, sensory features are extracted through a series of spatiotemporal filters along the ascending

cortical hierarchy (DiCarlo et al., 2012; Hubel and Wiesel, 1968; Riesenhuber and Poggio, 1999). Recently, however, predictive models have received a lot of attention. In the last two decades, predictive processing has been involved in many neurocognitive domains, such as visual (Arnal and Giraud, 2012; Stefanics et al., 2014) and auditory perception (Heilbron and Chait, 2018), language (Gina R. Kuperberg and Jaeger, 2017; Hakonen et al., 2017) and music (Koelsch et al., 2019), both in humans as well as in animals (Richter et al., 2017).

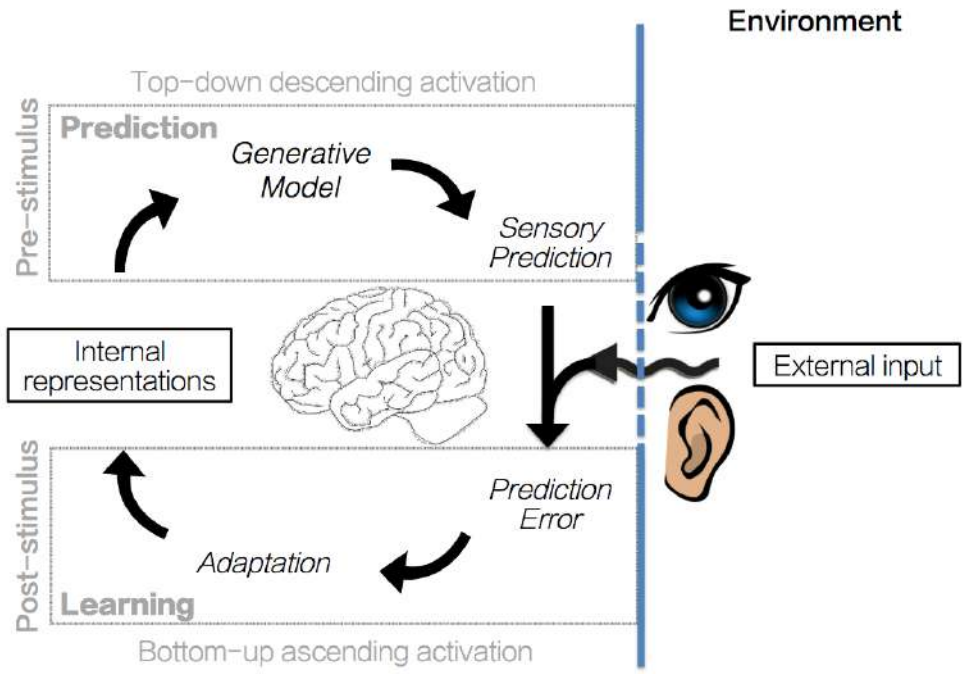


Figure 1: A schematic illustration of Predictive Processing framework. The external sensory input reaching the human brain is predicted before it appears (in pre-stimulus time period from top-down regions) and then compared to the actual input. The difference between actual input and prediction (called *prediction error*) is used to update the internal generative model which again generates prediction about the world around it.

Experimental approaches to Predictive Processing

Predictive Processing models became initially popular in the research domain of visual perception (Kimura et al., 2009; Mumford and Lee, 2003; Rao and Ballard, 1999; Stefanics et al., 2011). However, more recently research on auditory processing became a popular bed for

testing these models (Auksztulewicz and Friston, 2016; Dürschmid et al., 2016; Garrido et al., 2009; Näätänen et al., 2007; Pakarinen et al., 2007; Recasens et al., 2015; Ulanovsky et al., 2003). Several experimental paradigms have been developed in recent years to unfold how the brain processes the sensory input (both visual and auditory) and the role of top-down predictions. These paradigms were based on ground-breaking research questions about sensory processing. For example: How does brain react when something appears suddenly, surprisingly or unexpectedly? To address this type of questions researchers mainly focused on the so-called odd-ball paradigm (Costa-Faidella et al., 2011; Egner et al., 2010; Garrido et al., 2009; Grill-Spector et al., 2006; Kimura et al., 2009; Kok et al., 2012a). However other experimental paradigms have been used. Overall, I grouped those experimental paradigms in the following categories:

1. Surprise-based paradigms (e.g., odd-ball experiments): Sensory input coming from the external environment may be predictable or unpredictable in nature. Surprise based paradigms are used to investigate the effect of an unexpected/unpredictable stimulus while the brain was predicting/expecting something else (Egner et al., 2010; Garrido et al., 2009; Kimura et al., 2009; Stefanics et al., 2014). There is an abundance of literature showing a classical mismatch negativity effect emerging in the odd-ball experiments. In the predictive processing framework, these paradigms mainly focus on the *prediction error*

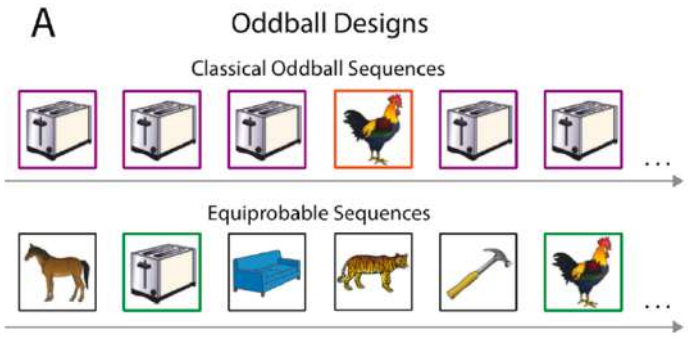


Figure 2A: Classical oddball paradigm with sequence of a commonly-appearing standard stimulus is interspersed with a rarely-appearing deviant. In equiprobable conditions, the same stimuli are interspersed with a large number of different images, which each appear with same probability. Figure taken from (Feuerriegel et al., 2021)

2. Repetition and roving paradigms: These paradigms mainly focus on brain responses to expectation generated by repeating the stimuli and altering this expectation by changing the stimuli (Grill-Spector et al., 2006; Recasens et al., 2015; Utzerath et al., 2017). In roving paradigms (Costa-Faidella et al., 2011; Stefanics et al., 2018), a deviant tone (odd-ball/ unexpected stimuli) appears in a sequence of standard tones (tones) and then the deviant tones become standard.

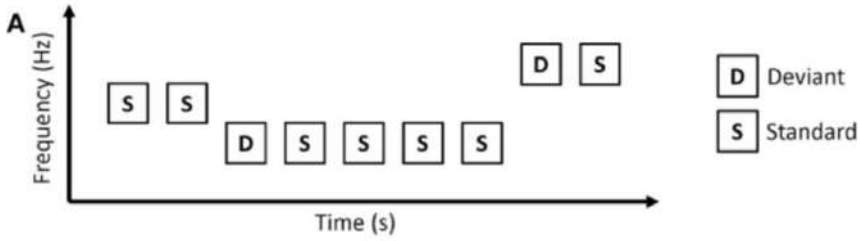


Figure 2B: Roving Paradigm in which probability of stimuli depends on the length of repetitive standard stimuli. Figure taken from (Kirihara et al., 2020)

3. Probabilistic cueing based experiments: An expectation about a stimuli can also be generated by manipulating the probability of occurrence in a block (Han et al., 2019; Kok et al., 2012b). For example, a diamond cue in the figure may be followed by a lion in 75 % of trials in a block (making it an expected stimulus), whereas a horse might appear in 25 % of trials (making it an unexpected stimulus). This can also be counterbalanced through another cue where two stimuli will appear with equal probability (50 % probability each).

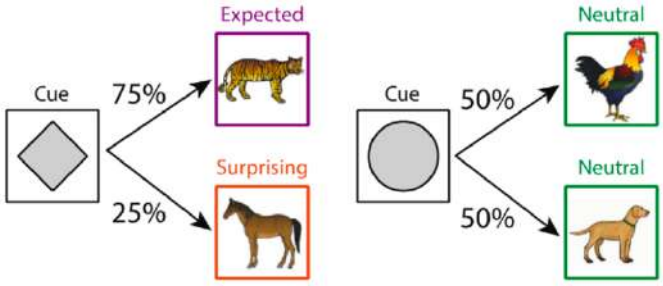


Figure 2C: In probabilistic cueing paradigm, a cue stimulus provides information that a specific stimulus is more likely to appear (diamond cue), or two or more stimuli are equally likely to appear (circle cue). Figure taken from (Feuerriegel et al., 2021).

4. Motion-based paradigms: These paradigms (Blom et al., 2020) involve motion of a specific item in a predictable fashion (for example, position changing linearly or randomly). The expectation can be generated either for a specific spatial location (position) or for a time point when the stimulus is expected.

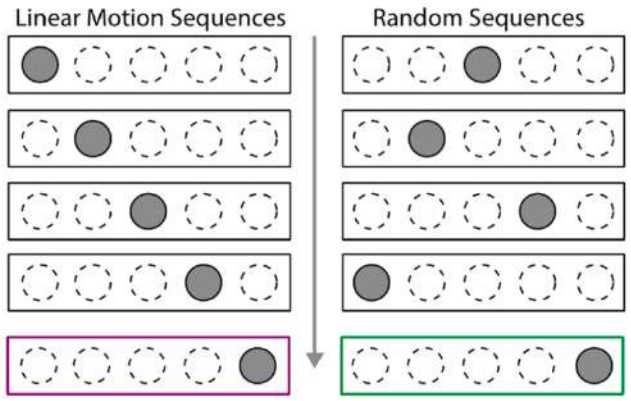


Figure 2D: In motion-based paradigms, the appearance of a stimulus at a particular location is either expected (in a linear motion sequences) or preceded by a randomly-generated sequence of stimulus locations (in random sequences). Figure is taken from (Feuerriegel et al., 2021).

All these match/mismatch paradigms either compare the expectation effect with the violation of that expectation (i.e., *prediction* vs. *prediction errors*) or compare the violation of an expectation with a neutral condition (i.e., *prediction error* vs. *no prediction errors*). In the present study we deviate from this trend, by designing an experimental paradigm that focuses on the *prediction* vs. *no prediction* contrast, in which no violation of expectations is involved. In our opinion, the violation of an expectation involves cognitive processes behind the pure

neurocognitive mechanisms involved in perception, such as changes in attention, arousal, or learning mechanisms.

Importantly, prediction effects are often confounded with attention-dependent cognitive changes triggered by an experimental task, in which participants could be made explicitly aware of the predictability manipulation. The neurocognitive mechanisms involved in attention and prediction are highly overlapping but potentially dissociable using neuroimaging methods (Summerfield and Egner, 2009). Both prediction and attention are top-down processes that affect human perception. However, while predictive processing has been proposed as an intrinsic mode of brain functioning that explains neural activity in general (Friston, 2005), attention is viewed as a more controlled mechanism, where neural resources are consciously invested to focus/unfocus on a stimulus. Prediction and attention are related and not always clearly distinguishable (All and James, 2015; Kok et al., 2012b). Neurophysiological evidence indicates that neural activity is enhanced for an attended stimulus compared to an unattended one. If attention and prediction would trigger similar neural responses, then we should observe an increase in neural activity for predicted compared to unpredicted stimuli. However, this is not the case since predicted stimuli suppress neural activity compared to the unpredicted ones, a phenomenon termed as *Expectation Suppression*. Summerfield & Egner (2009) underscore the importance of investigating attention and prediction in an orthogonal fashion in order to isolate the role played by prediction on perceptual processing. As indicated by these researchers, the proper experimental paradigm should dissociate these two cognitive components: participants should pay attention to the processing of one sensory feature of a stimulus, while the prediction/expectation manipulation should rely orthogonally on another feature.

The role of Time in Predictive Processing

In the predictive framework, the input stimuli received by the lower order brain areas (i.e., primary sensory cortices) from the external sensory periphery (like the eyes or the ears) are decomposed into several features. For example, a visual stimulus may have features like edges, orientations, color, shape, and other properties, and an auditory input might have features like pitch, volume, frequency, etc. These features (that constitute the identity of a stimulation) are the building bricks of the overall percept and determine its identity (often called "*what*" features). Importantly, in real-life situations, objects in the surrounding environment are "temporally dynamic." While playing tennis, you may see a ball coming toward you on the tennis court: but even after having all the "*what"* information features (like color, shape), for making a successful shot, it is important to estimate "*when*" the ball will approach you and impact your tennis racket. While certain stimuli are temporally regular and hence predictable, others could be highly unpredictable in time. We will define this property as temporal predictability.

From the perspective of visual perception (as compared to auditory perception, see below), temporal predictability (*when* information) has raised relatively low interest compared to the study of stimulus content predictability (*what* information). The role of timing for predictive processing in the visual domain has been mainly studied for the perception of moving objects. The visual system employs a certain amount of time to process the incoming sensory information, a delay defined as "neural transmission delay" (Blom et al., 2020; Maunsell and Gibson, 1992). When considering the perception of moving objects, this delay could cause potential problems for the visual system that should anticipate some information to provide valuable cues for the cognitive system to plan action. Predictive processing models explain how our brain deals with this delay through a compensatory mechanism that requires the

generation of predictions about the incoming stimuli. Studies from both animals and humans have shown that predictable visual stimuli are represented with shorter latencies in the visual cortex compared to unpredictable stimuli (Hogendoorn and Burkitt, 2018; Jancke et al., 2004; Subramaniyan et al., 2018). Moreover, studies have shown that stimulus-specific information induces sensory templates from prior expectations in both visual and auditory domains which can be decoded from neural signals even in pre-stimulus time intervals (Demarchi et al., 2019; Kok, P., Failing, M. F., & de Lange, 2014; Kok et al., 2017).

These studies thus underscore that these *what* and *when* features are strongly interwoven and shape the human perception. Thus, understanding the interaction of *what* and *when* in human perception is crucial for understanding human cognition. Studies on visual perception have paid relatively less attention to temporal (*when*) than to content (*what*) predictability (Demarchi et al., 2019; Kok et al., 2017; Nobre et al., 2007). The effect of temporal jittering has been explored intensively in visual attention studies that focused on the estimation of the spatial location of a stimulus (Coull and Nobre, 1998). Coull and Nobre (1998) used PET and fMRI to manipulate the cued attention to spatial locations based on their temporal intervals in factorial design. An hemispheric asymmetric effect was reported in the parietal cortex, with the right parietal cortex activations related to spatial attention whereas the left parietal cortex activations mainly linked to temporal attention.

Temporal expectations in literature are reported to have interaction with other cognitive processes like spatial attention or stimulus features. While only temporal expectations are modulated (in lack of expectation about the content), no effects on neural responses were reported (Correa and Nobre, 2008; Doherty et al., 2005; Rohenkohl et al., 2012; Rohenkohl and Nobre, 2011).

Aims of the present thesis

This thesis aims to answer some of the critical research questions which still lack experimental evidence. Firstly, we aim to investigate "*how expectation suppression is affected by temporal uncertainty?*" Expectation suppression is defined as a reduction in a specific measure of neural activity (i.e., EEG/MEG signal amplitudes, Local field potential (LFP) or bandlimited LFP amplitudes, single-cell or multicell unit firing rates or even fMRI BOLD signals) following the presentation of a stimuli that is expected/predicted compared to a neutral condition (i.e., when the stimuli is presented but its neither expected nor surprising) (Arnal and Giraud, 2012; Egner et al., 2010). The term "expectation suppression" derives from the repetition suppression effect which describes the reduction in neural activity for repeated stimuli (Desimone, 1996). Importantly, the notion of *suppression* has been used in different predictive coding frameworks (Egner et al., 2010; Summerfield and De Lange, 2014) to imply that activity in stimulus selective neurons is reduced when the perceptual expectations of an observer is fulfilled (Feuerriegel et al., 2021). Expectations can be related to stimulus location, stimulus content and stimulus timing (Auksztulewicz et al., 2018). It is worth noting that different sensory modalities (i.e., visual and auditory) are classically assumed to be similarly sensitive to expectation suppression. Since this cannot be taken for granted, here we ask: Do both modalities present similar suppression effects? Or do they show different neural patterns? And how are the neural patterns affected by temporal uncertainty?

Secondly, "*how do the visual and the auditory modality show differential sensitivity to task-relevant and task-irrelevant stimulus features?*" Do these modalities process the attended and the predicted features of a stimulus in a similar or different way. This question will provide critical information about how the processing of a specific sensory property is inherently interwoven with the processing of a separate feature and how this changes across modalities.

Thirdly, "*does any of these modality show prestimulus activation when a stimulus is predictable?*" In other words, will we find evidence of the neural correlates of *prediction* (as opposed to *prediction error*)? This issue has been largely investigated in the neurocognitive literature without providing any conclusive evidence about the reliability of pre-stimulus prediction neural correlates.

In the present work I used magnetoencephalography (MEG), which has excellent temporal resolution and good spatial resolution to measure the brain correlates of both *what* and *when* aspects of stimulus processing. I designed two experiments, one focused on visual perception (chapter 3) and another on auditory perception (chapter 4). In the visual experiment, the participants were presented with a series of Gabor patches (no match - mismatch) having two experimentally manipulated features (*orientation angle* and *cycles per degree*: cpd). Based on the first four Gabors, the orientation angle of the target could be predictable or not, whereas the participants were asked to evaluate the other feature (i.e., cpd), which was intermediate during the first four Gabor patches and was higher or lower at the target (orthogonal relationship between prediction feature and task-related feature). In the auditory experiment, four tones with two experimentally manipulated features (pitch and tone length) were presented. These four tones could either predict or not the pitch of the target, whereas the participants were asked to report the length of the target, which was either longer or shorter compared to the previous tones (orthogonal relationship between pitch and length of tone).
In this thesis, I present results by measuring evoked responses (ERF) both at the sensor as well as the source space. This is followed by machine learning-based multivariate pattern analysis techniques like time-resolved decoding (Carlson et al., 2019; Cichy et al., 2014).

1.2 Neuroanatomy of Visual and Auditory system

Neuroanatomy of Visual System

Vision depends on the light reflected from the objects passing through the cornea and the lens. The light combines to produce a clear image of the objects around us on a sheet of photoreceptors called "retina." Similar to a camera, which projects a flipped image, i.e., objects above the center are projected to the lower part, and objects below the center are projected to the upper part, images on the retina are also reversed. The retina transmits information via the optic nerve in the form of electrical signals to different parts of the brain (including the occipital cortex), which process the image and allow us to see the visual objects around us.

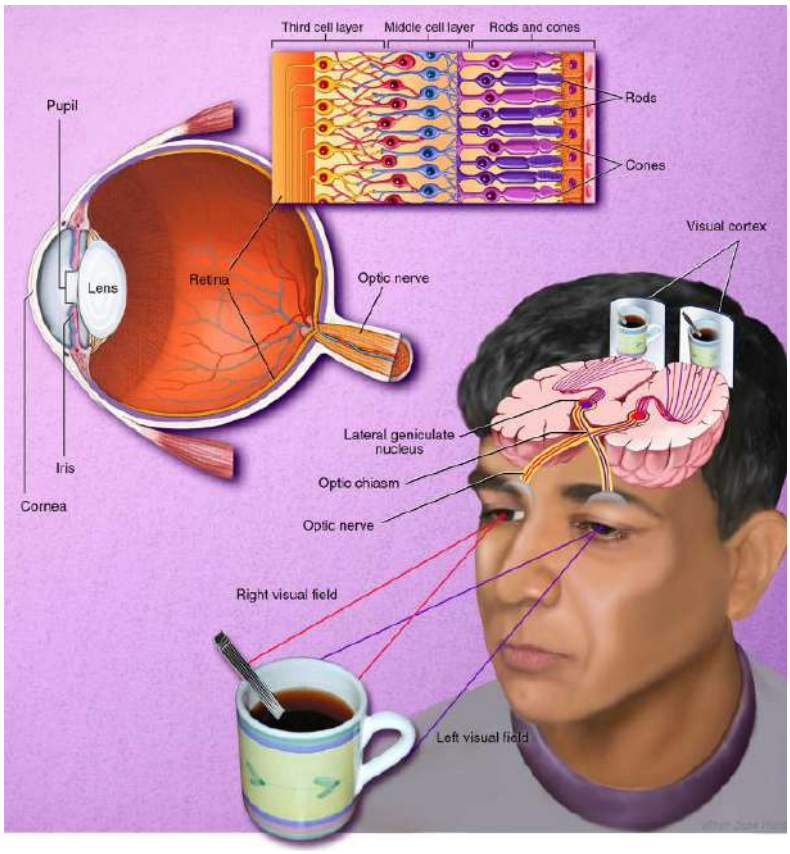


Figure 3: Visual system showing the beginning of vision process through cornea and lens, to retina and then to LGN via optic nerve. Figure taken from BrainFacts.org

Thus, the visual system processing begins by comparing the amount of light impacting on a small part of the retina with the amount of surrounding light. Visual information from the retina is then transmitted through the lateral geniculate nucleus (LGN) of the thalamus to the primary visual cortex (often called V1) located in the backside of the human brain (occipital lobe). Although the visual processing mechanisms are not still completely understood, it has been argued that visual processing takes place in two cortical pathways called the *dorsal* pathway and *ventral* pathway. The dorsal pathway, which stretches from the occipital lobe to the parietal lobe, also shares some of the areas with the ventral pathway. The dorsal pathway has been mainly reported to play a role in visually guided behavior; it is very sensitive to high temporal frequencies, motion, maintains low consciousness level, and process information faster compared to the other pathway. The dorsal pathway mainly deals with objects' locations and motions, so it's often called the "where" stream. The ventral pathway, which is associated with object recognition and form representation, is also called the "what" stream. This pathway is connected to the medial temporal lobe (related to memory storage), the limbic system (related to emotions), and eventually with the dorsal pathway (related to object locations). Neurons in LGN project to V1 sublayers 4Cβ, 4A, 3B, and 2/3a. From there, the information goes to V2 and V4 to the areas of the inferior temporal lobe (i.e., PIT: posterior inferotemporal, CIT: central inferotemporal, and AIT: anterior inferotemporal). Moving along the mainstream from V1 to AIT, receptive fields increase their latency, size, and the complexity of their tuning. All these brain areas follow a processing hierarchy.

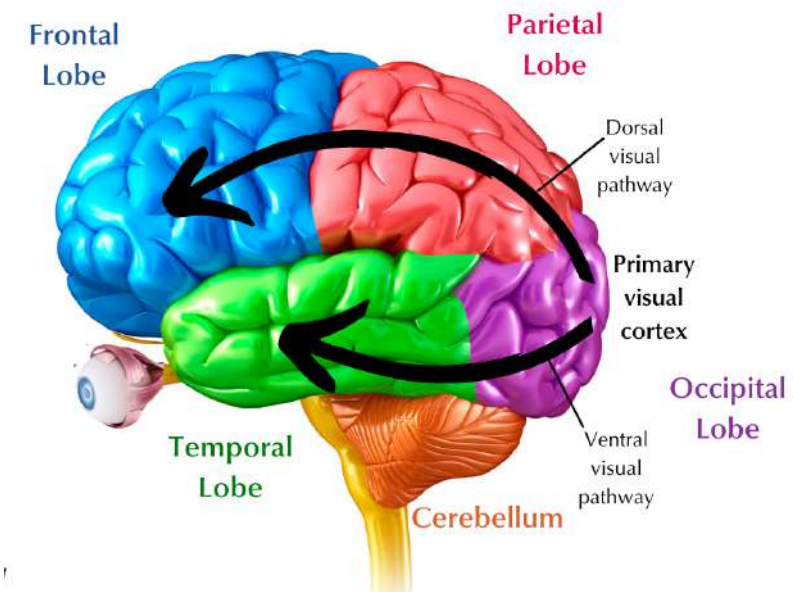


Figure 4: Dorsal and Ventral visual pathway. Image taken from (https://www.perkins.org/higher-order-visual-pathways-and-the-cvi-brain/).

Neuroanatomy of Auditory system:

The human ear is grouped into three parts: the outer ear, the middle ear, and the inner ear. The outer ear is the external portion of the ear, consisting of the pinna and the ear canal. This part gathers sound waves and guides them to the middle ear. The middle ear has three tiny bones (malleus, incus, and stapes), called the ossicles. These three bones form a connection from the eardrum to the inner ear. As sound waves hit the eardrum, the eardrum moves back and forth, causing the ossicles to move. As a result, the sound wave is converted to a mechanical vibration transferred to the cochlea. The cochlea is part of the inner ear with a watery fluid (called perilymph), which moves in response to the vibration. As the fluid moves, thousands of hair cells located on the basilar membrane in the cochlea sense the vibration and convert that movement to electrical signals communicated via neurotransmitters to thousands of nerve cells.

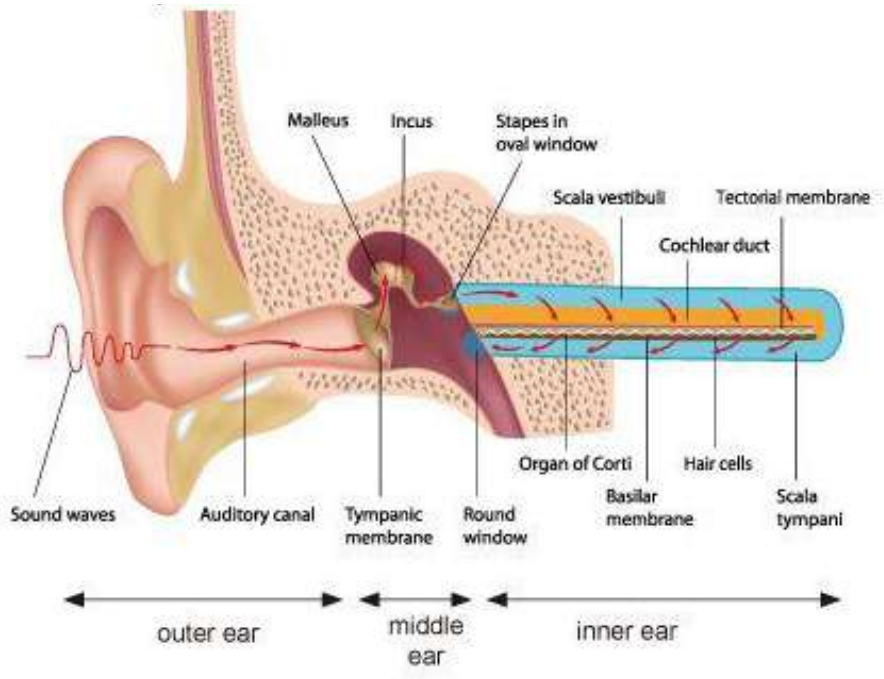


Figure 5: Different parts of Ear and cochlear tonotopy. Figure taken from https://es.yamaha.com/es/products/contents/proaudio/docs/audio_quality/04_audio_quality.html)

The hair cells are tuned to specific frequencies based on their location in the cochlea. In this way, lower frequencies cause movement in the base of the cochlea, and higher frequencies work at the apex. This characteristic is known as cochlear tonotopy. The cochlea is capable of interpreting sound in terms of frequencies (between 20 Hz and 20.000 Hz) and intensity (between 0 decibels (dB) sound pressure level (SPL) and 120 dB SPL). Nerve impulses generated in the inner ear travel along the cochlear nerve (acoustic nerve) and enter the brainstem at the lateral aspect of the lower pons. Auditory fibers from more basal areas of the cochlea reach dorsomedial parts, and neurons from more apical parts project to the ventrolateral parts of the medial geniculate nucleus in the human brain.

The human auditory cortex represents around 8% of the surface of the cortex. The auditory cortex (figure 6) is located along the superior temporal gyrus (STG). Human auditory cortex is hierarchically organized with a core or primary auditory cortex, surrounded by non-primary belt and parabelt regions.

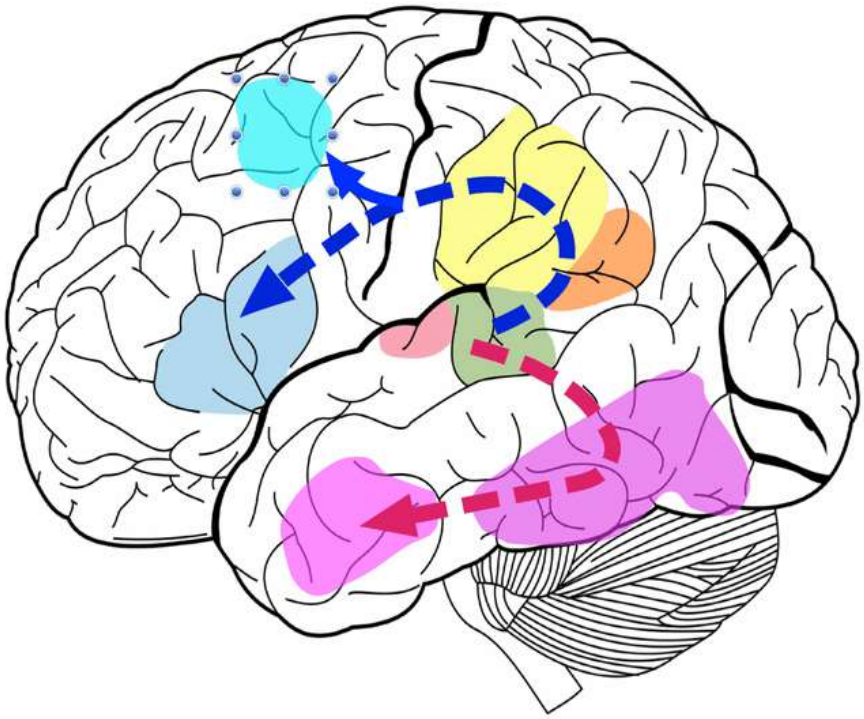


Figure 6: The Auditory cortex, reflecting Primary auditory cortex (Peach color adjacent to green color), Angular gyrus (orange), Supramarginal gyrus (yellow), Broca area (sky blue) and Wernicke area (green). Image taken from https://mriquestions.com/language.html

The primary auditory cortex lies in STG and extends into the lateral sulcus and the transverse temporal gyri (also called Heschl's gyri). Primary auditory cortex passes the auditory information to Wernicke area, where the auditory information is passed to two cortical pathways: *ventral* and *dorsal* (similar to visual domain). The *ventral* cortical pathway (shown by red dotted line in figure 6) carries semantic information, which is important for determining the meaning of language and travels from the rostral pole of the temporal lobe to the occipitotemporal cortex. The *dorsal* cortical pathway (shown by blue dotted line in figure 6) carries phonological information about sounds from the superior temporal cortex to the inferior frontal cortex, which aids the individual in understanding segments of speech, learning vocabulary, and understanding articulation of words.

Chapter 2: Methods

2.1 Magnetoencephalography

This section will mainly describe the neuroimaging modality (i.e., Magnetoencephalography, MEG) used in the present thesis. I will explain the MEG instrumentation, data preprocessing methods (maxfilter, artifact rejection & ICA), and data analysis methods (evoked analysis, source reconstruction, and MVPA) used in this thesis.

Brain waves were first described in humans by Hans Berger (Berger, 1929) at the end of the 1920s, and their dependence on the behavioral state (e.g., wakefulness, sleep) and sensory and cognitive processing has since been elucidated by a large body of research.

Many neuroimaging methods have been developed, and many of them are currently used in different cognitive and clinical neuroscience applications. These methods can be classified into broader categories including *invasive* methods such as Electrocorticography (ECoG), Low Field Potentials (LFPs), Single Cell recording, Positron Emission Tomography, stereo Electroencephalography (sEEG). All these invasive methods provide direct recording of brain activity but require a lot of surgical expertise, animal care, maintenance, risk of mortality, and recovery time after surgery. Some methods like PET even need a radiotracer that has harmful effects on the body in a few cases.

The opposite category of such methods is defined *non-invasive* and include scalp Electroencephalography (EEG), which records the electrical activity of the brain; Magnetoencephalography (MEG), which measures the magnetic field generated from neuronal communication in the brain; Magnetic resonance Imaging (both functional and structural) whose functional component measures the change in oxygenated and deoxygenated blood (called BOLD response); Computer tomography Scanning (CT-SCAN) which measures the anatomy in the brain using x-rays and generates a tomographic image.

All these methods can be further classified in terms of their resolution, either spatial resolution or temporal resolution. *Spatial* resolution is the capability of any imaging modality in discriminating between brain activity between two brain areas. The lower the spatial resolution of a modality, the better that modality can differentiate between activity from close regions. For example, if a modality has a spatial resolution of 5 mm, that means it can differentiate the brain activity generated from brain areas that are at least 5 mm apart. Similarly, *temporal* resolution is the capability of a modality to discriminate between brain activity generated at two different time points. For example, if there is any modality with a temporal resolution of 5 milliseconds, that means that such modality can differentiate the brain activity that emerged at least five milliseconds apart. PET and MRI are considered to have an excellent spatial resolution (~ 1 mm), but they have a very low temporal resolution as the oxygenated blood takes few seconds to reach brain areas that are active during a task. However, EEG is considered to have an excellent temporal resolution (~ 1 millisecond) but has a poor spatial resolution (~2 centimeters or more). Of all these neuroimaging modalities, MEG provides the best compromise over the other methods as it is non-invasive, has an excellent temporal (~ 1 millisecond) and good spatial resolution (~ 5 mm). Moreover, MEG provides a direct estimate of the measured neural activity. On the contrary, other methods like fMRI and PET provide an indirect estimate of neural activity derived from neurovascular coupling.

In short, Magnetoencephalography is a non-invasive brain imaging technique for investigating human brain activity. It allows the measurement of ongoing brain activity on a millisecond-by-millisecond basis, and it unveils *where* brain activity is produced. MEG signals were first measured in 1968 by Dr. David Cohen at the University of Illinois. He used a copper induction coil as the detector in a magnetically shielded room, which reduced the magnetic background noise. However, the measured signal was still very noisy. Dr. Cohen later moved to

Massachusetts Institute of Technology, Massachusetts, USA, and built a better magnetic shielded room and measured the new superconducting quantum interference devices (SQUIDs) developed by James E Zimmerman. Using SQUIDs in a shielded room proved fruitful and is currently being used by most MEG manufacturers.

Figure 7: The figure shows the first MEG measurement in a shielded room by Dr. Cohen (left) and the modern MEG machines (center and right).

What do we measure with MEG:

The MEG and EEG are mainly used for measuring the temporal dynamics of brain activity. They are unrivaled in terms of the precision of their temporal resolution, i.e., milliseconds or lower. But what makes EEG and MEG so different? One instrument (EEG) can be purchased for a few thousand euros, whereas a MEG machine costs millions of euros. MEG machines yield very high maintenance costs since they require helium refills to cool down the sensor superconducting materials once in a while. The helium gas adds to the cost of MEG, and the natural concern is *if* this higher cost of equipment and maintenance is justified. Does MEG provide substantial brain-related evidence compared to EEG? The answer to such questions is "YES." EEG measures the change in electrical potential related to activity in the human brain,

which travels to the scalp. Source localization of this activity is mainly limited by the distortive effects of the intervening structures, such as dura, skull, and the different tissue types in the human brain, which severely hamper efforts to localize the signal source precisely. On the other side, MEG measures the magnetic field changes induced by intracellular current flow. The direction of this field obeys the 'right-hand rule' in Ampère's law applications. These magnetic fields that pass through the dura, skull, and scalp are relatively unaltered. This technique thus offers a non-invasive method to 'listen' to brain activity during rest and task conditions. This - from the subject's perspective - feels safe, painless, and quick to set up. This technique does not require any gel or impedance checking method as used in EEG, despite measuring at several hundred channels at a time. MEG data combined with mathematical modeling and advanced signal processing methods enable localization of sources while recording at several kilohertz sampling frequencies.

The magnetic field measured by MEG is relatively tiny (femtotesla unit) in strength. To get an idea of the order of magnitude of this signal, below I report a comparison showing different magnetic field strengths available around us. An MRI machine has a magnetic field of 1 Tesla (nowadays, this goes up to 3T, 7T for humans, and 9.4T for animals). If we reduce this by 1000 magnitude, we have one millitesla (10^{-3}), which is the strength of a refrigerator magnet. If we further reduce this by a magnitude of 1000, we have one microtesla (10^{-6}), which is the strength of a microwave oven at 1 foot. Reducing this magnitude by another 1,000,000, we have one picotesla, the magnetic field of human brain activity: it ranges from picotesla (10^{-12}) to femtotesla (10^{-15}), depending upon different parameters of the source activity.
Measuring such a small magnetic field requires highly sophisticated instrumentation, which I will explain in the following subsection.

2.2 MEG Instrumentation

The MEG system used in the present thesis has the following components: First, a gantry that carries the sensor (SQUIDS); second, a magnetically shielded room (Faraday cage); third, an operator console. I will explain these in detail.

1. MEG Sensors

Superconducting Quantum Interference Device (SQUIDs) are extremely sensitive to magnetic field changes at very low temperatures, to be measured and converted into digital signals (‘quantization’). The first SQUIDs used for brain application were read via inductive coupling to an external circuit operating at radio frequencies. These single-junction SQUIDs are called rf-SQUIDs. Later, the two-junction dc-SQUID became popular when the semiconductor industry grew up and outperformed the rf-SQUID in many aspects, since it had a lower noise level, simpler electronics, and minimal crosstalk between channels. All these factors are essential for the success of the contemporary MEG systems.

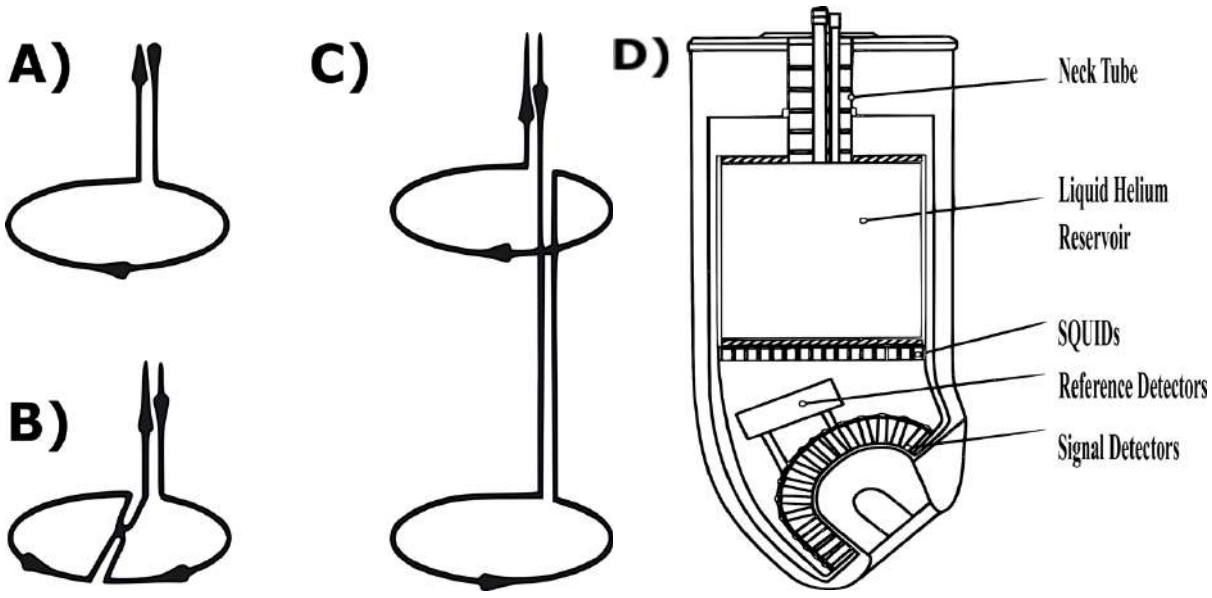


Figure 8: A) Magnetometer flux coil B) Planar Gradiometer coil C) Axial Gradiometer coil D) MEG dewar sketch design.

All modern MEG systems are equipped with dc-SQUIDs. SQUIDs are made relatively small to optimize their sensitivity, usually less than 1 mm in diameter.

Due to this small surface area, SQUIDs sensors have poor coupling to the magnetic field. In MEG applications, SQUIDs cannot be used in their native form. Flux transformers are used to enhance the coupling that "squeeze" more magnetic flux into the SQUID loop by collecting it from a much larger area. This allows measuring different components of the magnetic field without altering the SQUID geometry. Flux transformers are also made of superconducting material and comprise a pick-up coil close to the brain. Based on the pick-up coils, the sensors are classified into different types. One such type is a "magnetometer" (Figure 8 A). Magnetometers measure the magnetic field component along the direction perpendicular to the surface of the pick-up coil. These are very sensitive to nearby sources (neural currents in the brain) as well as susceptible to faraway sources. An additional compensation coil may be added to reduce sensitivity to distant sources, which measures the interfering signal primarily. This type of configuration is called a "gradiometer" (Figure 8 B, 8C).

Gradiometers are less sensitive (almost blind) to distant sources. The gradiometers can be classified as "axial gradiometers" (Figure 8B) or "planar gradiometers" (Figure 8C) based upon the arrangement of two coils of a gradiometer. The peak signal from an axial gradiometer comes from the sources around the rim of the sensor, whereas planar gradiometers signal peaks from the sources right beneath them. All these sensors and helium reservoirs are kept together and constitute the "Dewar." Figure 8 D shows an outline of a dewar. The MEG facility available at BCBL consists of 102 pairs of triplet sensors. Each sensor comprises one magnetometer and two planar gradiometers.

2. **Magnetically Shielded room**

The brain signal we aim to measure has a tiny magnetic field. These tiny magnetic fields can be contaminated by environmental magnetic fields (considered noise) such as power interference generated from electronic devices, personal computers (PC), and moving metallic carts. SQUIDs are very sensitive to such ambient noise. To reduce this noise, brain signals are measured inside a magnetically shielded room which suppresses most of the ambient noise present around the room. Below is an image of a magnetically shielded room installed at BCBL.

Figure 9: Magnetically shielded room and MEG facility installed at BCBL

3. **Operator Console**

This area is designated for all the hardware related to data acquisition. This mainly comprises MEG data acquisition computers, stimulus presentation computers and synchronization hardware. This area also includes a participant preparation section where the participant is briefed about the experiment, the head of the participant is digitized, and participants can change their clothes, etc.

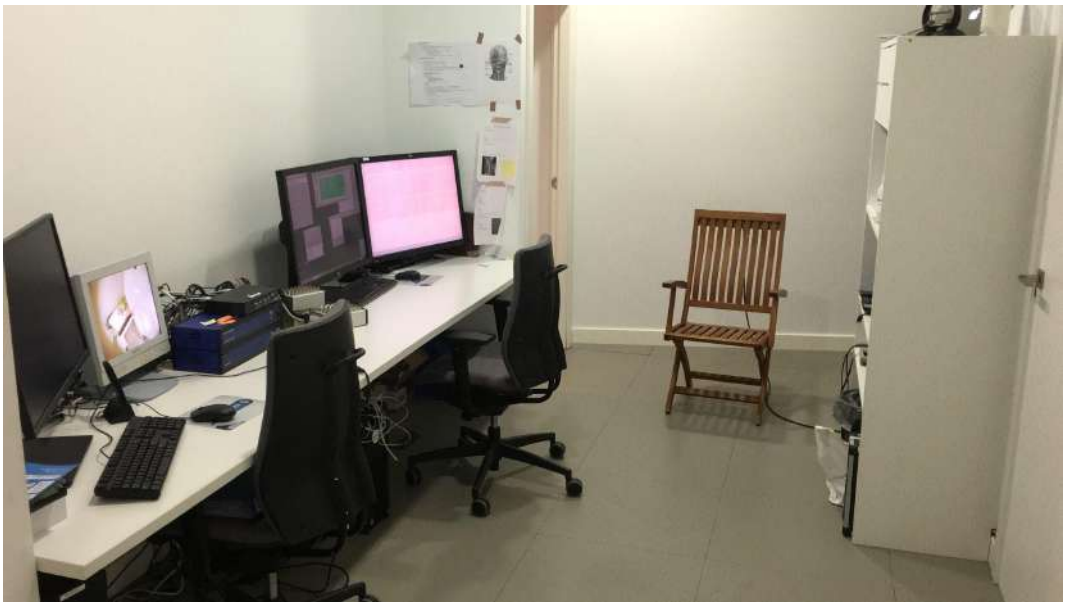


Figure 10: MEG operator console setup at BCBL

2.3 MEG data acquisition and processing

Participants for this thesis were selected from the web Participa database (https://www.bcbl.eu/participa/) maintained by the BCBL lab staff. This database consists of more than 3000 participants' data stored with their relative details (such as handedness, linguistic profile, education, etc). Participants having any metallic implant in their body, such as dental implants, cochlear implants, dental braces, metallic rods in bones, pacemaker, permanent nose pins, permanent earrings, etc., are excluded from participating in the MEG

experiments. Selected participants were free from any neurological disorder and had normal or corrected to normal vision through MEG compatible eyeglasses. The participants were also given clean cotton clothes to wear to avoid any trapped metallic components from the environment. For the data acquired in the post-COVID situation, the participants were also provided with MEG compatible masks and non-metallic hand sanitizer. The MEG area was sanitized and ventilated properly between successive participants.

After changing the clothes, the participants were asked to sit on a head-digitization chair. Five head positioning coils (HPI) were fixed around the head to detect any head movement during the experiment. The data from these HPI coils is used to correct the head motion of the participant using maxfilter 2.1 (proprietary software from MEGIN, Finland) offline. Three fiducial points (Nasion, right, and left auricular points) were also digitized along with ~ 300 points on the participant's head. All these digitization steps were performed to have an estimate of participant head shape. This head shape is later used to combine MEG data with forward models derived from MRI data and perform source localization (Magnetic Source Imaging).

Since the eye movements and activity from the heart have a very strong electric signal, this largely affects the MEG data. Two pairs of electrooculogram (EOG) electrodes were placed horizontally and vertically around the eyes to capture the eye movements (both eye blinks and saccades) during the experiment. Two lead electrocardiogram (ECG) electrodes were also placed on the participant's body to monitor heart-beat.

Although the MEG machines are designed with proper earthing safety measures to protect the participants from any shock during hardware failure, as an extra safety measure, two electrodes are put behind the participants' ears and connected with the MEG machine, which acts as a ground.

Artifacts correction in MEG data:

Artifacts in MEG data can be classified broadly into two types. First, the artifacts generated from the sensors and the ambient noise. Second, the artifacts generated from biological signals (EOG, ECG, and muscle artifacts). For reducing the sensor noise, Maxfilter software (version 2.2.12) is used. Maxfilter is specific for Elekta (now MEGIN) MEG machines only. The name comes from the famous Maxwell filtering method used as a spatial filtering technique. Maxfilter (i) suppresses magnetic interference coming from both outside and inside of the sensor arrays, (ii) helps in reducing measurement artifacts, (iii) transforms data between various head positions, and finally (iv) plays an important role in correcting head movements. Maxfilter uses the signal space separation (SSS) method (Taulu and Simola, 2006) that utilizes the basic properties of electromagnetic fields and harmonics functions to decompose the MEG signals into three main components:

a) B_{in}: The brain signals originating within the sensor array (in space S_{in})
b) B_{out}: External disturbances arising out the sensor array (in space S_{out})
c) N: noise and artifacts generated by sensors and sources of interference close to the sensor array (in space S_T)

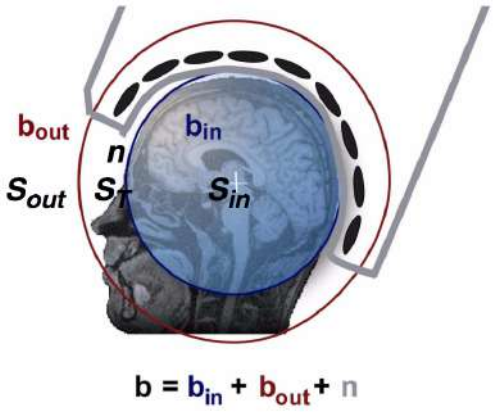


Figure 11 : The three different magnetic fields present during the data acquisition and used in the Maxfilter process. (Figure adapted from Elekta Maxfilter user manual)

The disturbing magnetic interference is suppressed by omitting the harmonic function components corresponding to high spatial frequencies, by neglecting the S_{out}-space component B_{out}, and by reducing the S_T-space component N. The temporal extension of the SSS method (tSSS (Taulu and Simola, 2006)) significantly widens the software shielding capability of MaxFilter, because tSSS can also suppress internal interference tS^ arises in the sensor space or very close to it.

Other artifacts which are generated from the participant's body are muscle artifacts, EOG, and ECG artifacts. Muscle artifacts arise mainly from the neck muscles, which are close to the MEG helmet. These artifacts have higher frequency and amplitude compared to the brain signal. These artifacts can be easily detected by z-transforming the whole signal and thresholding such z-transformed data. Below is an example of a muscle artifact detected using z-transformation in Fieldtrip software ("Automatic artifact rejection - FieldTrip toolbox,").

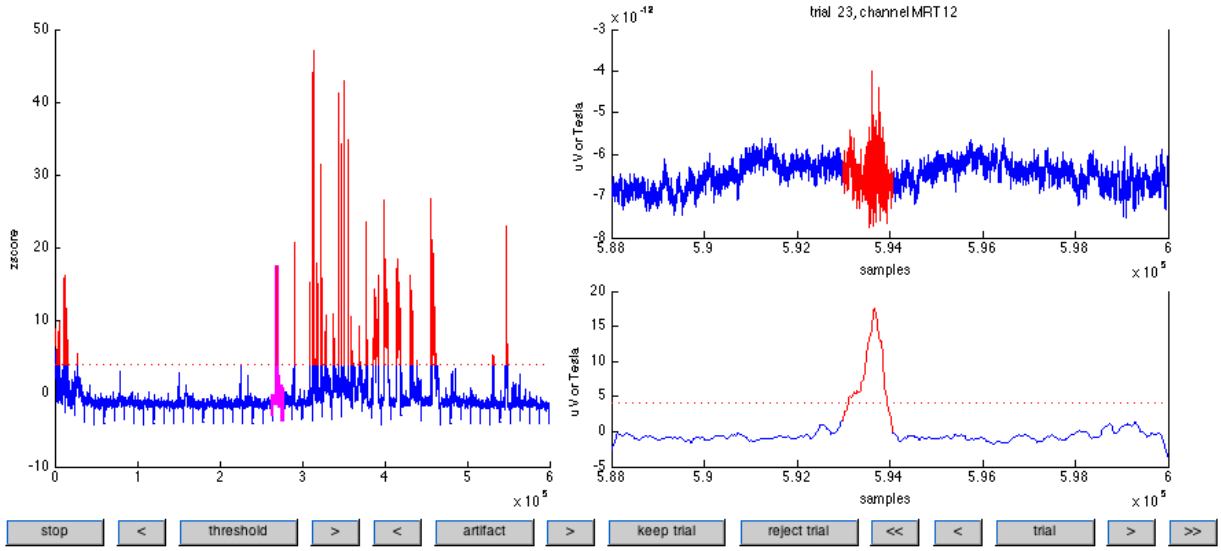


Figure 12: Muscle artifact detection based on z-transform (figure adapted from Fieldtrip toolbox website)

The same approach is also used for removing the jump artifacts which usually occur when a sensor or group of sensors have an amplification gain and the baseline of the sensor/sensors shifts.

Two important artifacts which are quite common are EOG and ECG artifacts. These artifacts are mainly removed using signal separation techniques like Independent Component Analysis: FastICA (Hyvärinen and Oja, 2000). First, the MEG data from all the channels is decomposed into independent components. ICA has two main assumptions. First, that the input signal is composed of statistically independent components; second, these independent components should be non-gaussian in nature. ECG and EOG components derived from MEG data used in this thesis are shown in figure 13.

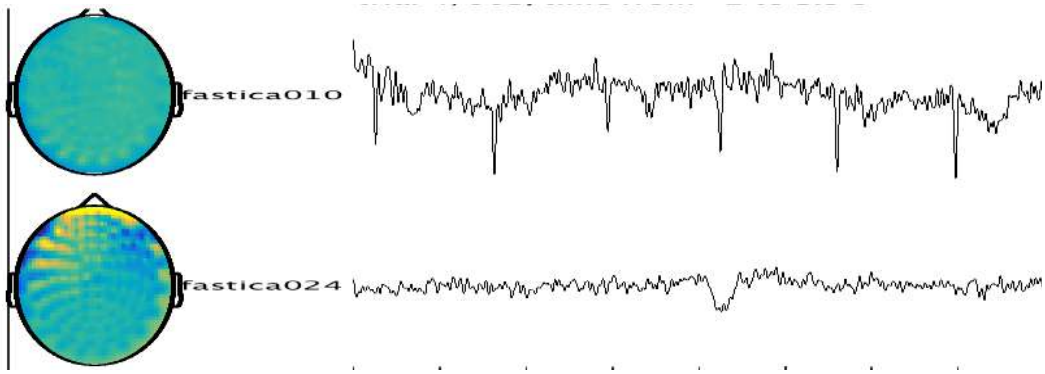


Figure 13: ECG and EOG artifacts with topographic plots from the data used in the thesis

Data analysis methods:

Behavioral Analysis

Behavioral analysis (i.e., a measure of reaction times: RTs) has been a classical and widely used method in psychology and neuroscience for noninvasively studying mental processing: RTs are assumed to reflect the time required to complete the perceptual and motor-planning computations and to plan a response (Donders, 1969; Friston et al., 1996; Sternberg, 1969). This assumption appears justified by evidence that RTs are modulated by factors such as stimulus complexity, stimulus-response compatibility, number of potential responses, or required response accuracy (Fischman, 1984; Fitts, 1954; Henry and Rogers, 1960; Kaswan and Young, 1965). Generally, the mean response for a participant is calculated by averaging the reaction for all the trials. Later, the average responses for all participants can be statistically compared across conditions. The mean and standard deviation across conditions is generally reported. In fact, beyond the mean, an important parameter to consider is *variance*. ANOVA (analysis of variance) is a statistical method that splits an observed aggregate variability found in a data set into two components. First, the systematic factors that have a statistical influence on the data set, and second, the random factors that don't have a statistical influence on the data set. Generally systematic factors affecting variance are typically the ones manipulated in

the experimental paradigms. Another important approach used in behavioral literature is linear mixed models that allow both fixed and random effects to be considered jointly and are particularly used when there is non-independence in the data, such as data arising from a hierarchical structure.

In this thesis, I have reported the mean reaction times across the main experimental conditions, analyzed with linear mixed models (*lmer* toolbox implemented in R software).

Evoked Analysis:

Event-related fields (ERF) reflect neural activation that occurs at the same time (time-locked and phase-locked) with respect to stimulus onset or task onset for all the trials of an experiment. Evoked responses typically develop within a few hundred milliseconds from the stimulus presentation time or close to the execution of the task. Evoked single-trial responses may be detectable in some cases (such as open eyes vs closed eyes); however, some tens or hundreds of trials are generally used to yield an average evoked response with a good signal-to-noise ratio. The earliest short-latency salient responses are typically transient (short-lasting) and tightly time-locked to the stimulus. This yields sharp responses even when averaged across several trials. The longer latency responses tend to progressively increase in duration and exhibit more jitter to the stimulus timing; while averaged, they appear as sustained responses with a slow fade in and out phases. There are several evoked components (such as M100, M300, M400, etc.), which are already established as neural markers for different brain processing stages. In this thesis, I will only investigate M100 response (i.e., response to a stimulus peaking at ~100 milliseconds). In predictive processing literature, M100 has been related to the prediction error signals (Den Ouden et al., 2012; Phillips et al., 2015). If a stimulus is predictable, it is considered to have a reduced M100 peak (more prediction, less prediction error) compared to unpredictable stimuli (less prediction, more prediction error).

Such potential differences have been assessed statistically using a repeated-measure ANOVA (Dien, 2017) and cluster-based permutation test approach (Maris and Oostenveld, 2007).

Source Reconstruction:

A time window of interest is selected based on the statistically significant differences observed in the sensor space data, and then source reconstruction is computed in this time window only. Source reconstruction of the sensor data involves two stages. The first stage is generating a *forward model* based on either participant's individual MRI or using a standard brain template. The forward model helps to understand how does a given individual source contributes to the sensor data. The second stage is creating an *inverse model*, which does an inverse mapping from sensor space data to individual sources by using the weights of the forward model.

Forward Model:

The initial step in solving the forward problem is to generate an individual volume conduction model of the patient's head using her/his structural T1 MRI image. The most common head models are spherical shell head models (de Munck and Peters, 1993). These models assume that the human brain is sphere-shaped. On the opposite, the realistic head model makes use of electrically conductive properties and geometric properties of the human brain. The information about the geometry of the brain is extracted from the structural MRI. In case there is no individual MRI, a brain template/atlas can also be used. While choosing a head model, it should be considered that electrophysiological signals (EEG) are affected by different tissue types (gray matter, white matter), skin, and hair, as the electric current is distorted by all these factors. In contrast, the magnetic field is unaffected by the anisotropic conductance of different tissues. So, a single shell head model is sufficient for creating a forward model for MEG. There is a wide range of realistic head models such as Boundary Element Method (BEM) (Fuchs et

al., 1998; Hämäläinen and Sarvas, 1989), Finite Element Method (FEM) (Thevenet et al., 1991), and finite difference method (FDM) (Hallez et al., 2005). Before computing a forward model, the structural data (i.e., MRI) and functional data (i.e., MEG) must be aligned in a common spatial frame. The process is called MEG-MRI *coregistration*. While recording the functional data, the participant's head is digitized for three fiducial points (i.e., Nasion, Right Pre-Auricular (RPA), and Left Pre-Auricular (LPA)), and around three hundred (300) points are marked virtually on the participant's head. These points help to get a shape of the head which is later coregistered with structural data. This coregistration can be done either manually by matching fiducial points and head points or can be done via automated methods like the Interactive closest point algorithm (ICP) (Rusinkiewicz and Levoy, 2001).

The next step after generating a head model is to compute the lead field. The lead field operator (L) embodies all the biophysical and anatomical assumptions one needs to account for in the forward model. The L links the current density J in the brain at location rJ with orientation θJ to the magnetic field B measured at sensor location r. To define the location (x, y, z) of each current, it is necessary to segment the volume of the brain (often called the source spaced) in voxels of constant size (e.g., 5 × 5 × 5 mm voxels). The ε models an additive measurement noise at sensor location r, which is usually assumed to follow a Gaussian distribution with zero mean and a parameterized variance structure (Mattout et al., 2006).

$$B(r)=L(r,rJ,\theta J)J(rJ,\theta J)+\varepsilon(r)$$

Significantly, the magnetic field varies linearly with current amplitude, and magnetic fields produced by several dipoles are simply additives due to the linearity of Maxell's equations. Therefore, if B is an NB × 1 vector containing the magnetic field measured in all NB sensors,

is a Nε × 1 vector containing the noise measured in all Nε sensors and J is a NJ × 1 vector containing the amplitude of all NJ active sources, one can write

$$B = LJ + \varepsilon ,$$

where L is a NB × NJ lead field matrix.

Inverse Model:

Inverse modeling is an ill-posed problem, i.e., it may have non-unique (more than one) and unstable solutions which are highly sensitive to noise. Brain signals like MEG are naturally contaminated by noisy nuisance components (such as environmental noise and human body physiological artifacts), which should not be accounted for the inverse modeling of the brain activity. Several methods have been used to compute sources of MEG activity in sensor space. One such method assumes that the measured magnetic signal is generated by a single dipole, e.g., so-called equivalent current dipole (ECD), which is characterized by a set of parameters. Precisely, the ECD's position, orientation, and amplitude are interactively estimated to explain the measured sensor-level MEG signal at best. The main parameter assessing the certainty of an ECD model is the goodness of fit (GoF), defined as:

$$\text{Goodness of Fit (GoF)} = \quad 1 - \frac{\|B - \hat{B}\|2}{\|B\|2}$$

The GoF quantifies the agreement between the measured MEG signals B and the $\hat{B}$ signals that would be produced by this ECD at a given time.

Another popular approach to solve the inverse problem is to assume that the recorded brain MEG signal is generated by multiple sources distributed through the source space. One of the challenges for distributed inverse methods is that the number of currents (sources) by far exceeds the number of MEG sensors. Therefore, an infinite number of current distributions can explain the observed MEG signals. The non-uniqueness is a situation where an inverse problem is said to be ill-posed. This question has been addressed with the physics of ill-posedness and inverse modeling, which formalize the necessity of including additional mathematical and physical constraints in the model to find a unique solution. Different contextual information assumes a family of inverse solution methods, e.g., among others, minimum norm (MN) and beamforming estimations.

In the case of the beamforming approach (used in the present thesis, (Veen et al., 1997)), it is assumed that all sources are uncorrelated. For that, a weight vector w(rJ) to apply to B is estimated through the following minimization problem.

$$w(rJ) = \text{argmin}_w E(\|wB\|2) \text{ constrained to } wL(rJ) = I$$

In this minimization problem, the constraint ensures that the activity coming from the source located in rJ is reconstructed with unit gain while minimizing the power from other sources. If C denotes the NB × NB covariance matrix of the magnetic field (B) and L(rJ) the NB × Nθ leadfield matrix corresponding to sources at location rJ with Nθ orthogonal source orientations (Nθ ε {1,2,3}),

$$w(rJ) = [L(rJ)TC-1L(rJ)]-1L(rJ)TC-1$$

By evaluating the activity in all sources positioned on a grid covering the brain, one can compute a topographic map of current densities. Source reconstruction algorithms project sensor space data to source space to localize neural activity within the brain. In this way, spatiotemporal maps of cerebral activity can be produced to visualize the brain regions involved in performing a specific task.

MVPA: Time-resolved decoding

Evoked responses have been widely used to explore the neurocognitive foundations of a given effect. However, evoked responses do not provide information about how the brain is sensitive to specific features of a stimulus, a factor that is usually removed while averaging the data. Multivariate pattern analysis techniques (MVPA or "decoding methods") have gained popularity as they consider neural responses as a pattern of activity that enables the investigation of different time-varying brain states that a system or cortical field can generate. These methods provide an advance over the univariate methods. The first MVPA study which became popular came from Haxby and colleagues (Haxby et al., 2001) and backed-up by two studies in nature neuroscience (Haynes and Rees, 2005; Kamitani and Tong, 2005) drastically changed the field of MVPA methods application in functional Magnetic Resonance Imaging (fMRI). Using MVPA, cognitive scientists became capable of decoding stimulus information in the human brain. The techniques combined with non-invasive imaging modality (fMRI) enabled scientists to scrutinize the stimuli specific (or feature specific) information in different brain areas. However, human brain processing is very efficient in terms of temporal processing. It takes us a few milliseconds to recognize a friend, family members, objects, etc. Non-invasive imaging techniques having a good temporal resolution (EEG and MEG) combined with MVPA techniques have gained a lot of popularity in recent years and are used in several applications like Brain-Computer Interfaces, lie detection, and detection of brain disorders. There are a

variety of domains where MVPA methods can be used such as time series, time-frequency representation, wavelet, and ICA-based decomposition, connectivity, etc. To limit the scope of this thesis, I will limit my MVPA analysis only to time-series (decoding evoked responses).

The core concept behind time-resolved decoding is explained in figure 14. I will explain time-resolved decoding with a hypothetical experiment. First, consider an imaging modality (either EEG or MEG) which records electrical currents (in case of EEG) or magnetic field (in case of MEG) continuously. In every trial, two images, either house or a human face, are recorded, and each image is presented 40 times. The data is preprocessed to remove the artifacts (eye blinks, ECG, muscle, etc., described above). Every participant has a data matrix containing the number of trials (n_{trials}, 80 in given example), sensors ($M_{sensors}$, 204 gradiometers in given example), and time points (t_k where k refers to the number of time points, 100 time points in given example). To understand, let's assume the data is recorded with just one sensor. For a single sensor classification, the amplitudes at a given time point (let’s assume first time point) act as classification features. So, for a given time point, the amplitude values of 80 trials at a given sensor and given time. The data is divided into different chunks called k-fold (generally 5-fold) to avoid overfitting of classifiers, and the classifier uses most of the data (4-folds) for training and the rest of the data for testing. This process is repeated over all the sensors and all the time points and averaged across sensors. This gives a classification accuracy curve which represents a classification accuracy at every time point. The concept of time-resolved decoding is illustrated in the figure below.

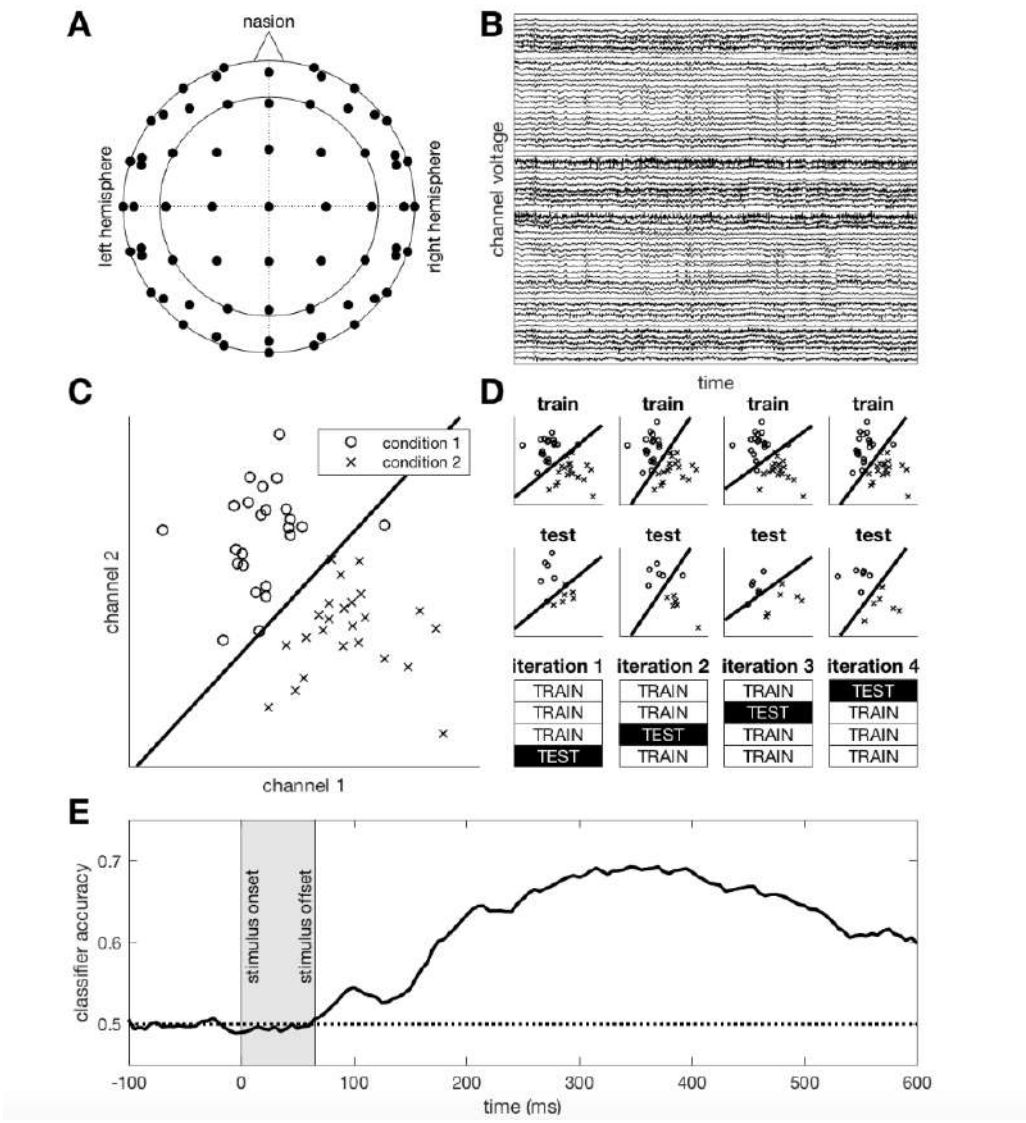


Figure 14: A) A standard EEG recording electrode template. B) All the channels showing data as a function of time C) Plot of data from two different conditions and separating using a linear decision boundary D) Standard 4 fold cross validation illustration. E) Time resolved decoding accuracy curve. (Figure taken from (Carlson et al., 2019)).

Chapter 3: Expectation Suppression in Vision

3.1 Introduction:

In this experiment, I aim to investigate the expectation suppression in the visual domain, and how expectation suppression is affected by the context ("*what*") and timing (*"when"*). To do so, I designed a 2-by-2 factor experimental paradigm. The first manipulated factor was the context/identity of the presented stimuli ("*what*" manipulation), which either provides stronger predictions about the target (by presenting a series of cues rather than one cue reported as in previous studies) developing an expectation about the next stimuli. The second factor was the timing ("*when*" manipulation), so that the participants could either predict the onset of next incoming stimuli, or not. To avoid the interacting effects of prediction and attention driven by the task assigned to the participants, an orthogonal relationship between different features was manipulated (i.e., predictions generated for one feature and task associated to another feature). This also allows us to generate a prediction versus no prediction paradigm, as there was no violation of predictions involved. Detailed explanation about the experimental paradigms is provided in the Methods and Material section of this chapter.

The issues addressed by this experiment are the following:

i) How expectation suppression is affected by "*what*" manipulation. Will the suppression increase as the predictability about the next stimuli becomes stronger and stronger in a sequence? What are the brain areas that are showing this suppression in neural responses?

ii) How does the temporal uncertainty (*when* factor) affect the expectation suppression? Will the predictable timing increase the suppression carried out by the *what* manipulation or attenuate the suppression effect? What are the brain areas that will show this suppression?

iii) How does the temporal affect the stimulus specific decoding? Can we decode the predictable feature across the sequence of Entrainers? Can the predictions generated by the Entrainers be decoded in the pre-stimulus time region ?

Considering all the goals mentioned above, we used the MEG, which has an excellent temporal resolution and good spatial resolution, making it a perfect choice for mapping the spatio-temporal manipulation in our experimental design. Details about MEG modality have already been explained in chapter 2.

3.1 Methods and Materials:

This section's methods, results, and discussion were published in the NeuroImage (Nara et al., 2021) journal in May 2021.

Participants:

Twenty healthy participants were recruited from the BCBL Participa website (https://www.bcbl.eu/participa/). Out of twenty, four participants were rejected from the study for having a lot of artifacts (more than 50 % of trials). Sixteen participants (7 females) were included in the present study (age range: 19 – 31 years old; M = 24.8 y.o. ; SD = 3.6 y.o.). The ethical committee and scientific committee of the Basque Center on Cognition, Brain and Language (BCBL) approved this experiment (following the principles of the Declaration of Helsinki). Participants were informed about the experiment and MEG imaging modality; all the participants gave written informed consent and were financially compensated for their valuable contribution to the experiment.

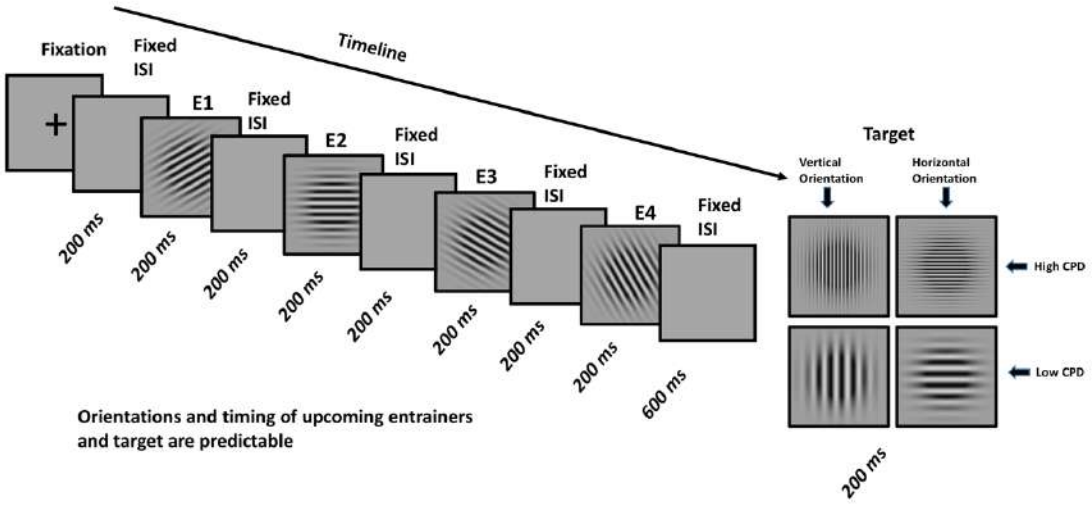
Experimental Design:

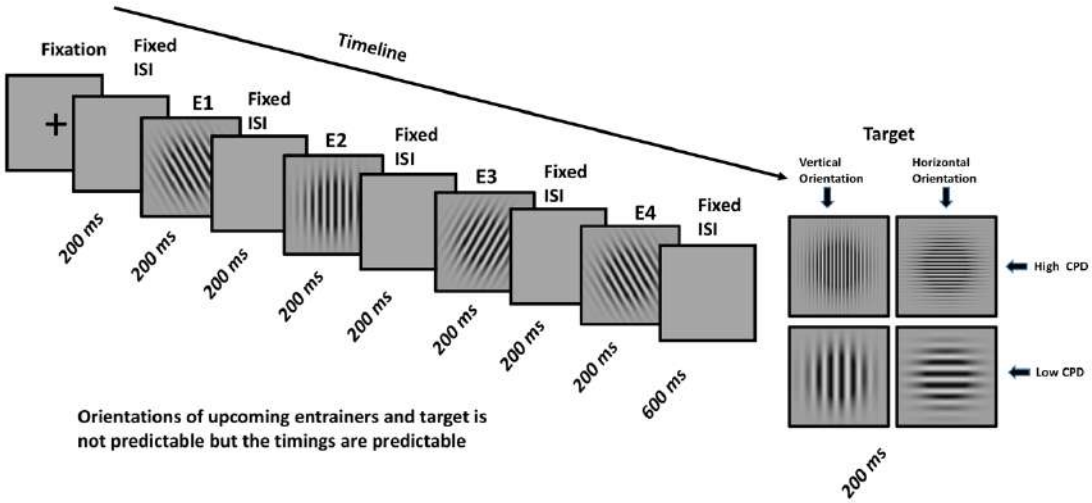
We presented a series of Gabor patches manipulating two dimensions, i.e., orientation (expressed in degrees) and spatial frequency (expressed in cycles per degree of visual angle, CPD). All the Gabors were presented in the center of the screen with a gray background using a projector screen, while optimum luminance was maintained in the MEG gantry. Every trial began with a fixation cross followed by a sequence of four Gabor patches (called Entrainers herewith) presented for 200 ms each sequentially with an inter stimulus interval of 200 ms. A fifth Gabor (called Target herewith) was presented for 200 ms (after an inter stimulus interval from the fourth Gabor of 600 ms), and participants had to perform a task associated with it. The Entrainers had an intermediate number of cycles per degree (40 CPD) for visual angle:

participants had to detect if the target had either a higher (60 CPD) or a lower (20 CPD) with respect to the Entrainers. There were four manipulations in the experimental design (Figure 15): a) first, the orientation of the Target Gabor could be either horizontal or vertical; b) second, the Target Gabor's CPD could be higher or lower compared to the CPD of the the previous four Entrainers; c) third, the orientation angle of the Entrainers could be either predictable (i.e., scaled in fixed step angle of either 15 or 30 degrees: for example, 30, 45, 60 75 and 90 degrees of orientation angle) or not (for example 60, 30, 75, 45 and 90 degree); d) fourth, the timing between all the Entrainers and the Target could be either predictable (i.e., fixed interstimulus interval-ISI of 200 ms in between the first four Entrainers and 600 ms between Entrainer 4 and Target) or unpredictable (varying interstimulus interval ranging between 70-330 ms for first four Entrainers and 450-770 ms between Entrainer 4 and Target).

Participants' task was to determine whether the Target Gabor had a higher or a lower CPD with respect to the Entrainers. Participants responded by pressing left or right MEG compatible button presses (counterbalanced and randomized across participants). Depending on the orientation and interstimulus timing between all the Entrainers and the Target, the trials were divided into four conditions (Figure 15): *what + when*, *when*, *what,* and *random*. In the first condition (i.e., *when + what*), all the Entrainers had a predictable timing (i.e., fixed ISI of 200 ms), and the orientation angle of the Target Gabor was also predictable based on the previous Entrainers. In the second condition (i.e., *when* condition), the timing between all the Entrainers and the Target Gabor is predictable, but the Target Gabor's orientation was unpredictable. In the third condition (i.e., *what* condition), the Target Gabor's orientation was predictable, but the timing was unpredictable between all the Entrainers and the Target Gabor. In the fourth condition (i.e., *random* condition), both the Target Gabor's orientation and the timing between Entrainers and the Target Gabor are not predictable. A total of 160 trials were presented for

each condition (40 horizontal targets with higher CPD, 40 horizontal targets with lower CPD, 40 vertical targets with higher CPD, 40 vertical targets with lower CPD), leading to a total of 640 trials per participant.





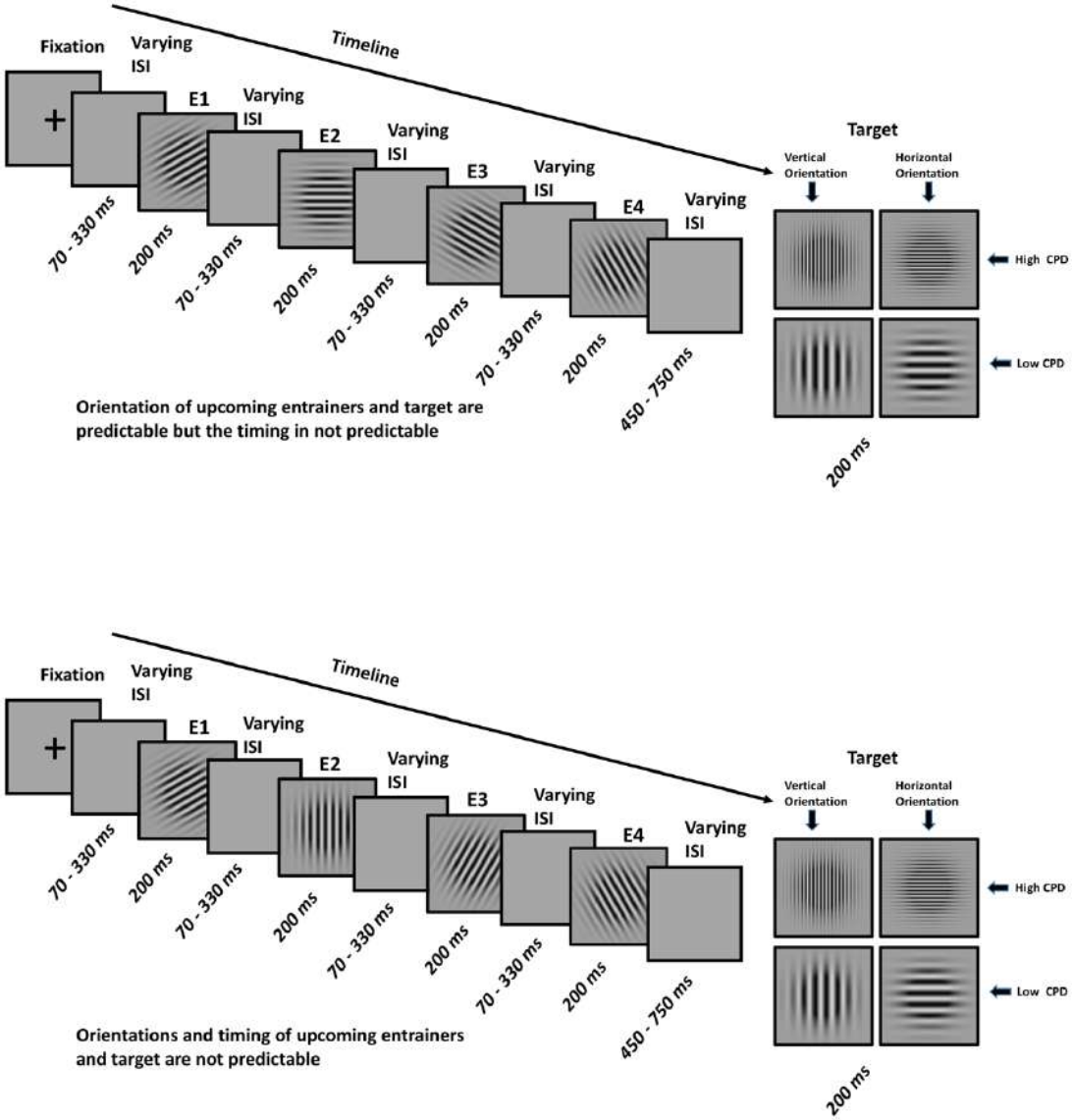


Figure 15: Experimental design. A) Orientations and timing of upcoming Entrainers and target are predictable (what+when condition). B) Orientations of upcoming Entrainers and target are not predictable but the timings are predictable (when condition). C) Orientation of upcoming Entrainers and target are predictable but the timing is not predictable (what condition). D) Orientations and timing of upcoming Entrainers and target are not predictable (random condition). Abbreviations: E1 – Entrainer 1, E2 – Entrainer 2, E3 – Entrainer 3, E4 – Entrainer 4, ISI – Inter Stimulus Interval.

Data acquisition and preprocessing:

MEG data were acquired in a magnetically shielded room using the whole-head MEG system. The MEG system contains 102 sensor triplets (each comprising a combination of one magnetometer and two orthogonal planar gradiometers) uniformly distributed around the participant's head. Head position inside the MEG helmet was continuously monitored using a set of five Head Position Indicator (HPI) coils. Each coil's location was defined relative to the anatomical fiducials (nasion, left and right preauricular points) with a 3D digitizer (Fastrak Polhemus, Colchester, VA, USA). This procedure was critical for head movement compensation during the data recording session. Digitalization of the fiducials points along with ~300 additional points evenly distributed over the scalp of the participant's head was used during data analysis to spatially align the MEG sensor coordinates with T1 MRI images recorded on a 3T MRI scanning facility installed at BCBL(Siemens Medical System, Erlangen, Germany). MEG recordings were acquired with a sampling rate of 1 kHz and a bandpass filter at 0.01-330 Hz. Eye movements were recorded with two pairs of electrodes in a bipolar montage, one placed on the external chanti of each eye (horizontal electrooculography (hEOG)) and another, above and below the right eye (vertical EOG). Similarly, cardiac rhythm (ECG) was also monitored using two electrodes, one placed on the right side of the participants' abdomen/chest and another, below the left clavicle.

Continuous MEG data were preprocessed using the temporal Signal- Space-Separation (SSS) method (Taulu and Simola, 2006) which suppresses external interferencing noise. MEG data were also corrected for head movements, and bad channel time courses were reconstructed using interpolation algorithms implemented in the maxfilter software. Subsequent data analyses were performed using Matlab R2014b (Mathworks, Natick, MA, USA).

Reaction times:

Participants were asked to indicate if the CPD of the Target Gabor was higher or lower compared to the previous Entrainers. We made sure that accuracy was high and comparable across conditions. The response time (RT) was calculated for all the four conditions (i.e., *what + when, when, what and random*). The response times longer than 1500 ms were considered as outliers and were removed from the analysis. The mean reaction time and standard deviation is computed for each experimental condition.

Sensor level ERFs:

MEG trials were first corrected for jump and muscle artifacts using standard coherence-based automated scripts using Fieldtrip toolbox (Oostenveld et al., 2011) implemented in MATLAB 2014B. EOG and ECG artifacts were identified using Independent Component Analysis method (ICA) and the linearly subtracted from the MEG recordings. The ICA decomposition (30 components extracted per participant) was performed using the FastICA algorithm implemented in the Fieldtrip toolbox. ICA components maximally correlated with ECG and EOG recordings were automatically removed. On average, two components were removed per participant. The artifact-free data were bandpass filtered between 0.5 and 45 Hz. Trials were segmented time-locked to each of the Entrainers (Entrainers 1, 2, 3, and 4) and the target. The trial segments were grouped together for each entrainer and target, and then averaged to compute the ERFs. For each planar gradiometer pair, ERFs were quantified at every time point as the Euclidean norm of the two gradiometer signals. Baseline correction was also applied to the evoked data based on the 400 ms of data prior to the onset of the fixation cross presented at the beginning of each trial.

In brief, I first performed the analysis of the sensor level data to establish expectation suppression and its interactions with different kinds of predictability. We then move on to a

more detailed analysis of the functional anatomy of expectation suppression using source constructed data.

We applied an ANOVA to sensor-level data to explore the influence of our experimental factors on visual ERFs. First, we extracted ERF amplitudes in the set of five occipital sensors that had shown maximum response to the visual localizers. We then selected the time window classically associated with the initial visual evoked response (85–135 ms post stimulus). A three-way repeated measures ANOVA was computed in JASP (JASP Team, 2020) with these amplitude values as dependent variables and the following factors: entrainer (four levels; corresponding to Entrainers 1, 2, 3, 4); *what* (two levels; predictable/unpredictable entrainer and target orientations); and *when* (two levels; predictable/unpredictable timing of Entrainers and target). Significant interactions (specifically, the triple interaction "entrainer * *what* * *when*") were further investigated through theoretically relevant pairwise comparisons.

Pairwise comparisons between conditions were performed using a cluster-based permutation test (Maris and Oostenveld, 2007). A randomization distribution of cluster statistics was constructed for each subject over time and sensors and used to evaluate whether conditions differed statistically over participants. In particular, t-values were computed for each sensor (combined gradiometers) and each time point during the 0–270 ms time window, and were clustered if they had t-values that exceeded a t-value corresponding to the 99.99th percentile of Students t-distribution, i.e. a two-tailed *t*-test at an alpha of 0.01, and were both spatially and temporally adjacent. Cluster members were required to have at least two neighboring channels that also exceeded the threshold to be considered a cluster. The sum of the t-statistics in a sensor cluster was then used as the cluster-level statistic, which was then tested by permuting the condition labels 1000 times.

Four different comparisons were carried out. In the first comparison, we contrasted ERFs for the *when* and the *what+when* conditions. This comparison evaluated the effect of orientation predictability when the timing of the Entrainers and target were predictable. In the second comparison, we compared ERFs for the *random* and *what* conditions. This comparison evaluated the effect of orientation predictability when timing was unpredictable. These two comparisons mainly focused on the main effect of orientation predictability (i.e., the *what* manipulation) revealed by expectation suppression. We then compared the ERFs for the *what+when* and *what* conditions. Here we directly contrasted these two predictable orientation conditions to evaluate the effect of temporal predictability on stimulus predictability. The final comparison contrasted ERFs in the *when* and *random* conditions. This comparison was performed to analyze the effect of temporal predictability in the absence of orientation predictability.

Source level ERFs:

MEG-MRI co-registration was performed using MRIlab software (Elekta Neuromag Oy, version 1.7.25). Individual T1-weighted MRI images were segmented into scalp, skull, and brain components using the segmentation algorithms implemented in Freesurfer software (Martinos Center of Biomedical Imaging, MQ) (Dale et al., 1999). The source space was defined as a regular 3D grid with a resolution of 5 mm and the lead fields were defined using a realistic single-shell Boundary element model (BEM) head model (Fuchs et al., 1998). Both the planar gradiometers sensors and magnetometers sensors were used for inverse modelling. The covariance matrix was estimated combining the information embedded in the pre-stimulus time range (i.e., -400–0 ms prior to the cross fixation) and the post-stimulus time range (i.e., 0–400 ms after the presentation of Gabor). Whole brain source activity was estimated using linearly constrained minimum variance (LCMV) beamformer approach (Veen et al., 1997).

LCMV beamformer was computed on the evoked data in the time period 85–125 ms post-stimulus and in the pre-stimulus interval for all the four conditions at Entrainer 4. Source reconstruction was performed only at Entrainer 4 as the strongest effect of prediction were present at E4. For each source in the brain, the neural activity index was calculated as the ratio of the mean power in the post-stimulus and the pre-stimulus interval. Brain maps containing neural activity index (NAI) values were transformed from the individual MRIs to the standard Montreal Neurological Institute (MNI) for computing the group level analysis. For that, we applied a non-linear transformation using the spatial-normalization algorithm implemented in Statistical Parametric Mapping (SPM8) (Friston et al., 1994). The Freesurfer's *tksurfer* tool was used for visualizing the brain maps in the MNI space. For each condition and Entrainer, we obtained the NAI value and the MNI coordinates of local maxima (sets of contiguous voxels displaying higher power than all other neighboring voxels) (Bourguignon et al., 2018). NAI values were compared between conditions (e.g., *when* vs. *what + when, random* vs. *what, what vs. what + when* and *when vs. random*) using t- tests.

MVPA: time-resolved decoding:

Time-resolved within-subjects multivariate pattern analysis was performed to decode the feature specificity (i.e., the orientation angle and CPD of Gabor) from MEG data. This within-subject classification has an advantage over other methods: the classification algorithm may leverage individual subject specific characteristics in neural patterns since the classifiers do not need to generalize across different subjects. For E1, E2 and E3, data were segmented from 50 ms prior to 250 ms after the onset of the Entrainers. The time interval between E4 and the Target was longer than the time interval between the rest of the Entrainer. For this reason, E4 data were segmented from 50 ms prior to 600 ms after the onset of the Entrainer. The artifact free clean data was segmented from -400 ms to 550 ms time locked to the Target. The data

were classified using a linear support vector machine (SVM) classifier with L2 regularization. The classifier was implemented in Matlab using the LibLinear package (Fan et al., 2008) and the Statistics and Machine Learning Toolbox (Mathworks, Inc.). We performed a binary classification of the orientation of each Gabor depending on the orientation of the subsequent Target. In other words, the class labels (i.e., horizontal or vertical) were derived from the Target orientation: if the Target orientation was horizontal, all the preceding orientation in the corresponding conditions were labelled as horizontal and vice-versa. Same approach was used for labelling the classes for CPD feature decoding.

The data were down-sampled by a factor of five (i.e., new sampling frequency 200 Hz) prior to the classification. Pseudo trials were generated to improve the SNR by creating trials' bins, resulting in a set of 10 trials for each bin (Dima and Singh, 2018). This pseudo trial generation was repeated 100 times to generate trials with a higher signal to noise ratio. The data were then randomly partitioned using 5-fold cross-validation. The classifier was trained on 4 folds and tested on 1 fold and this process was repeated until each fold was left out once. The procedure of generating pseudo trials, dividing the data into 5 folds, and training and testing classifiers at every time point was repeated 25 times; classification accuracies were then averaged over all these instances to yield more stable estimates. To improve data quality, we also performed multivariate noise normalization (Guggenmos et al., 2018). The time-resolved error covariance between sensors was calculated based on the covariance matrix of the training set and used to normalize both the training and test sets in order to down weight MEG channels with higher noise levels. Cluster corrected sign permutation tests were applied to the accuracy values obtained from the classifier with cluster-defining threshold $p < 0.05$, corrected significance level i.e., cluster-alpha $p < 0.01$.

3.3 Results

Reaction times:

what + *when* condition showed faster response time (727.5 ± 204 ms) while the *random* condition slower (752 ± 207 ms). In the *when* and the *what* conditions participants responded respectively after (737.4 ± 203) ms and (741.4 ± 204) ms. These data indicate that temporal predictability determines quicker responses (*what* + *when* > *what* & *when* > *random*). We fit a Linear mixed model (LMER) considering subjects and observations as *random* effects, and the fixed effects of *what* (orientation predictable or not), *when* (timing predictable or not) and their interaction as fixed effects. We observed a main effect of *when* ($p<0.05$).

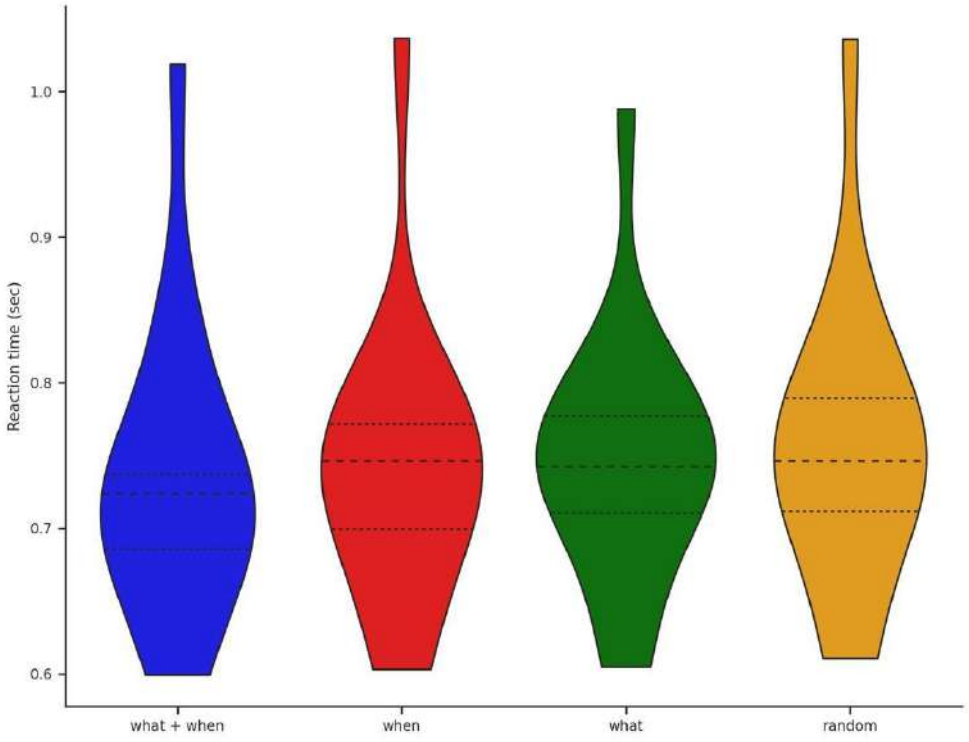


Figure 16: Reaction times distribution for CPD detection task across conditions

Sensor level ERFs:

We first analysed the amplitude of the initial visual evoked response for 5 occipital sensors to determine how the two-by-two experimental design modulated visual responses across Entrainers. In the three-way ANOVA (details in Table 1), we observed a significant main effect of entrainer ($p < 0.001$) on the peak amplitudes of the ERFs. Interestingly, this factor interacted with the factor *what* ($p < 0.001$), suggesting that Gabor orientation predictability affected visual-evoked responses differently across Entrainers. We should point out that a main effect of *what* ($p < 0.001$) supported the observation that orientation predictability influenced visual processing. Importantly, the interaction between the three factors, i.e., Entrainers, *what* and *when*, was significant ($p = 0.004$). This triple interaction underlines the fact that timing uncertainty influenced the development of visual predictions across the sequence of four Entrainers.

Figure 17 shows the sensor-level Event Related Fields (ERFs) time-locked to the onset of each Entrainer (E1, E2, E3 and E4) and Target (T) for the *when* and the *what + when* conditions. The amplitude of the ERFs was significantly (cluster *p*- value < 0.01) higher for the *when* compared to the *what + when* condition for E2, E3 and E4, but not for E1 or Target. The amplitude enhancement for the *when* compared to the *what + when* condition emerged within the [95 – 105] ms, [96 – 110] ms and [97 – 121] ms time intervals for the E2, E3 and E4 respectively. This effect was located in occipital sensors for all the Entrainers. Figure 18 shows the sensor-level ERFs for the *random* and the *what* conditions. The amplitude of the ERFs was significantly (cluster p-value < 0.01) higher for the *random* compared to the *what* condition for E2, E3 and E4, but not for E1. The amplitude enhancement for the *random* compared to the *what* condition emerged within the [95 – 119] ms, [94 – 123] ms and [96 – 127] ms time intervals for the E2, E3 and E4 respectively. This effect was located in occipital sensors for all the Entrainers. It is worth noting that the comparison having temporal predictability (i.e., *when*

vs *what* + *when*) involves more sensors compared to the comparison lacking temporal predictability.

Since both the comparisons are significant from Entrainer 2 onward, we compared the orientation predictability with (*what* + *when*) and without temporal predictability (*what*). Figure 19 shows that the initial very early evoked activity (0–75 ms) at E1 is similar for both conditions. As we move across Entrainers such early difference increases and is statistically significant ($p < 0.001$) but it vanishes at the Target. Since our focus in this analysis is in the early evoked response to the visual stimulus showing expectation suppression effects, the time window for statistical comparison was selected from 75 to 135 ms. The amplitude enhancement for the *what* compared to the *what* + *when* condition emerged within the [106 – 124] ms time interval only at E4. This effect was located in occipital sensors. Figure 20 shows the sensor level results of the ERFs for the *when* and the *random* conditions. This last comparison also shows the difference in the very early evoked activity ($p<0.001$). This comparison does not show any significant results for the early visual response at any Entrainer.

Table 1: Repeated measure ANOVA with the factors entrainer (four levels, one for each entrainer), *what* (two levels: orientation predictable or not) and *when* (two levels: timing predictable or not).

	Sum of Squares	df	Mean Square	F	p
entrainer	2.607e -22	3	8.691e -23	42.299	< .001
what	3.430e -23	1	3.430e -23	65.203	< .001
when	2.969e -24	1	2.969e -24	2.376	0.144
entrainer * *what*	2.572e -23	3	8.572e -24	18.503	< .001
entrainer * *when*	8.851e -25	3	2.950e -25	0.860	0.469
what * *when*	3.032e -25	1	3.032e -25	0.833	0.376
entrainer * *what* * *when*	2.277e -24	3	7.589e -25	5.073	0.004

Note. Type III Sum of Squares

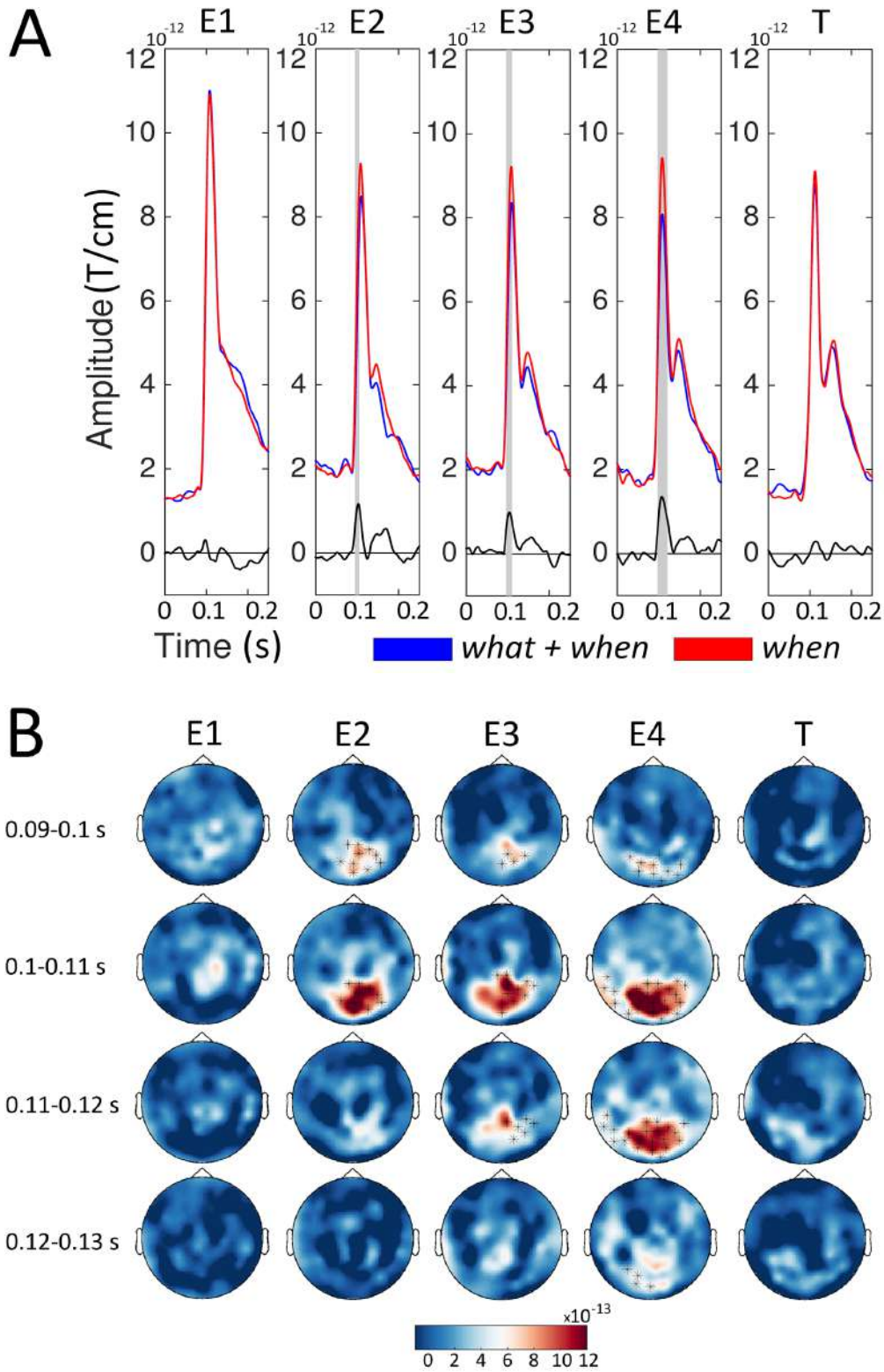


Figure 17: Sensor level ERFs for the when and what+when conditions. A) For each condition (red, when; blue, what+when) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located above occipital regions (MEG02042/3, MEG2032/3, MEG2342/3, MEG2122/3, and MEG1922/3). Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.01$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the when and what+when conditions in temporal windows around the amplitude peak value. Sensors showing significant differences ($p < 0.01$) are highlighted

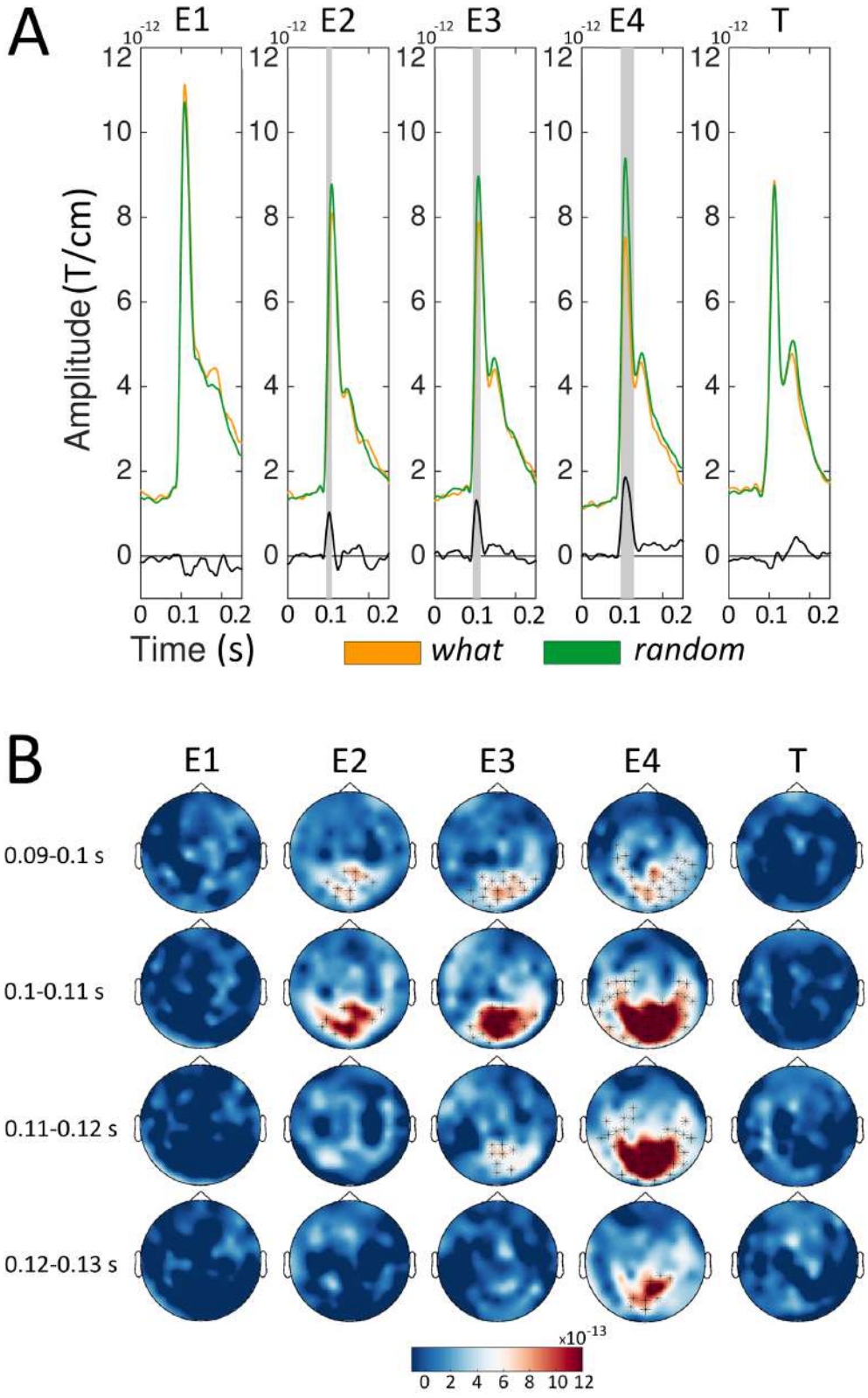


Figure 18: Sensor level ERFs for the random and what conditions. A) For each condition (orange, what; green, random) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located above occipital regions (MEG02042/3, MEG2032/3, MEG2342/3, MEG2122/3, and MEG1922/3). Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.01$) for the random than

the what condition. B) Sensor maps of the ERF difference between the random and what conditions in temporal windows around the amplitude peak value. Sensors showing significant differences ($p < 0.01$) are highlighted.

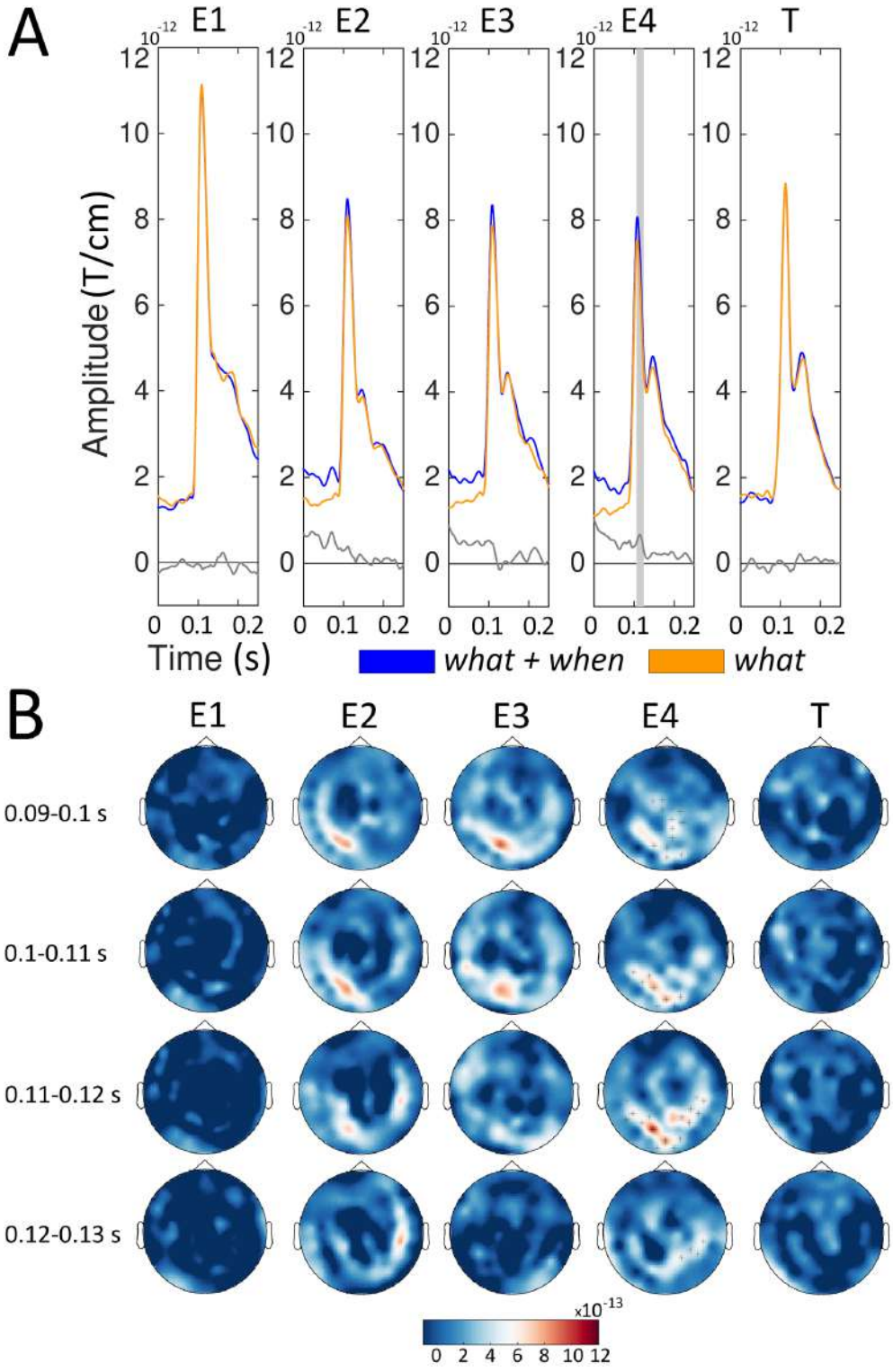
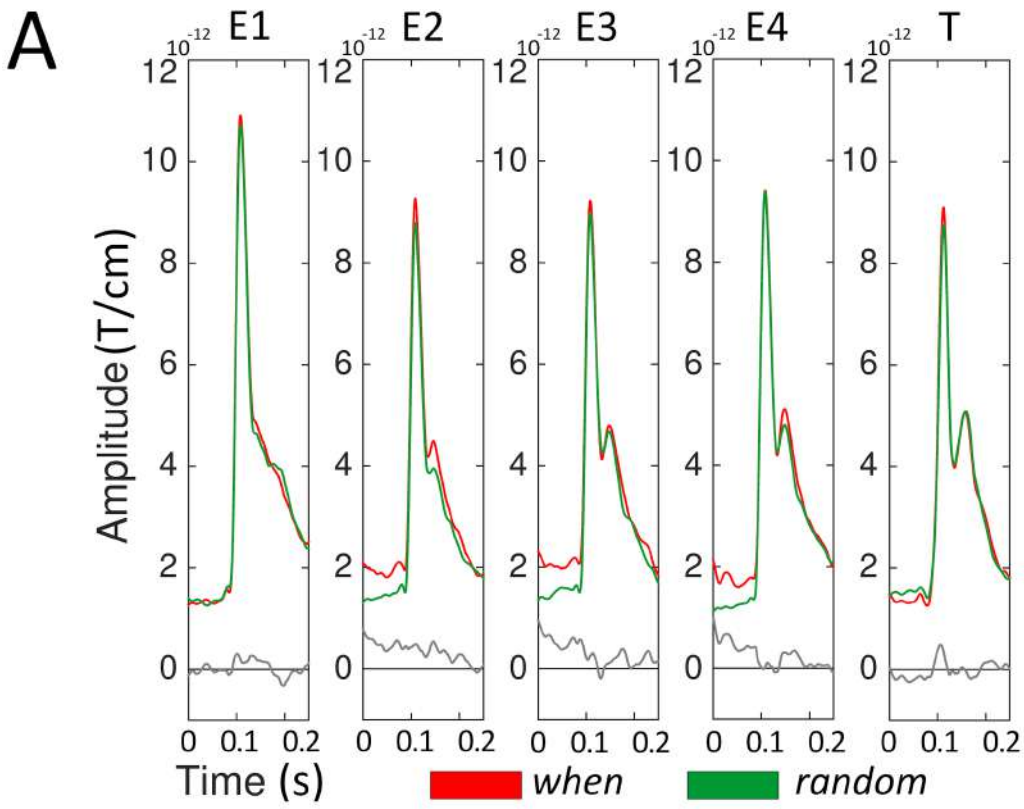


Figure 19: Sensor level ERFs for the when and what+when conditions. A) For each condition (orange, what; blue, what+when) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located above occipital regions (MEG02042/3, MEG2032/3, MEG2342/3, MEG2122/3, and MEG1922/3). Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.05$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the what and what+when conditions in temporal windows around the amplitude peak value. Sensors showing significant differences ($p < 0.01$) are highlighted



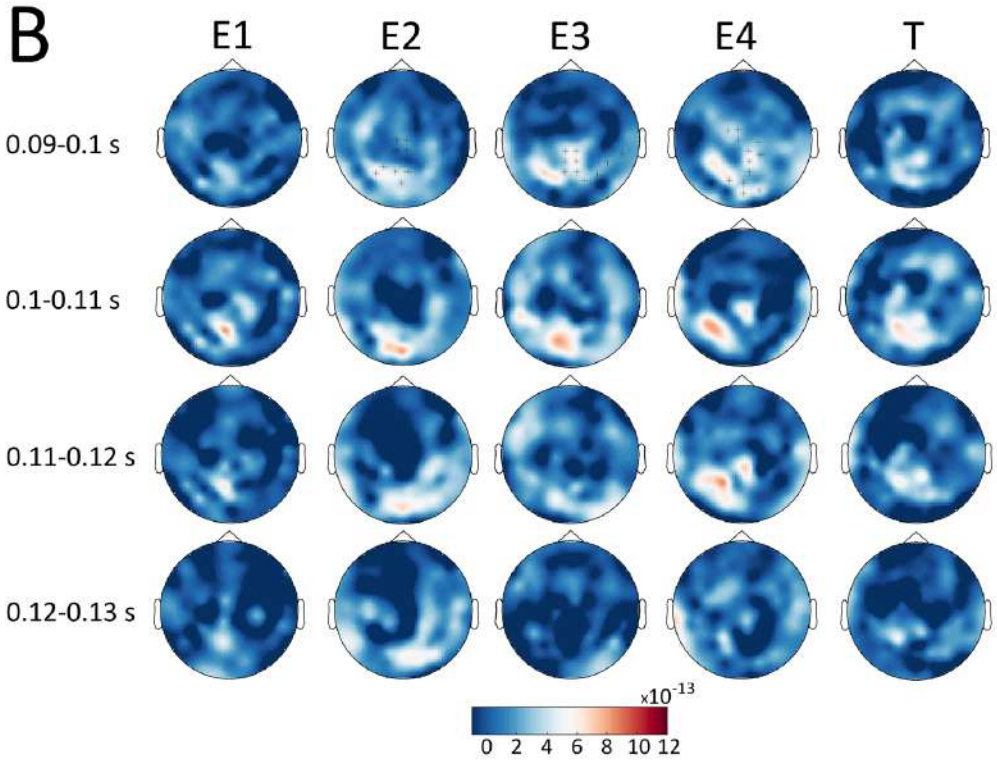


Figure 20: Sensor level ERFs for the when and what+when conditions. A) For each condition (red, when; green, random) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located above occipital regions (MEG02042/3, MEG2032/3, MEG2342/3, MEG2122/3, and MEG1922/3) B) Sensor maps of the ERF difference between the when and what+when conditions in temporal windows around the amplitude peak value.

***Source Level ERFs**:*

We then identified the brain regions underlying the significant effects observed at the sensor space. Source activity was estimated around the peak amplitude of the sensor-level ERFs in the 85–125 ms interval. Whole-brain maps of source activity were created for each condition (*what+when, when, what* and *random*) at Entrainer 4 (E4), i.e., the stimulus where the difference between *what+when* and *what* was statistically significant (Figure 21). We identified the brain regions underlying the significant effects observed in the sensor space. Source activity was estimated around the peak amplitude of the sensor-level ERFs in the time range 85–125 ms interval. Whole-brain maps of source activity were created for all the four conditions (i.e., *what+when, when, what*, and *random*) at Entrainer 4 (E4), i.e., the stimulus where the difference between *what+when* and *what* was statistically significant (shown in figure 19).

Source activity was localized in bilateral occipital regions for all conditions and compared to baseline at the group level. The first local maxima emerged in visual association areas (i.e., Brodmann Area 18: BA 18, average MNI coordinate [–3, –76, –2]) of the left occipital cortex in all the conditions.

For this local maxima, we evaluate the amplitude of source activity across conditions, following the same rationale described for the sensor-level analyses. The brain maps representing the maximum peak activity in the source space are shown in figure 21 A. Figure 21 (B) shows the source amplitudes (also called Neural Activity Index: NAI) with the corresponding standard error. Source amplitude was significantly higher for the *when* than the *what+when* condition ($t = 4.16, p < 0.05$) and for the *random* compared to the *what* condition ($t = 5.20, p < 0.05$) at E4. Crucially, these values were higher for the *what+when* than the *what* condition ($t = 2.38, p < 0.05$), while no difference emerged between the *when* and *random* conditions. Overall, the present results confirm the effects observed at the sensor level, providing a candidate location for the generation of the expectation suppression effects reported at the sensor level.

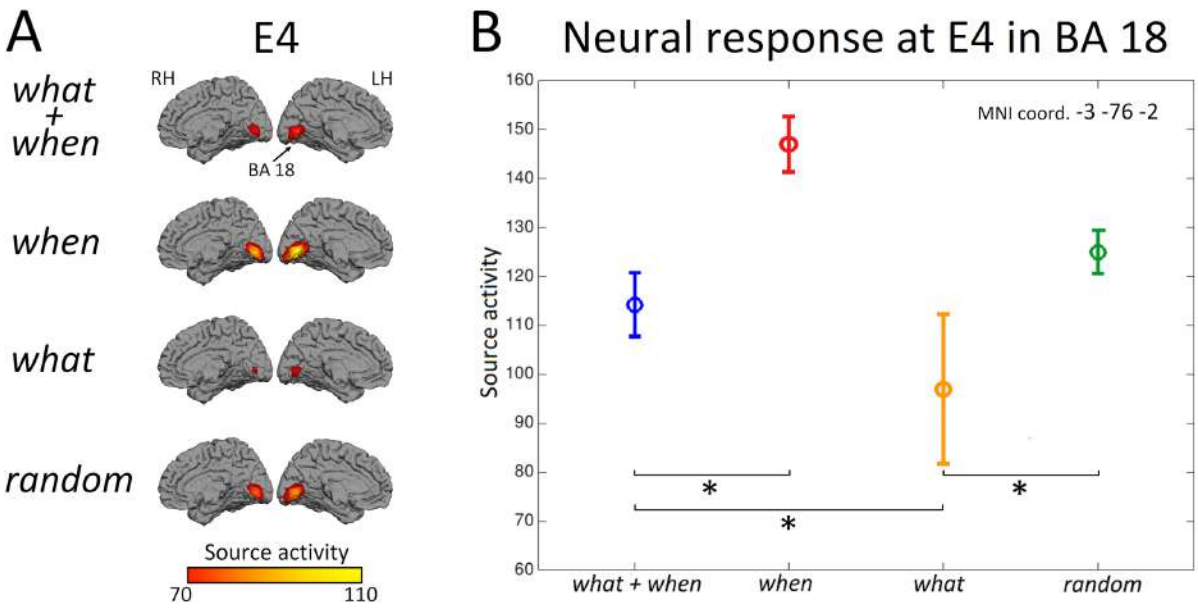


Figure 21. A) Brain maps representing source activity for each condition (what+when, when, what, and random) at Entrainer 4 (E4). We included a view of the medial surface and the occipital lobe of the left (LH) and the right (RH) hemispheres. B) The mean source activity in BA18 (Brodmann Area 18: xyz MNI coordinate: –3, –76, –2) in the four conditions at E4. Asterisks indicate significant differences across conditions.

MVPA: time-resolved decoding:

The ERF analyses showed that the expectation suppression effect grew incrementally larger across the four Entrainers, demonstrating that orientation predictability reduced visual processing costs, possibly due to increased reliance on internally generated expectations. We performed the following analyses to corroborate the hypothesis further that the visual system developed expectations for successive Gabor orientations during the entrainer sequence. We first checked whether the orientation of perceived Gabors could be decoded by applying a time-resolved decoding approach to each entrainer (as a control measure, we also performed the same analysis to decode the spatial frequencies, i.e., CPD of these Gabors). MVPA showed that only those conditions with predictable orientations (*what+when* and *what*) revealed above-chance and statistically significant decoding accuracy values compared to the conditions where orientation was not predictable (see Figures 22 and 23 for the comparison *what+when* vs. *when* and *what* vs. *random*, respectively). Figure 24 shows the decoding accuracy of predictable orientations in conditions with (*what+when*) and without (*what*) temporal predictability. Here, we see how the target orientation becomes increasingly decodable across Entrainers (especially at E3 and E4; *t*-test between peak decoding accuracy at E1 and E4 across participants and conditions, $p < 0.01$) and is most robust at the target. While comparing the decoding results, we showed that decoding accuracy was higher for Entrainer 4 in a later time interval (525–595 ms, $p < 0.05$) for the *what + when* than the *what* condition. By contrast, decoding values for spatial frequencies (high vs. low CPD) were significant only at the target. This effect was expected since the CPD of the target could not be

predicted from the Entrainers, which had intermediate spatial frequencies; this measure provided a good baseline for target orientation decoding effects.

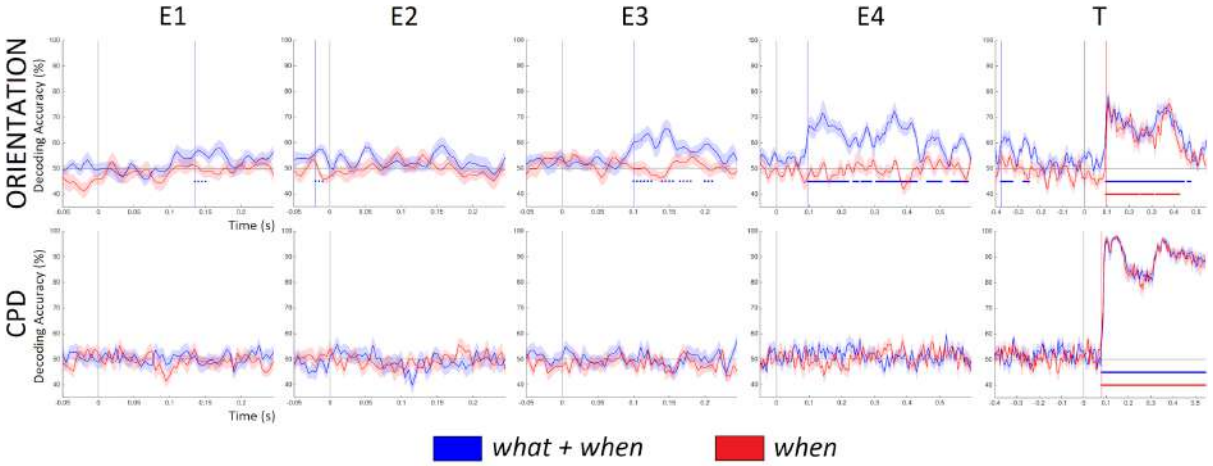


Figure 22. Time-resolved decoding accuracy for the what+when condition (blue line) and when condition (red line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time.

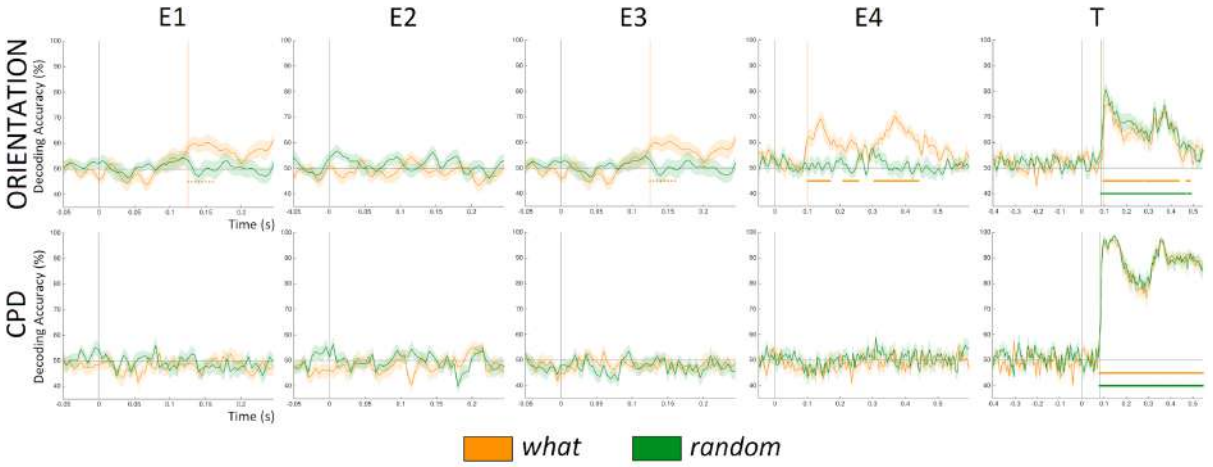


Figure 23. Time-resolved decoding accuracy for the what condition (orange line) and random condition (green line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time.

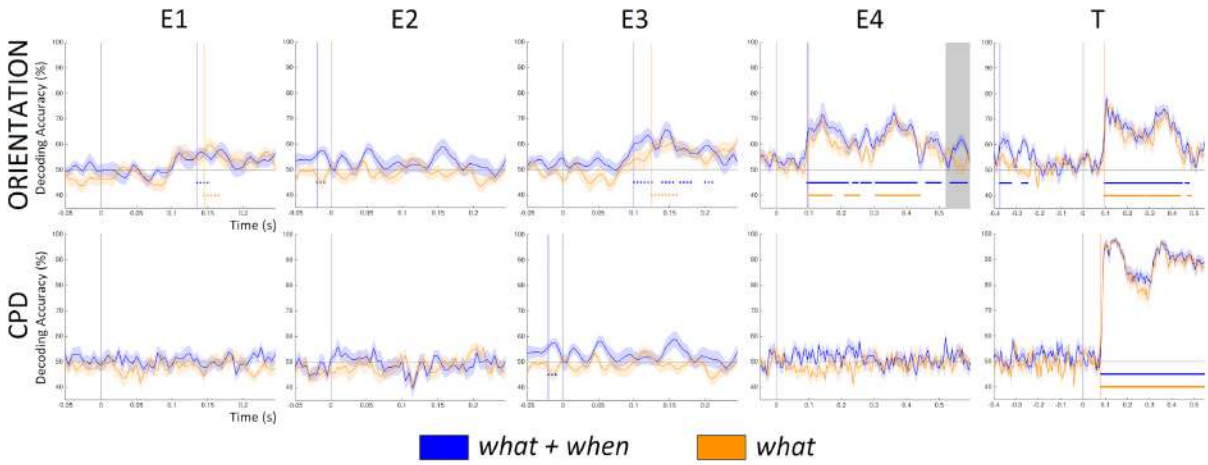


Figure 24. Time-resolved decoding accuracy for the what+when condition (blue line) and what condition (orange line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time. The gray box at E4 in Orientation shows the statistical significant differences ($p < 0.05$) between what+when and what.

3.4 Discussion

1. Expe*ctation Suppression for "what" factor*

We report robust expectation suppression effects for predictive processing of visual Gabors. Across a series of four Entrainers, we observed incrementally stronger suppression of the visual evoked response peaks when Gabor orientations were predictable. This is in line with previous literature for expectation suppression (de Lange et al., 2018; Feuerriegel et al., 2021; Larsson and Smith, 2012). This visual evoked response possibly originated in visual area 2 (V2), which showed reduced activity for predictable stimuli compared to unpredictable stimuli. The source location of the present effect could reflect some top-down activity generated in an extrastriate region projecting to the primary visual cortex (V1)

It is worth noting that expectation suppression effects evident during the entrainer sequence vanished at the presentation of the target Gabor (where participants had to perform the task). The effect was not mirrored in the behavioral responses as well, which probably reflect later decision processes. This strengthens the view that expectation suppression effect in the visual domain is sensitive to the task in which participants are involved (St. John-Saaltink et al., 2015). Within the predictive processing literature, the expectation suppression could be understood as reduction in prediction error (i.e., increase in predictions).

2. Effect of "when" factor

We show that expectation suppression effects were modulated by temporal predictability: expectation suppression of evoked responses was more prominent (larger) when the timing of the Entrainers was jittered/unpredictable. Also, there were very initial differences in the evoked responses to the stimuli (0–60 ms). This differential effect indicates that the neurocognitive system invested fewer resources in visual processing in temporally uncertain scenarios due to some form of precision weighting, i.e., higher reliance on internal predictions.

3. Stimulus specific feature decoding

Our results show that the decoding accuracy for Gabor orientations increased across Entrainers when successive entrainer and target orientations were predictable. This indicates that stimulus predictability is a crucial factor in enhancing the accuracy of orientation decoding during the presentation of Entrainers. In fact, when the stimulus is not predictable, decoding accuracy remains at the chance level. Our results show that while the input stimulus is temporally predictable, the human brain maintains the representation of the stimulus for a longer time compared to unpredictable timing ones. We also show the robustness of our decoding pipeline by analyzing a baseline/control decoding model focused on the CPD features. The classifier was only sensitive to the orientation angles which were manipulated in the Entrainer sequence.

Chapter 4: Expectation suppression in Auditory processing

4.1 Introduction

In this experiment, we aim to investigate the expectation suppression in the auditory domain, and similar to the visual domain how expectation suppression is affected by the context ("*what*") and timing (*"when"*). We inherited the experimental design structure from our visual experiment and used a similar "2-by-2 factors" experimental paradigm. Here, the first manipulation factor was the context/identity of the presented stimuli ("*what*" manipulation), which either provides gradually stronger predictions about the target across Entrainers. The second factor was the timing ("*when*" manipulation), so that participants could either predict the onset of next incoming stimuli. We also maintained the orthogonal relationship between different features (i.e., predictions generated for one feature and task associated with another feature). This allowed us to generate a prediction versus no prediction paradigm, where no violation of predictions was involved. Detailed explanation about the experimental paradigms is provided in the Methods and Material section of this chapter.

The goals of this experiment are the following:

i) How expectation suppression in the auditory domain is affected by "*what*" manipulation? Does this effect mirror the effects in the visual domain or the two modalities deal with expectation suppression in a different manner? Also, it will be interesting to observe the brain areas that are showing this potential suppression in neural responses. Will this suppression be evident in both the hemispheres?

ii) How will the temporal uncertainty (*when* factor) affect the expectation suppression in the auditory domain? Will the predictable timing increase the suppression carried out with predictable context of the target or attenuate the suppression effect? What are the brain areas that will show this suppression?

iii) How does the temporal uncertainty affect the stimulus specific decoding? Can we decode the predictable feature across Entrainers? Can the predictions generated by the Entrainers be decoded in the pre-stimulus time region?

Considering all the goals mentioned above, we used MEG to meet the research goals explained above. Details about MEG modality have been explained in chapter 2.

4.2 Methods and Materials

Participants:

Twenty-one healthy participants were recruited for this experiment. Five participants were excluded from the study due to artifactual data. Remaining sixteen participants (13 females; age range: 19 - 31; M = 26.12; SD = 5.52) took part in our experiment. It is worth mentioning that the participants who took part in this experiment were different from the ones who took part in the visual experiment of this thesis. Participants gave written informed consent and were financially compensated. The participants were recruited from the BCBL Participa website

Experimental procedure:

A series of tones with variable pitch and length was presented. The tones were played to participants using MEG-compatible headphones and the instructions were presented on a screen placed 60 cm from each participant's nasion. The instructions were presented in the centre of the screen on a gray background. Each trial began with a fixation cross (black color) followed by four tones (Entrainer tones), each presented for 200 ms followed by an interstimulus silent interval of 200 ms. After a longer interstimulus interval (600 ms), a fifth tone (Target tone) was presented. The entrainer tones had fixed length (200 ms), while the target could have either a longer length (300 ms) or shorter length (100 ms). Participants were required to indicate if the target tone had a longer or shorter duration compared to the entrainer tones by pressing the correct button.

Four properties of these entrainer sequences were experimentally manipulated (Figure 25): a) the length of the target tone was either long or short; b) the pitch of the target was either high or low; c) the pitch of the target was either predictable based on the pitch of previous entrainer

tones (i.e., increasing or decreasing e.g., pitch of 246.94, 261.63, 293.66, 329.63 Hz for entrainer tones and then target of 349.23 Hz) or unpredictable (a random pitch selection from a set of 12 different pitch levels ranging from low to high (details provided in supplementary materials); e.g., 261.63, 246.94, 329.63, 293.66 for entrainer tones and then target of 349.23 Hz); d) the timing of the interstimulus intervals (blank period) between the four entrainer tones and between the last entrainer (entrainer tone 4) and the target was either predictable (i.e., fixed interstimulus intervals of 200 ms between Entrainers and 600 ms between entrainer 4 and the target) or unpredictable (varying interstimulus intervals ranging between 70-330 ms between Entrainers and 350-850 ms between entrainer 4 and the target).

Depending on the timing and pitch of the entrainer tones and target, trials were divided into four conditions (Figure 1): (i) in the *what+when* condition, both the timing and the pitch of successive Entrainers - and the final target Gabor - were predictable; (ii) in the *when* condition, timing was predictable but pitch were unpredictable. (iii) in *what* condition, successive Entrainers and target pitch were predictable but timing was unpredictable; (iv) in the *random* condition both pitch and timing were unpredictable.

A total of 160 trials were presented in each condition (80 low pitch and 80 high pitch, randomly assigned 80 long and 80 short length) for a total of 640 trials per participant. 80 localizer trials for long and short tones (40 high and 40 low pitch) targets were also acquired while participants simply fixated the centre of the screen. These 640 trials are selected randomly from two pre-defined presentation sequences.

On each trial, participants had to indicate whether the target tone had a longer or a shorter length than the preceding entrainer tones. Participants responded by pressing a button with their left or right hand, with the response hand counterbalanced across participants. To keep the participants relaxed, a longer mandatory break took place every 60 trials (the MEG researcher pressed a button from the operating console to pause and restart the presentation).

Also, a short optional break (participants pressed the button when they were ready to continue) was available every 12 trials.

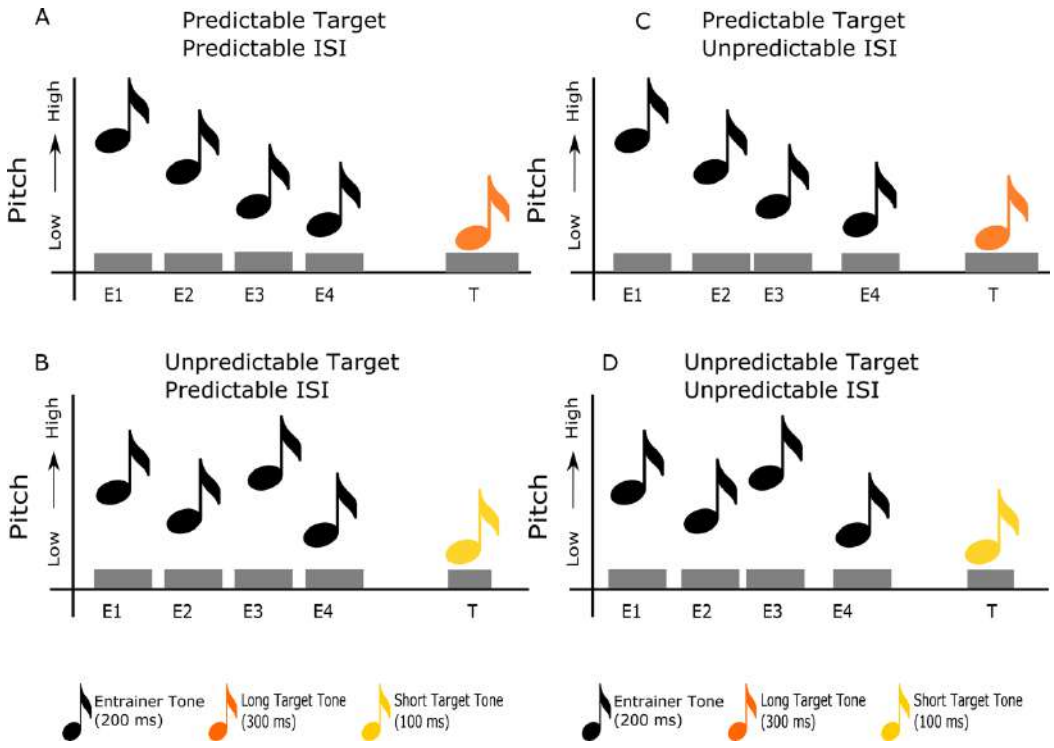


Figure 25: Experimental design. A) Target pitch and timing of upcoming entrainer tone and target tone are predictable (what+when condition). B) Pitch of upcoming entrainer tone and target tone are not predictable but the timings are predictable (when condition). C) Target pitch of upcoming entrainer tone and target tone are predictable but the timing is not predictable (what condition). D) Target tone and timing of upcoming entrainer tone and target tone are not predictable (random condition). Abbreviations: E1 – Entrainer 1, E2 – Entrainer 2, E3 – Entrainer 3, E4 – Entrainer 4, T - TargetISI – Inter Stimulus Interval.

Methods:

The methods used in the auditory experiment of this thesis followed the same methods used in the visual experiment.

Data acquisition and preprocessing:

MEG data were acquired in a magnetically shielded room using the whole-head MEG system installed at the BCBL. The data acquisition parameters and procedures such as head movement tracking using HPI coils, fiducial point markings, digitization of head shape, band-pass filters, EOG and ECG collection etc. were the same as explained in the visual experiment. Continuous MEG data were preprocessed using the temporal Signal- Space-Separation (SSS) method (Taulu and Simola, 2006) to suppress external interference noise. Subsequent data analyses were performed using Matlab R2014b (Mathworks, Natick, MA, USA).

Reaction times:

Participants were asked to indicate if the length of the Target tone was longer or shorter compared to the previous Entrainers tones. We made sure that accuracy was high and comparable across conditions. The response time (RT) was calculated for all the four conditions (i.e., *what + when, when, what and random*). The response times longer than 1500 ms were considered as outliers and were removed from the analysis. The mean reaction time and standard deviation is computed for each experimental condition.

Sensor level ERFs:

MEG trials corrected for jump, muscle artifacts, EOG and ECG artifacts were identified using the same pipeline which was used in the visual experiment. In this experiment, the EOG and ECG components were identified manually and later linearly subtracted from the MEG data. On average, two components were removed per participant. The artifact-free data were bandpass filtered between 0.5 and 45 Hz. Trials were segmented time-locked to the presentation of each of the entrainer tones (Entrainers 1, 2, 3, and 4) and the target tone. The

trial onsets were computed from the MISC channel which recorded the tones' envelope online. The trial segments were grouped together for each entrainer and target, and then averaged to compute the ERFs. Baseline correction was also applied to the evoked data from baseline time period which was 1000 ms before the fixation cross (i.e., at the beginning of a trial).

We used a similar approach to the visual experiment by performing the analysis of the sensor level data to establish expectation suppression and its interactions with different kinds of predictability. We applied an ANOVA to sensor-level data to explore the influence of our experimental factors on visual ERFs. We extracted ERF amplitudes in the set of ten sensors (five on each hemisphere) from sensors close to temporal brain regions. We then selected the time window classically associated with the initial auditory evoked response (75–125 ms post stimulus). A three-way repeated measures ANOVA was computed in JASP (JASP Team, 2020) with these amplitude values as dependent variables and the following factors: entrainer (four levels; corresponding to Entrainers 1, 2, 3, 4); *what* (two levels; predictable/unpredictable entrainer and target tone pitch); and *when* (two levels; predictable/unpredictable timing of Entrainers and target tone)[1]. Significant interactions (specifically, the triple interaction "entrainer * *what* * *when*") were further investigated through theoretically relevant pairwise comparisons. Pairwise comparisons between conditions were performed using a cluster-based permutation test (Maris and Oostenveld, 2007). T-values were computed for each sensor (combined gradiometers) and each time point during the 0–270 ms time window, and were clustered if they had t-values that exceeded a t-value corresponding to the 99.99th percentile of Students t-distribution, i.e., a two-tailed *t*-test at an cluster-alpha of 0.05, and were both spatially and temporally adjacent. Cluster members were required to have at least two

[1] I also computed a 4-factor ANOVA by adding the hemisphere as a fourth factor (two levels: right and left hemisphere), I did not observe interaction of the hemisphere factor with other factors.

neighboring channels that also exceeded the threshold to be considered a cluster. The sum of the t-statistics in a sensor cluster was then used as the cluster-level statistic, which was then tested by permuting the condition labels 1000 times.

Similar to the visual experiment, four different comparisons were carried out. In the first comparison, we contrasted ERFs for the *when* and the *what+when* conditions. This comparison evaluated the effect of pitch predictability when the timing of the entrainer tones and target were predictable. In the second comparison, we compared ERFs for the *random* and *what* conditions. This comparison evaluated the effect of pitch predictability when timing was unpredictable. These two comparisons mainly focused on the main effect of target tone's pitch predictability (i.e., the *what* manipulation) revealed by expectation suppression. We then compared the ERFs for the *what+when* and *what* conditions. Here we directly contrasted these two predictable pitch conditions to evaluate the effect of temporal predictability on stimulus predictability. The final comparison contrasted ERFs in the *when* and *random* conditions. This comparison was performed to analyze the effect of temporal predictability in the absence of pitch predictability.

Source level ERFs:

We used the same pipeline for performing the source reconstruction of significant effect at sensor space. First MEG-MRI co-registration was performed using MRIlab software, then Individual T1-weighted MRI images were segmented into different tissue types using Freesurfer software (Martinos Center of Biomedical Imaging, MQ) (Dale et al., 1999). The source space was defined as a regular 3D grid with a resolution of 5 mm and the lead fields were defined using a realistic single-shell Boundary element model (BEM) head model. Both the planar gradiometers sensors and magnetometers sensors were used for inverse modelling.

The covariance matrix was estimated combining the information embedded in the pre-stimulus time range (i.e., -400–0 ms prior to the cross fixation) and the post-stimulus time range (i.e., 0–400 ms after the presentation of Gabor). Whole brain source activity was estimated using linearly constrained minimum variance (LCMV) beamformer approach (Veen et al., 1997). LCMV beamformer was computed on the evoked data in the time period and brain maps containing neural activity index were transformed and normalized similarly to what has been explained for the visual experiment.

The mricron software was used for visualizing the brain maps in the MNI space. For each condition at Entrainer 3 (see Results below for the motivation in selecting this Entrainer), we obtained the NAI value and the MNI coordinates of local maxima (sets of contiguous voxels displaying higher power than all other neighboring voxels) (Bourguignon et al., 2018). NAI values were compared between conditions (e.g., *when* vs. *what + when, random* vs. *what, what vs. what + when* and *when vs. random*) using t-tests and repeated measures ANOVA.

MVPA: time-resolved decoding:

Time-resolved within-subjects multivariate pattern analysis was performed to decode the feature specificity (i.e., the pitch and length of the tone) from MEG data. Similar to visual experiment, data at E1, E2 and E3, were segmented from 50 ms prior to 250 ms after the onset of the Entrainers. For E4 data were segmented from 50 ms prior to 600 ms after the onset of the Entrainer. The artifact free clean data was segmented from -400 ms to 550 ms time locked to the Target. The data were classified using a linear support vector machine (SVM) classifier with L2 regularization. We performed a binary classification of the pitch of each tone depending on the pitch of the subsequent Target. In other words, the class labels (i.e., high pitch or low pitch) were derived from the Target pitch only: if the Target pitch was high all the

preceding pitch in the corresponding conditions were labelled as high pitch class and vice-versa. Same approach was used for labelling the classes for tone length feature decoding.

The data were down-sampled by a factor of five (i.e., new sampling frequency 200 Hz) prior to the classification. Pseudo trials were generated to improve the SNR by creating trials' bins, resulting in a set of 10 trials for each bin (Dima and Singh, 2018). This pseudo trial generation was repeated 100 times to generate trials with a higher signal to noise ratio. 5-fold cross validation was used and the classification process was repeated 25 times (details provided in visual experiment). The data were then randomly partitioned using 5-fold cross-validation. To improve data quality, we also performed multivariate noise normalization (Guggenmos et al., 2018). The noise normalization and cluster-based permutation test were computed similarly to the visual experiment.

4.3 Results

Reaction Time:

what + *when* condition showed the fastest response time (888.1 ± 236 ms) while the *random* condition was slowest (910.9 ± 242 ms). In the *when* and the *what* conditions participants responded respectively after (894.5 ± 286) ms and (882.7 ± 233) ms. We fit a Linear mixed model (LMER) considering subjects and observations as *random* effects, and the fixed effects of *what* (orientation predictable or not), *when* (timing predictable or not) and their interaction as fixed effects. We observed an effect of *when* ($p<0.05$).

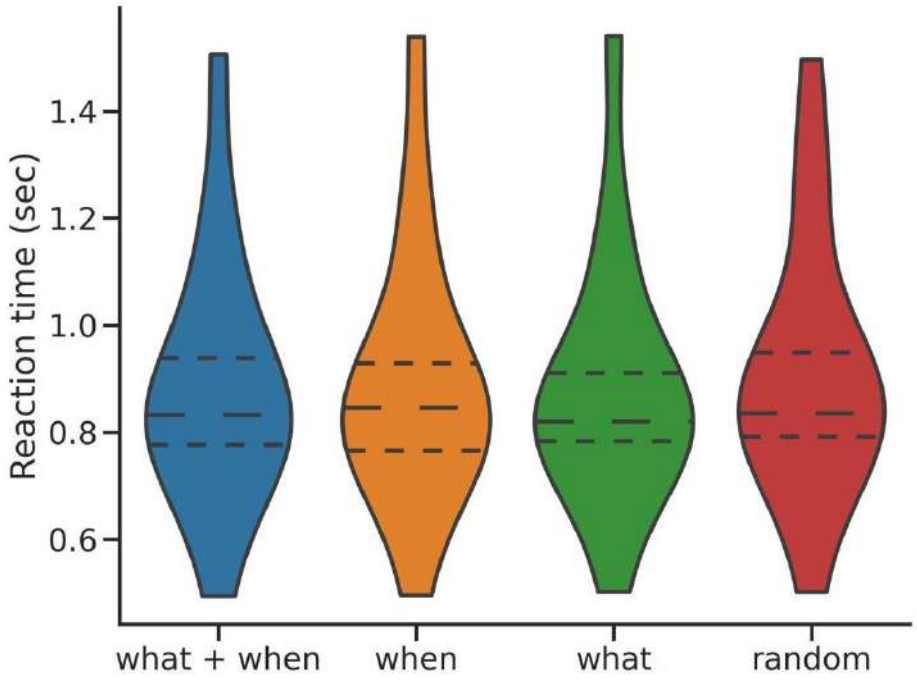


Figure 26: Reaction time distribution of duration task across conditions

Sensor level ERFs:

Similar to visual domain, we analysed the amplitude of the initial auditory evoked response for 10 temporal sensors (5 on each hemisphere) to determine how the two-by-two experimental design modulated auditory responses across Entrainers. In the three-way ANOVA (details in Table 2), we observed a significant main effect of entrainer ($p < 0.001$) on the peak amplitudes of ERFs. Interestingly, this factor interacted with the factor *what* ($p < 0.001$), suggesting that target pitch predictability affected auditory-evoked responses differently across Entrainers. We should point out that a main effect of *what* ($p < 0.001$) supported the observation that orientation predictability influenced auditory processing. The *when* factor did not affect the auditory processing but a two-way interaction of entrainer**when* was significant. Importantly, the interaction between the three factors, i.e., entrainers, *what* and *when*, was significant ($p = 0.017$). This triple interaction underlines the fact that timing uncertainty influenced the development of auditory predictions across the sequence of four entrainer tones.

Figure 27 shows the sensor-level Event Related Fields (ERFs) time-locked to the onset of each Entrainer tone (i.e., E1, E2, E3 and E4) and Target tone (T) for the *when* and the *what* + *when* conditions. The amplitude of the ERFs was significantly (cluster *p*-value < 0.01) higher for the *when* compared to the *what* + *when* condition for E2, E3 and Target, but not for E1 or E4. The amplitude enhancement for the *when* compared to the *what* + *when* condition emerged within the [88 – 125] ms, [92 – 125] ms and [83 – 108] ms time intervals for the E2, E3 and Target respectively. This effect was located in temporal sensors for all the Entrainers. Figure 28 shows the sensor-level ERFs for the *random* and the *what* conditions. The amplitude of the ERFs was significantly (cluster *p*-value < 0.01) higher for the *random* compared to the *what* condition for E2, E3 and E4 and Target, but not for E1. The amplitude enhancement for the *random* compared to the *what* condition emerged within the [77 – 125] ms, [81 – 122] ms, [103 – 121] and [79 – 108] ms time intervals for the E2, E3, Target respectively. This effect was located in temporal sensors, sensors close to motor and pre-motor sensor areas for the Entrainers. It is worth noting that the comparison having temporal predictability (i.e., *when* vs *what* + *when*) involves a larger number of sensors compared to the comparison lacking temporal predictability.

Since both the comparisons are significant from Entrainer 2 onward, we compared the orientation predictability with (*what* + *when*) and without temporal predictability (*what*). Figure 29 shows that the initial very early evoked activity (0–60 ms) at E1 is similar for both conditions. As we move across Entrainers such early difference is present at Entrainer 2 but then it vanishes for the rest of the Entrainers. Since our focus in this analysis is in the early evoked response to the auditory stimulus showing expectation suppression effects, the time window for statistical comparison was selected from 75 to 125 ms. The amplitude enhancement for the *what* compared to the *what* + *when* condition emerged within the [93- 108] ms time interval only at E2. This effect was located close to pre-motor area sensors. Importantly, this

effect is already evident since time 0 and could emerge from some ongoing difference preceding E2. The effect flips out at E3 in polarity and *what* + *when* has lesser amplitude compared to the *what* condition. I consider this difference genuinely driven by the temporal uncertainty interfering with the expectation suppression effect driven by pitch predictability. The effect emerged within [99 - 125] ms and was not significant across the subsequent Entrainers and Target. Figure 30 shows the sensor level results of the ERFs for the *when* and the *random* conditions. This last comparison also shows the difference in the peak amplitudes at E2 which emerged at [111 – 125] ms.

Table 2: Repeated measure ANOVA with the factors entrainer (four levels, one for each entrainer), *what* (two levels: target pitch predictable or not) and *when* (two levels: timing predictable or not).

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Entrainers	1.308e -22	3	4.360e -23	15.691	< .001
when	1.138e -25	1	1.138e -25	1.393	0.256
what	4.661e -24	1	4.661e -24	19.135	< .001
Entrainers ∗ when	1.607e -24	3	5.357e -25	4.995	0.004
Entrainers ∗ what	3.711e -24	3	1.237e -24	6.677	< .001
when ∗ what	1.300e -26	1	1.300e -26	0.110	0.745
Entrainers ∗ when ∗ what	9.682e -25	3	3.227e -25	3.784	0.017

Note. Type III Sum of Squares

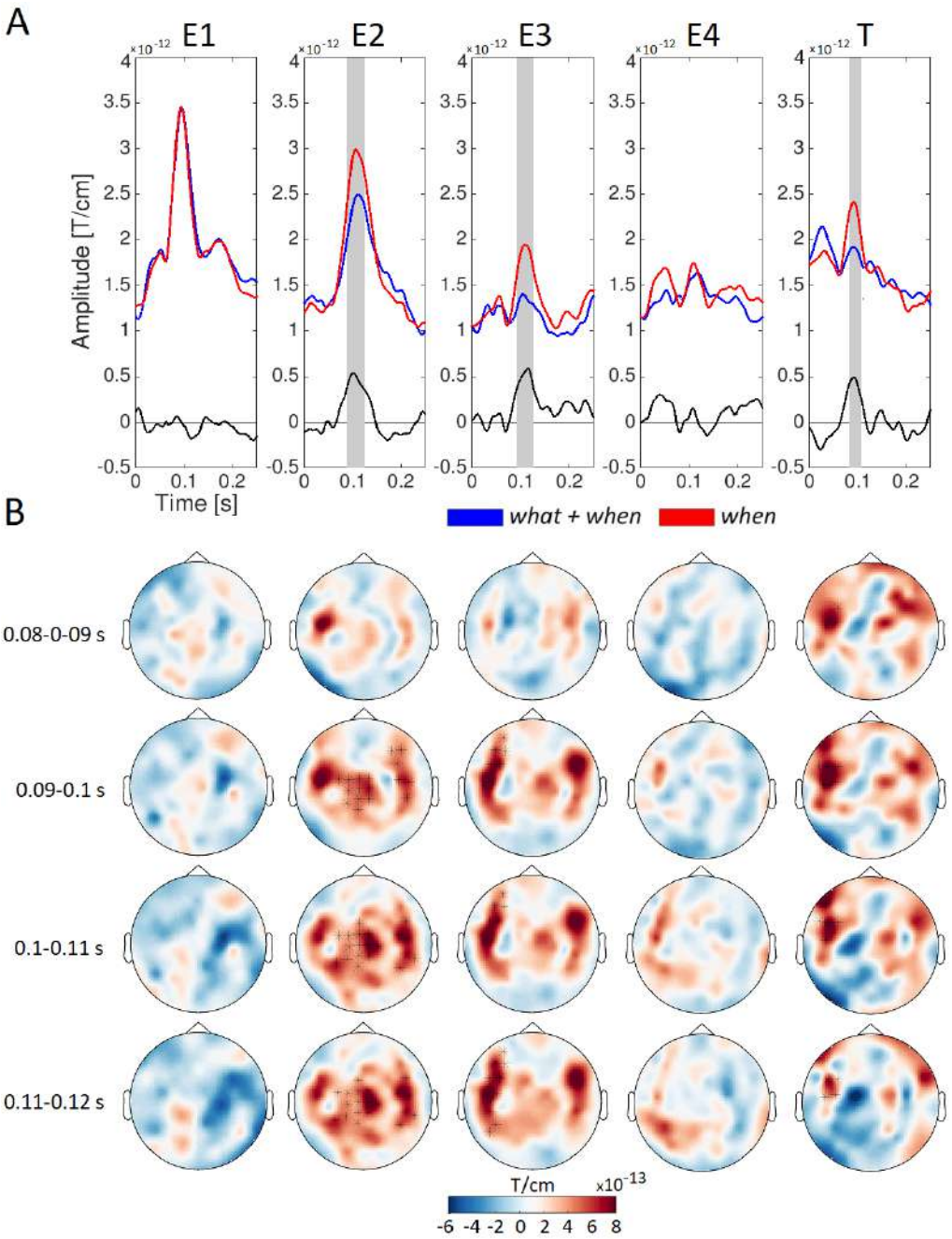


Figure 27: Sensor level ERFs for the when and what+when conditions. A) For each condition (red, when; blue, what+when) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located in bilateral auditory regions ('MEG0222+0223', 'MEG0232+0233', 'MEG0242+0243', 'MEG1312+1313', 'MEG1322+1323', 'MEG1342+1343', 'MEG1612+1613', 'MEG1622+1623', 'MEG2412+2413', 'MEG2422+2423') . Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.05$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the when and what+when conditions in the temporal window around the amplitude peak value. Sensors showing significant differences ($p < 0.05$) are highlighted.

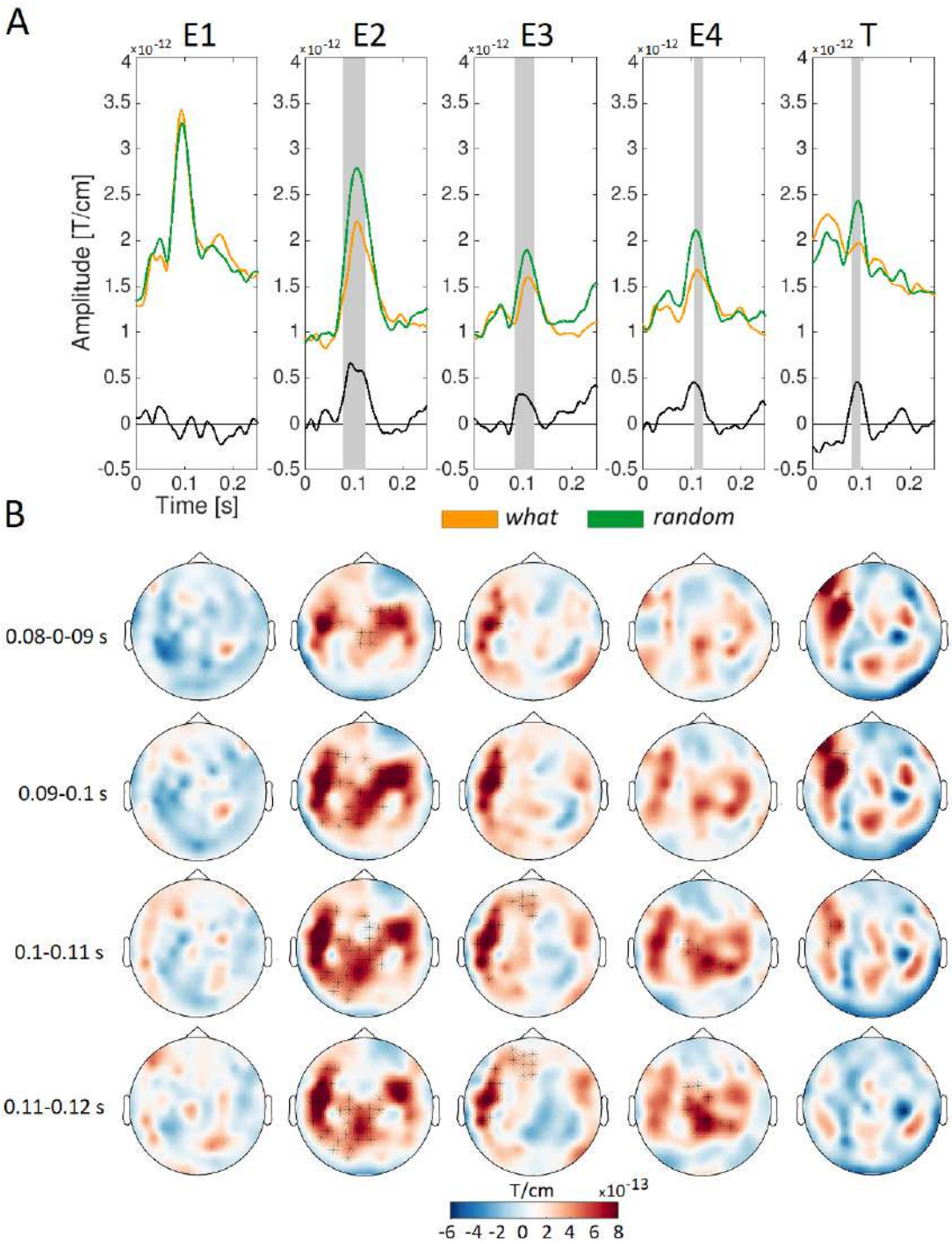


Figure 28: Sensor level ERFs for the when and what+when conditions. A) For each condition (orange, what; green, random) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located in bilateral auditory regions ('MEG0222+0223', 'MEG0232+0233', 'MEG0242+0243', 'MEG1312+1313', 'MEG1322+1323', 'MEG1342+1343', 'MEG1612+1613', 'MEG1622+1623', 'MEG2412+2413', 'MEG2422+2423') . Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.05$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the when and what+when conditions in the temporal window around the amplitude peak value. Sensors showing significant differences ($p < 0.05$) are highlighted.

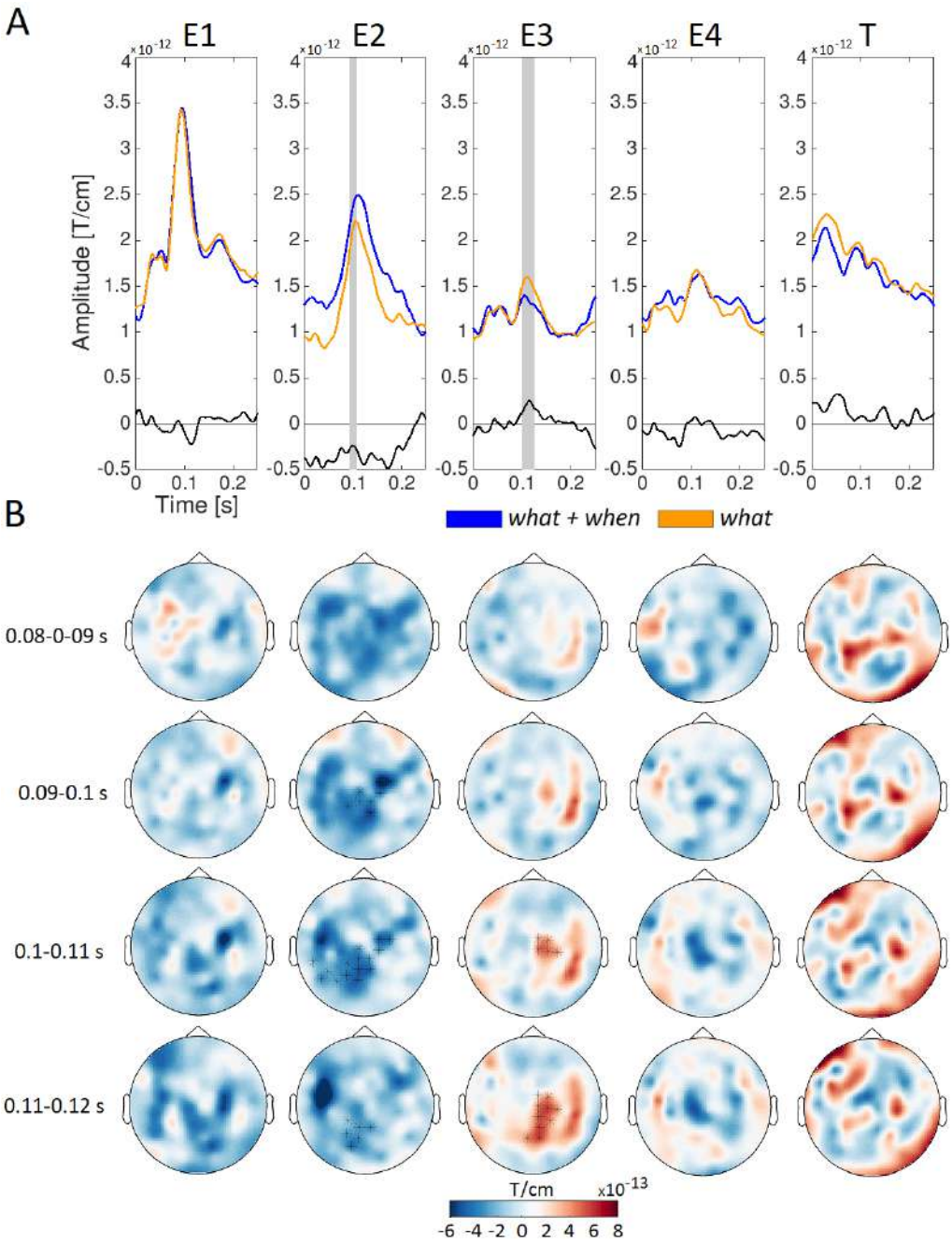


Figure 29: Sensor level ERFs for the when and what+when conditions. A) For each condition (blue, what+when; orange, what) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located in bilateral auditory regions ('MEG0222+0223', 'MEG0232+0233', 'MEG0242+0243', 'MEG1312+1313', 'MEG1322+1323', 'MEG1342+1343', 'MEG1612+1613', 'MEG1622+1623', 'MEG2412+2413', 'MEG2422+2423') . Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.05$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the when and what+when conditions in the temporal window around the amplitude peak value. Sensors showing significant differences ($p < 0.05$) are highlighted.

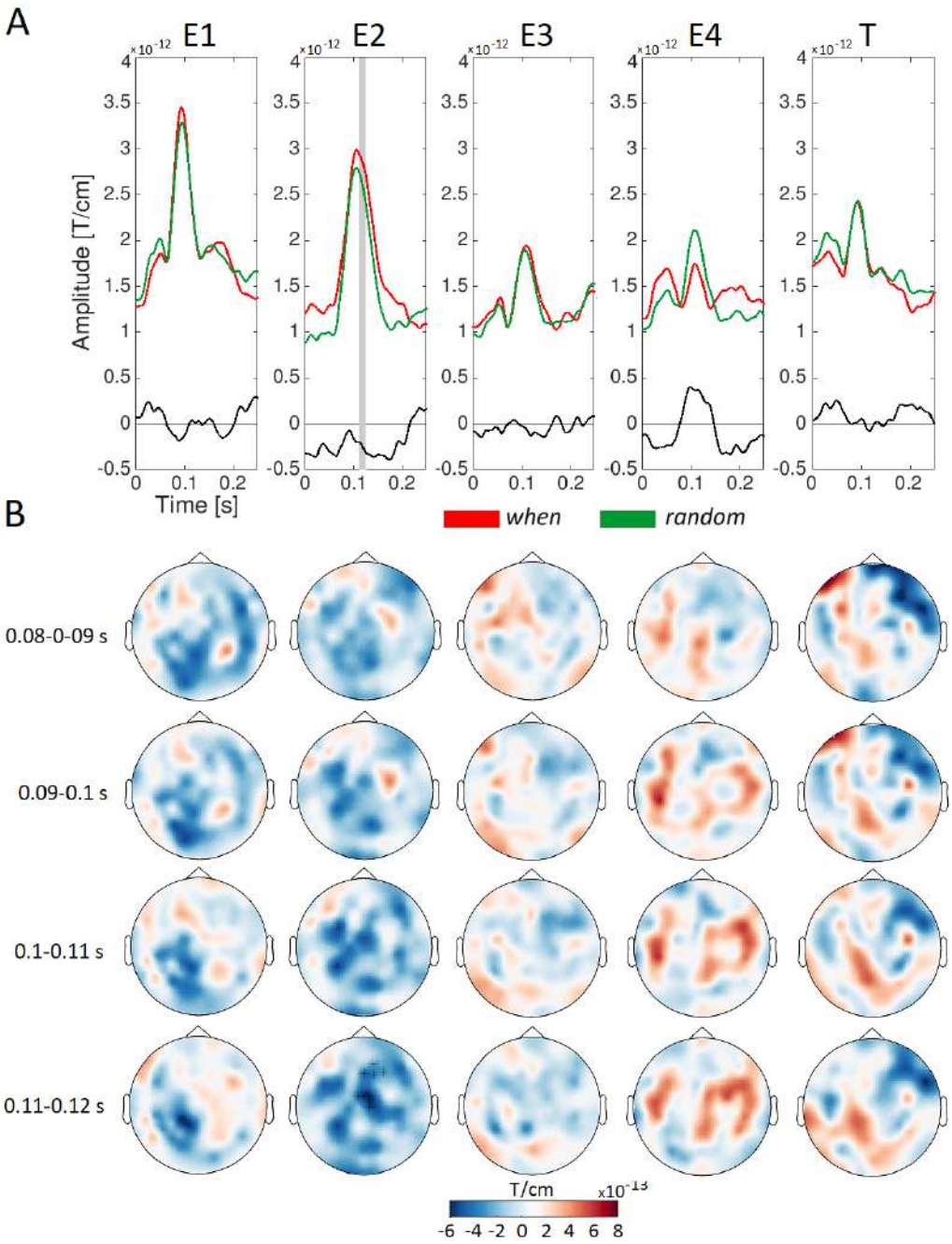


Figure 30: Sensor level ERFs for the when and what+when conditions. A) For each condition (orange, what; green, random) and stimulus (Entrainer 1 [E1], E2, E3, E4 and Target [T]), we show the average of the event related fields (ERFs) in representative channels located in bilateral auditory regions ('MEG0222+0223', 'MEG0232+0233', 'MEG0242+0243', 'MEG1312+1313', 'MEG1322+1323', 'MEG1342+1343', 'MEG1612+1613', 'MEG1622+1623', 'MEG2412+2413', 'MEG2422+2423') . Gray boxes represent time points where the amplitude of the ERFs was higher ($p < 0.05$) for the when than the what+when condition. B) Sensor maps of the ERF difference between the when and what+when conditions in the temporal window around the amplitude peak value. Sensors showing significant differences ($p < 0.05$) are highlighted.

Source Level ERFs:

Data from 15 participants were used in the source reconstruction (excluding a participant for which we did not have the T1 structural image). We identified the brain regions underlying the significant effects observed at the sensor space. Source activity was estimated around the peak amplitude of the sensor-level ERFs in the 75–125 ms interval. Whole-brain maps of source activity were created for each condition (*what+when, when, what* and *random*) at entrainer 3 (E3), i.e., the stimulus where the difference between *what+when* and *what* was statistically significant and theoretically more relevant.

Source activity was localized in bilateral temporal and parietal regions for all conditions and compared to baseline at the group level. Two local maxima located in both the hemispheres were selected for further analysis. The first local maxima emerged in close vicinity to the left primary auditory cortex (average MNI coordinate [–45, –24, 14]).

For this local maxima, we evaluate the amplitude of source activity across conditions, following the same rationale described for the sensor-level analyses. The brain maps representing the maximum peak activity in the source space are shown in figure 31 A. Figure 31B shows the source amplitudes with the corresponding standard error. Source amplitude was significantly higher for the *when* than the *what+when* condition (t = -1.774) and for the *random* compared to the *what* condition (t = -1.120) at E3. In this local maxima, no effects emerged between *what* compared to *what + when* and *random* compared to *when*.

The second local maxima emerged in the right primary auditory cortex (average MNI coordinate [49, –16, 12]). For this local maxima also, we evaluate the amplitude of source activity across conditions, following the same rationale described for the sensor-level analyses..

Figure 31 B shows the source amplitudes with the corresponding

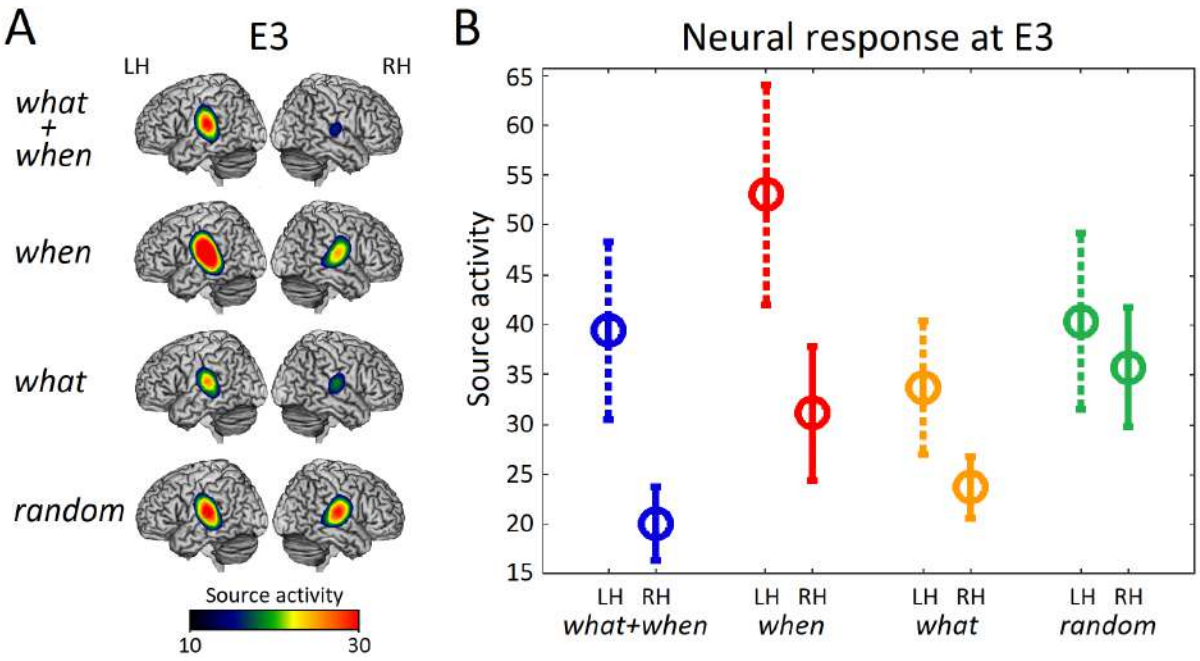


Figure 31 A. Brain maps representing source activity for each condition (what+when, when, what, and random) at Entrainer 3 (E3). We included a view of the medial surface and the occipital lobe of the left (LH) and the right (RH) hemispheres. B) The mean source activity MNI -45 -24 14 (LH) and 49 -16 12 (RH) in the four conditions at E3.

standard error. Source amplitude was significantly higher for the *when* than the *what+when* condition (t = -1.694) and for the *random* compared to the *what* condition (t = -2.781) at E3. In this local maxima, there is a significant effect emerging between what compared to *what + when* (t= -1.276), consistent with the effect observed at the sensor level.

MVPA: time-resolved decoding:

The ERF analyses showed that the expectation suppression effects in the auditory domain are much stronger compared to the visual domain. We performed the following analyses to further corroborate the hypothesis that the auditory system developed expectations for successive tone pitches during the entrainer tone sequence, given the fact that the neural responses are largely suppressed. We first checked whether the pitch of perceived tones could be decoded (as a control measure, we also performed the same analysis to decode tone length, i.e., CPD). MVPA showed that the decodable information was only available at the target tone only in both the

features. The classifiers were not able to pick the representational information of the pitch stimulus while the potential auditory expectations were building up during the entrainers' sequence.

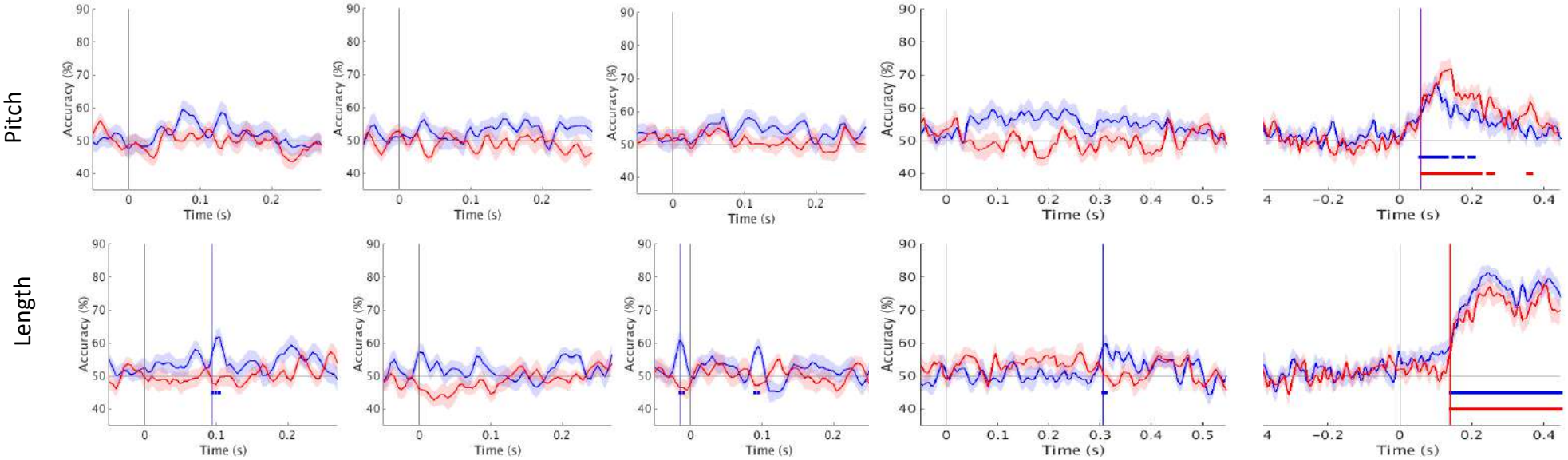


Figure 32. Time-resolved decoding accuracy for the what+when condition (blue line) and what condition (orange line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time.

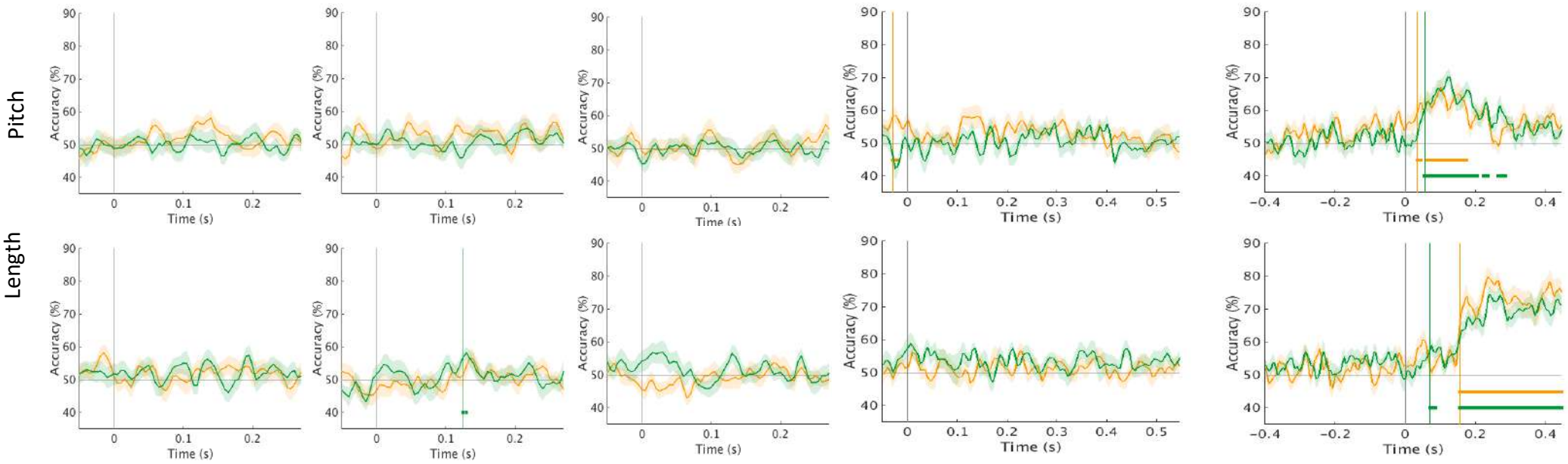


Figure 33. Time-resolved decoding accuracy for the what condition (orange line) and random condition (green line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time.

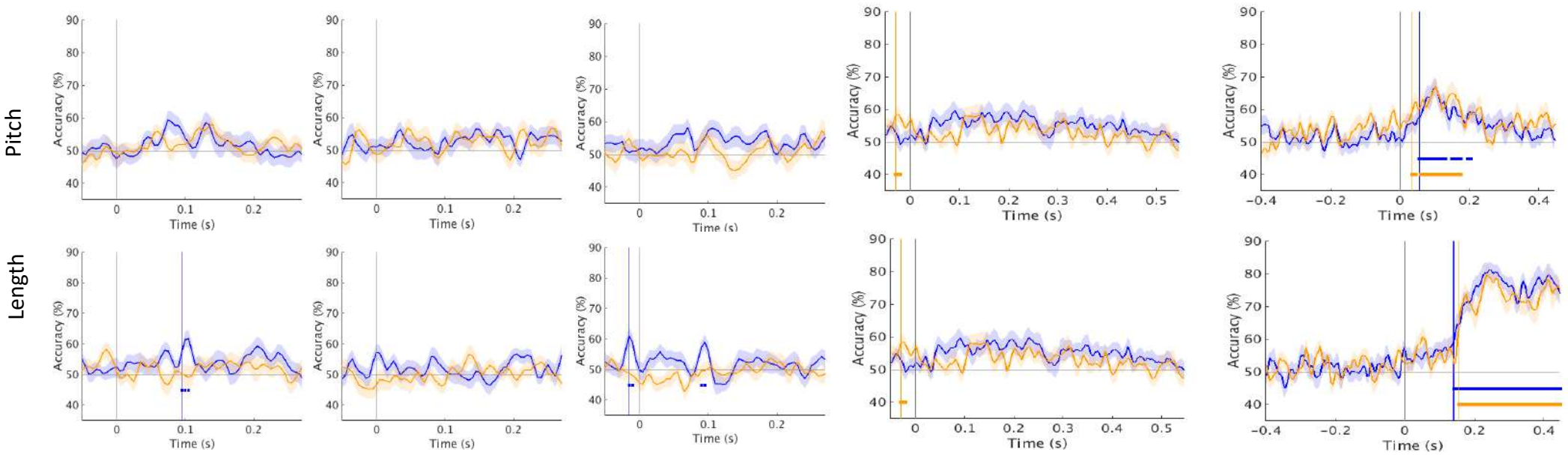


Figure 34. Time-resolved decoding accuracy for the what+when condition (blue line) and what condition (orange line) time-locked to Entrainer 1 (E1), E2, E3, E4 and Target (T). The coloured dots under the curves indicate the statistical significance of decoding accuracy across time.

4.4 Discussion

1. *Expectation Suppression for "what" factor*

Across a series of four entrainer tones, we observed extensive suppression of the auditory evoked response peaks, independently of the experimental condition: at E2, E3 and at E4, the suppression was so strong that we observed almost a floor effect in the last entrainer (E4) where the auditory response was in some conditions hard to discern.

Importantly, we report expectation suppression effects for predictive processing of auditory tones. Across the entrainer tones, we observed stronger suppression of auditory evoked response peaks when the pitch of the next tone was predictable.

This auditory evoked response originated in the left and right primary auditory cortex, showing reduced activity for predictable stimuli compared to unpredictable stimuli. The source location of the present effect is limited to the primary auditory cortex which is usually considered a low-level brain area in cortical auditory processing hierarchy. This could be mainly due to the fact that we used pure tones, whose processing does not require the involvement of higher auditory associative regions.

It is worth noting that expectation suppression effects evident during the entrainer sequence were also present at the presentation of the target tone (where participants had to perform the task). This is an interesting effect: When the prior information about the target tone pitch was available (in the *what+when* and *what* conditions), the neural response to the tone is largely suppressed (it was almost absent). In the other case, when no predictable prior pitch information was present (in the *when* and *random* conditions), the neural responses did not get so strongly suppressed. This shows that the predictions developed during the entrainer sequence were largely affecting the target processing even if participants were focusing on a different feature (tones' duration).

2. Effect of "when" factor

We also report that the expectation suppression effects for predictive processing of entrainer tones are not only sensitive to the frequency/spectral content of the tones but also to the timing at which the tones were presented. In fact, expectation suppression effects were highly modulated by temporal predictability: The *when* factor is increasing the suppression of evoked response given by the *what* factor (figure 29, *what* > *what* + *when* at E3). The neural signals at the second entrainer show a difference in peak response based on both the *when* and the *what* factor, as well as the differences in very initial evoked response to stimuli (0-50 ms). Importantly though, I am theoretically mainly interested in the interaction of *what* and *when* factors. To investigate further, I performed a two-way ANOVA (*what* and *when* factor) on peak evoked amplitudes extracted from channels close to temporal brain regions at both E2 and E3. The interaction of *what*when* factor is missing at E2 (p = 0.545), and is present at E3 (p = 0.017). So, I considered the effect at E3 theoretically more relevant, and selected this effect for source reconstruction.

The sources of evoked responses appear to arise from the primary auditory cortex in both the hemispheres. My experimental findings, first of all, recall the existing literature on auditory and speech processing. The 'asymmetric sampling in time model of speech processing' (AST model, (Poeppel, 2003)) suggests that the left auditory areas preferentially extract information from short (~20–40 ms) temporal integration windows. The right hemisphere homologs preferentially extract information from longer (~150–250 ms) integration windows like temporal expectations. In my experiment the left-hemisphere regions show more activation compared to the right hemisphere. Higher responses (independently of condition) on the left hemisphere indicate that in this paradigm (focused on pitch tones), there is larger involvement of the left hemisphere probably due to its specialization on shorter integration windows. The

effect of the *when* factor only emerges in the right hemisphere probably because the right hemisphere auditory regions, operating on longer temporal integration windows, can capture the temporal jittering across tones (in line to AST model of speech, Poeppel, 2003).

3. Stimulus specific feature decoding

Decoding accuracy for entrainer tones was at chance level, thus indicating that pitch specificity could not be identified from the neural responses during the presentation of the entrainers. This is a surprising finding because the perception of the incoming sensory data that match previously generated expectations is facilitated (more neural suppression, better perception). This neural facilitation is expected in the predictive processing framework: The brain does not directly pass the information about sensory input forward, but rather computes the difference between expectations and sensory input. The prediction errors provide the informative (decodable) neural representation of the sensory input mainly if these expectations are "uninformative" and the sensory input is sufficiently precise (in other words, if prediction error is large). In conditions that match prior (*what* and *when*) expectations, increasing the sensory expectation reduces the informativeness of multivariate neural representations (see also (Blank and Davis, 2016)).

At the target, I observed that the task related feature (length) has higher and long-lived decoding accuracy (figure 32) compared to the task unrelated/prediction feature (pitch). However, the conditions involving *what* and *when* expectations about the target show more neural suppression but less decoding accuracy compared to the condition lacking such expectations. This strengthens my claim that higher expectation suppression of the neural response does not yield higher decodable multivariate representational information.

Finally, it could be noted that evoked responses across entrainers are incrementally suppressed becoming in some conditions not recognizable. This suppression could also contribute to the

fact that such low evoked auditory responses do not have enough signal-to-noise ratio to boost the decodability of the eliciting stimulus pitch.

As a separate note, current Predictive Processing theories (Friston, 2005; Rao and Ballard, 1999) suggest that cortical regions involved in sensory processing contain two subpopulations of neurons: a) prediction error units that represent the unexpected part of the incoming sensory information and b) prediction units that represent the expected part of the incoming sensory information. My results could indicate that the neural representation of a stimulus in the human brain is mainly driven by prediction error units which have stronger magnitude compared to the weaker prediction units. Since I have used a population coding neuroimaging method (MEG), it is not possible to separate the activity from two subpopulations It could be that other neural measures, such as neurophysiological recordings with depth electrodes or laminar-specific ultra-high field strength fMRI, are better able to detect responses from different prediction units and could provide clearer evidence of laminar-specific representations of predictions and prediction errors.

Chapter 5: General Discussion

The relevance of the present experimental paradigm

The goal of these experiments was to evaluate how visual and auditory expectations differentially modulate sensory perception due to variable temporal and stimulus predictability. I would like to first remark how my experimental design for studying predicting *what* and *when* is different and has critical advantages compared to the previous literature. In both the experiments, I mainly focused on the expectation suppression effect (Walsh et al., 2020) and, in contrast to previous studies focused on repetition suppression (Auksztulewicz and Friston, 2016; Costa-Faidella et al., 2011; Stefanics et al., 2018; Utzerath et al., 2017) and expectation suppression (Auksztulewicz et al., 2018; Han et al., 2019; Kok et al., 2012a; St. John-Saaltink et al., 2015), (i) I did not use mismatching stimuli, and (ii) participants were not aware of the experimental manipulations in the study (based on post-experiment debriefing sessions). Given the much-debated interaction between attention and predictive processing (Kok et al., 2012b), I thus developed an experimental design that aimed to control for strategic effects related to the processing of an orthogonal feature (Egner et al., 2010).

Several studies that have reported reduced neural responses for predictable stimuli have used passive viewing (Alink et al., 2010) and stimuli that are fully task-irrelevant (Den Ouden et al., 2009). It thus seemed that expectation suppression effects emerge even in the absence of a task driving the attention of the participants to the prediction manipulation. However, some authors had reported no effects of sensory activity when stimuli were unattended (Larsson and Smith, 2012). These findings would indicate that contextually predictable stimuli may not necessarily suppress early visual neural responses in the absence of a prediction-focused task. But, opposite to these findings, there are studies which report that expectation suppression is mainly driven by the task associated with the stimulus (St. John-Saaltink et al., 2015). This variability about the associated task on one side and expectation suppression on the other side motivated us to employ an orthogonal relationship between task and the expectation introduced in our design.

My experimental design has been thought to evaluate (even if with separate groups of participants) prediction effects in two separate sensory domains. In the visual domain, the orthogonal relationship was established between Gabor orientation and spatial frequencies (CPD). While the orientation manipulation was noticeable, it is essential to underscore that our participants did not report having observed any temporal jitter of the visual stimuli in the temporally unpredictable conditions. In the auditory experiment, the orthogonal relationship was maintained between the target pitch and the length of the tones.

Expectation suppression of early sensory responses

Auksztulewicz et al. (2018) showed that predictability increased the evoked responses' amplitudes only at late latencies (~450 ms). In this thesis, the first finding I would like to underscore is that the expectation suppression affects a predictable stimulus' (i.e., *what* effect) response emerging as early as 100 ms (i.e., the first evoked response to a stimulus) both in the visual and auditory domain. I found the expectation suppression effect in the visual domain when the target feature was predictable and showed that these effects increased incrementally across the entrainer sequence. I interpret this effect as demonstrating that the visual system constantly develops expectations for the upcoming stimuli across the entrainer sequences: the stronger the expectation for a Gabor orientation, the larger the suppression of the visual response. This effect was significant in the evoked responses of the second, third, and fourth Entrainers and was present in conditions both with and without temporal predictability. It is worth noting that the suppression of neural responses was sensitive to the *what* manipulation (i.e., the context of the stimuli) from the second entrainer, whereas the effect of *when* (i.e., timing) emerged later in time, being present only at the fourth entrainer. This suggests that the visual domain is more sensitive to the content (*what*) than the stimuli's timing (*when*). In addition to this differential sensitivity to *what* and *when,* I also report that the timing (*when*

manipulation) only played a role when the content (*what* manipulation) was predictable (at the fourth entrainer); if there is no supporting contextual information present, there was no effect of timing (in line to Correa and Nobre, 2008; Doherty et al., 2005; Rohenkohl and Nobre, 2011; Rohenkohl et al., 2012).

In the auditory domain, I observed that the expectation suppression effect was larger when both the target feature and the timing were predictable. The suppression effect increased incrementally across the entrainer sequence being clearer at the third entrainer. I interpret this effect as demonstrating that the auditory system develops increasingly strong expectations for the upcoming stimuli across sequences both for the content (*what*) as well for the timing (*when*): the stronger the joint expectation about timing and context, the larger the suppression of the auditory response. The suppression effect was significant in the evoked responses of the second, third, and fourth Entrainers and was present in conditions both with and without temporal predictability.

The visual and the auditory systems are thus differently handling *what* and *when* information. To strengthen my claim about the differential sensitivity to *what* and *when* across modalities, I have also tested the 3-factor interaction (*what*, entrainers, and *when*). In the visual domain my results show that the entrainer factor and the *what* factor were mainly driving the suppression effect (Table 1). A two-way interaction of entrainers with *what* factor was also affecting the suppression effect. Finally, a three-way interaction of entrainer**what***when* was also significant. The visual domain evoked responses are highly sensitive to the *what* features, even if the *when* features also interact to shape the visual perception.

Similar to the visual domainIn the auditory domain, the entrainer and *what* factors modulated the suppression effect, but, differently from the visual domain, a two-way interaction between entrainer and *when* was significant. This can be interpreted as a stronger influence of timing

on expectation suppression even in the absence of *what* features. Similar to visual domain, the three-way interaction of entrainer**what***when* was also significant.

It can be inferred from these results that the auditory domain is more sensitive to the *when* manipulation (better integrating it with *what* expectations) compared to the visual domain. This can be explained with two points. First, the auditory domain has a stronger tendency of grouping multiple subsequent tokens in time together than the visual domain. For example, in the case of music and speech, the individual units (notes, syllables, phonemes) which have low perceptual meaning are grouped together to construct intelligible words and semantically more relevant units. This temporal structure of our auditory environment is less relevant in the visual domain, where the "perceptual grouping" becomes more relevant in the spatial dimension. The biological organization of the two sensory systems is relevant here: In the visual domain, the stimuli from external world create an image on the retina (retinotopic mapping), which critically represents more "spatially defined" features; on the other hand, in the auditory domain, the processing of the stimuli is mainly based on the pitch frequency (tonotopic mapping) giving less relevance to spatially defined features. Identifying features in the auditory domain (pitch) is more interwoven with time compared to the visual domain, because "timing" is an inherent property of the auditory domain.

Source-level inferences

The visual evoked response probably originated in visual area 2 (V2), showing reduced activity for predictable stimuli compared to unpredictable stimuli. The source location of the present effect could reflect some top-down activity generated in an extrastriate region projecting to the primary visual cortex (V1). However, this possibility should be further validated (possibly by employing direct brain recordings in non-human primates) with additional connectivity analyses to investigate the bidirectional interaction between V1 and V2 and determine if the

flow of information in the top-down direction is enhanced for predictable content conditions. The potential cortical sources of auditory evoked responses are present in both hemispheres (i.e., in close vicinity of the left and right primary auditory cortex). As mentioned in section 4.3, these sources mainly localize in low-level areas due to the fact that we used pure tones and one could infer that no top-down dynamics are going on in this sensory domain.

This different hierarchical location of the sources of the effects observed in the present paradigms is not easy to explain. It is important to note however that while primary auditory regions are usually involved in processing basic tones, Gabor patches are processed in a higher order non-primary visual region that could integrate pretty complex visual stimuli such as Gabor patches (Rowekamp and Sharpee, 2017). In the future, it would be interesting to use more complex auditory features such as complex and naturalistic stimuli (like speech) to evaluate the involvement of higher order associative auditory regions.

Decoding expected sensory features from early neural signals

My results show that decoding accuracy in the visual domain for Gabor orientations increased across Entrainers if the successive (entrainer and target) orientations were predictable. Temporal predictability did not affect the decodability of the predicted visual stimulus in the earlier time interval when the early visual evoked response emerged. This indicates that the neural representation of the Gabor orientation was stable and preserved independently of the amplitude of the related evoked response and independently of the timing of the visual stimuli. On the other hand, temporal predictability differently affected orientation decoding in a later time interval (525–595 ms, Figure 24), showing that the orientation representation of the fourth entrainer was maintained active for a longer period of time if the timing of the stimulus was predictable. In other words, this suggests that the visual system invests more resources and prolongs the processing of stimulus features when the temporal onset of the visual stimulus is

predictable. This difference mirrors the evoked effects, where I observed stronger visual responses to temporally predictable than temporally unpredictable conditions[2].

In the auditory domain, I could not decode the pitch predictions during the entrainer sequence, but only at the target the classifier was performing very well. This can be explained by the fact that in my experiment, the stimuli that closely match prior expectations about *what* and *when*, increases the sensory detail of the internal representation which, in turn, reduces the informativeness of multivariate representations. Moreover, the high amount of suppression observed across the entrainer sequence could also be involved in the fact that such low evoked auditory responses does not have enough signal-to-noise ratio to boost the decodability of the eliciting stimulus pitch. Thus, it is an open question if, with larger auditory evoked responses, the feature of the expected target tone could be detected or not across the sequence of the entrainers. My goal is to adapt this paradigm for future studies, trying to evaluate this conundrum.

It is worth noting that I did not find the pre-stimulus decoding effect (i.e., representations related to highly predictable stimulus can be decoded from neural responses before it actually appears) reported in previous studies employing expectation/prediction paradigms (in any of the sensory modality: Kok, P., Failing, M. F., & de Lange, 2014; Kok et al., 2017). There still is the possibility that I could observe some pre-stimulus effects in phase or oscillation measures that I have not explored in the present thesis. I am running specific analyses in this direction to address this issue.

[2] Please note that the condition showing larger evoked response is the one showing better decoding accuracy, reinforcing the observation reported in the auditory experiment that more neural suppression is associate to worse decoding accuracy

Prediction and perception

This thesis provided empirical evidence about predictive processing and temporal expectations across sensory modalities. First, my results reinforce the notion of predictive processing as an automatic mechanism (Bendixen et al., 2012). I observed expectation suppression to a predictable feature of sensory stimuli, even when the participants were engaged in processing a different feature of the incoming sensory stimuli.

Second, predictive processing theories (Friston, 2009; Summerfield and De Lange, 2014) associates the attenuation of sensory evoked responses with accurate predictions, and explain the effect in terms of reduced prediction error for predictable stimuli (conversely, it relates unpredicted stimuli with larger evoked responses due to the increased prediction error). My results also show a higher attenuation/suppression to expected stimuli and this suppression increases as the stimuli becomes more predictable in a sequence.

It has been reported that the suppression of evoked responses to predictable stimuli (which is a key notion of predictive processing theories) has recently been challenged by studies on attention and speech (Alink and Blank, 2021; Luthra et al., 2021). Alink and Blank (2021) argue that reduced responses to predictable stimuli can also be explained by a reduced saliency driven allocation of attention. Luthra and colleagues (2021) show that, similarly to the predictive coding theory, the TRACE model (McClelland and Elman, 1986) of speech perception shows reductions in total lexical activation, lexical feedback, and phoneme activation when the input matches prior expectations. Both of these proposals still require critical empirical replications and it is not clear how they dissociate from the predictive coding models. More importantly, no weighting of the expectations is addressed in these proposals. Under a predictive coding framework, the internal generative mental model actively infers the precision (the inverse variance) of prediction error (Feldman and Friston, 2010). Prediction error, generated from the comparison between top-down predictions and bottom-up sensory

signal, is used to update the prior beliefs. This, in turn, is crucial for weighting the "subsequent" prediction errors (assigning *precision*), so that more is learned from precise and reliable prediction errors (associated with higher precision) compared to noisy and unreliable prediction errors (associated with lower precision). Attentional selection could be identified as a precision-weighting mechanism, which increases/decreases the gain of prediction errors depending on the relative relevance of predictions, on one side, and the sensory evidence, on the other side. In the present thesis, using an orthogonal task to control for selective attention, I tried to isolate predictive coding from selective attention. Precision-weighting however could still play a role in the present results as explained below.

Third, although there is evidence that spatial and temporal expectations can be dissociable, it is still not understood how temporal expectations interact with spatial expectations in task-irrelevance situations. Experimental evidence on such interaction is missing. According to the predictive coding theory (Friston, 2009), when stimuli are presented after a short interval with greater frequency, the likelihood of stimuli occurring at that time increases, and therefore, prediction error for such stimuli should be reduced, resulting in smaller M100 amplitude (often called N1 effect in auditory EEG literature). There is however the possibility that temporal expectation might attenuate, boost or have additive effects to neural suppression determined by stimulus predictions. My results suggest that temporal expectations alone cannot elicit robust prediction errors. Instead, temporal prediction could determine the precision associated with pre-existing stimulus-specific predictions, and this associated precision is domain-specific. The visual domain is less affected by precision modulation from temporal expectation, whereas the auditory domain experiences a higher precision modulation arising from temporal expectations. This thesis underscores the variable role of temporal expectation in perception. My findings thus contribute to the expansion of predictive coding frameworks by directly addressing the role of timing in sensory processing (Sherwell et al., 2017).

Finally, I provide evidence to support the idea that the multivariate representations decodable from the neural signals are mainly driven by the prediction error signals and are highly dependent on the signal-to-noise ratio of the signals. Higher suppression to predictable stimuli improves the perceptual processing of stimuli, but this does not necessarily increase the stimulus-specific decodable information (Blank and Davis, 2016).

Conclusions

In the present thesis I explored expectation suppression across sensory modalities (vision and audition). I conclude that the sensory modalities deal differently with the contextual expectations and temporal predictability. This suggests that while investigating predictive processing in the human brain, the modality specific differences should be considered, since the predictive mechanism at work in one domain should not necessarily be generalised to other domains as well. Importantly, predictive processing across other modalities (like tactile and olfactory) in animal models should also be specifically studied since they do not necessarily follow similar predictive principles as visual and auditory perception.

Chapter 6: Appendix

Publications from the thesis:

1. **Nara, S.**, Lizarazu, M., Richter, C.G., Dima, D.C., Cichy, R.M., Bourguignon, M., Molinaro, N., 2021. Temporal uncertainty enhances suppression of neural responses to predictable visual stimuli. Neuroimage 239, 118314. https://doi.org/10.1016/j.neuroimage.2021.118314

2. **Nara, S.**, Lizarazu, M., Klimovich-Gray, A., Bourguignon, M., Molinaro, N., 2021. Role of temporal uncertainty in auditory processing. (in preparation)

3. **Nara, S.**, Zarraga, A., Bourguignon, M., Molinaro, N., 2018. Oscillatory representations of pre-stimulus visual Predictions in Hierarchical Predictive Coding framework, Poster presentation at Neurogune Workshop, Vitoria, Spain.

4. **Nara, S.**, Zarraga, A., Bourguignon, M., Molinaro, N., 2018. Pre and Post stimulus correlates of Prediction in Visual domain, Poster presentation at BIOMAG 2018, Philadelphia, USA

5. **Nara, S.**, Zarraga, A., Bourguignon, M., Molinaro, N., 2018. Neural correlates of Predictable and Non-Predictable targets in Visual Predictions. Poster presentation at 1st International Workshop on Predictive Processing (WoPP), Donostia,San Sebastian, Spain.

6. **Nara, S.**, Zarraga, A., Bourguignon, M., Molinaro, N., 2017. Predicting visual percepts: MEG evidence. Poster presentation at the BrainModes 2017, NBRC, Manesar, India.

Invited talks:

1. “Temporal uncertainty and predictive processing in visual domain” at TEQIP-III sponsored International Conference on Atamnirbhar Bharat: Technological Transformation and Preparedness in the Post COVID World" scheduled to be held on March 22-23, 2021.

2. “Predictive Processing account of Temporal Uncertainty in visual processing” at Society for Neurochemistry India (SNCI) Virtual meeting on "Cognitive and Neurodevelopmental Disorders" from 30th October to 1st November, 2021.

References:

Alink, A., Blank, H., 2021. Can expectation suppression be explained by reduced attention to predictable stimuli? Neuroimage 231, 117824. https://doi.org/10.1016/J.NEUROIMAGE.2021.117824

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus predictability reduces responses in primary visual cortex. J. Neurosci. 30, 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

All, E.B. V, James, W., 2015. Bridging prediction and attention in current research on perception and action 1626, 1–13. https://doi.org/10.1016/j.brainres.2015.08.037

Arnal, L.H., Giraud, A.-L., 2012. Cortical oscillations and sensory predictions. Trends Cogn. Sci. 16, 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Auksztulewicz, R., Friston, K., 2016. Repetition suppression and its contextual determinants in predictive coding. Cortex 80, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Auksztulewicz, R., Schwiedrzik, C.M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A.C., Schroeder, C.E., Friston, K.J., Melloni, L., 2018. Not all predictions are equal: "what" and "when" predictions modulate activity in auditory cortex through different mechanisms. J. Neurosci. 38, 8680–8693. https://doi.org/10.1523/JNEUROSCI.0369-18.2018

Ballard, D.H., Jehee, J., 2012. Dynamic coding of signed quantities in cortical feedback circuits. Front. Psychol. 3. https://doi.org/10.3389/fpsyg.2012.00254

Bendixen, A., SanMiguel, I., Schröger, E., 2012. Early electrophysiological indicators for predictive processing in audition: a review. Int. J. Psychophysiol. 83, 120–131. https://doi.org/10.1016/J.IJPSYCHO.2011.08.003

Berger, H., 1929. Über das Elektrenkephalogramm des Menschen. Arch. Psychiatr. Nervenkr. 87, 527–570. https://doi.org/10.1007/BF01797193

Blank, H., Davis, M.H., 2016. Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. PLoS Biol. 14, 1–32. https://doi.org/10.1371/journal.pbio.1002577

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., Hogendoorn, H., 2020. Predictions drive neural representations of visual events ahead of incoming sensory information. https://doi.org/10.1073/pnas.1917777117

Bourguignon, M., Molinaro, N., Wens, V., 2018. Contrasting functional imaging parametric maps: The mislocation problem and alternative solutions. Neuroimage 169, 200–211. https://doi.org/10.1016/j.neuroimage.2017.12.033

Carlson, T.A., Grootswagers, T., Robinson, A.K., 2019. An introduction to time-resolved decoding analysis for M/EEG.

Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17, 455–462. https://doi.org/10.1038/nn.3635

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. https://doi.org/10.1017/S0140525X12000477

Correa, Á., Nobre, A.C., 2008. Neural modulation by regularity and passage of time. J. Neurophysiol. 100, 1649–1655. https://doi.org/10.1152/JN.90656.2008

Costa-Faidella, J., Baldeweg, T., Grimm, S., Escera, C., 2011. Interactions between "what" and "when" in the auditory system: Temporal predictability enhances repetition suppression. J. Neurosci. 31, 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Coull, J.T., Nobre, A.C., 1998. Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18, 7426–7435. https://doi.org/10.1523/jneurosci.18-18-07426.1998

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical Surface-Based Analysis. Neuroimage 9,

179–194. https://doi.org/10.1006/nimg.1998.0395

de Lange, F.P., Heilbron, M., Kok, P., 2018. How Do Expectations Shape Perception? Trends Cogn. Sci. 22, 764–779. https://doi.org/10.1016/j.tics.2018.06.002

de Munck, J.C., Peters, M.J., 1993. A Fast Method to Compute the Potential in the Multisphere Model. IEEE Trans. Biomed. Eng. 40. https://doi.org/10.1109/10.245635

Demarchi, G., Sanchez, G., Weisz, N., 2019. Automatic and feature-specific prediction-related neural activity in the human auditory system. Nat. Commun. 10, 1–11. https://doi.org/10.1038/s41467-019-11440-1

Den Ouden, H.E.M., Friston, K.J., Daw, N.D., McIntosh, A.R., Stephan, K.E., 2009. A dual role for prediction error in associative learning. Cereb. Cortex 19, 1175–1185. https://doi.org/10.1093/cercor/bhn161

Den Ouden, H.E.M., Kok, P., de Lange, F.P., 2012. How prediction errors shape perception, attention, and motivation. Front. Psychol. 3, 548. https://doi.org/10.3389/FPSYG.2012.00548/BIBTEX

Desimone, R., 1996. Neural mechanisms for visual memory and their role in attention. Proc. Natl. Acad. Sci. U. S. A. 93, 13494–13499. https://doi.org/10.1073/pnas.93.24.13494

DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How Does the Brain Solve Visual Object Recognition? Neuron 73, 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dien, J., 2017. Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. Int. J. Psychophysiol. 111, 42–56. https://doi.org/10.1016/J.IJPSYCHO.2016.09.006

Dima, D.C., Singh, K.D., 2018. Dynamic representations of faces in the human ventral visual stream link visual features to behaviour.

Doherty, J.R., Rao, A., Mesulam, M.M., Nobre, A.C., 2005. Synergistic effect of combined temporal and spatial expectations on visual attention. J. Neurosci. 25, 8259–8266.

https://doi.org/10.1523/JNEUROSCI.1821-05.2005

Donders, F.C., 1969. On the speed of mental processes. Acta Psychol. (Amst). 30. https://doi.org/10.1016/0001-6918(69)90065-1

Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.J., Kirsch, H.E., Dalal, S.S., Deouell, L.Y., Knight, R.T., 2016. Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc. Natl. Acad. Sci. U. S. A. 113, 6755–6760. https://doi.org/10.1073/pnas.1525030113

Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and surprise determine neural population responses in the ventral visual stream. J. Neurosci. 30. https://doi.org/10.1523/JNEUROSCI.2770-10.2010

Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J., 2008. LIBLINEAR: A library for large linear classification. J. Mach. Learn. Res. 9, 1871–1874. https://doi.org/10.1145/1390681.1442794

Feldman, H., Friston, K.J., 2010. Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4. https://doi.org/10.3389/fnhum.2010.00215

Feuerriegel, D., Vogels, R., Kovács, G., 2021. Evaluating the evidence for expectation suppression in the visual system. Neurosci. Biobehav. Rev. 126, 368–381. https://doi.org/10.1016/j.neubiorev.2021.04.002

Fischman, M.G., 1984. Programming time as a function of number of movement parts and changes in movement direction. J. Mot. Behav. 16. https://doi.org/10.1080/00222895.1984.10735329

Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. 47. https://doi.org/10.1037/h0055392

Friston, K., 2009. The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 13. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. 360, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J. -P, Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210. https://doi.org/10.1002/hbm.460020402

Friston, K.J., Price, C.J., Fletcher, P., Moore, C., Frackowiak, R.S.J., Dolan, R.J., 1996. The trouble with cognitive subtraction. Neuroimage 4. https://doi.org/10.1006/nimg.1996.0033

Fuchs, M., Wagner, M., Wischmann, H.A., Köhler, T., Theißen, A., Drenckhahn, R., Buchner, H., 1998. Improving source reconstructions by combining bioelectric and biomagnetic data. Electroencephalogr. Clin. Neurophysiol. 107, 93–111. https://doi.org/10.1016/S0013-4694(98)00046-7

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: A review of underlying mechanisms. Clin. Neurophysiol. https://doi.org/10.1016/j.clinph.2008.11.029

Gina R. Kuperberg, Jaeger, T.F., 2017. What do we mean by prediction in language comprehension? Physiol. Behav. 176, 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Guggenmos, M., Sterzer, P., Cichy, R.M., 2018. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. Neuroimage 173, 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Hakonen, M., May, P.J.C., Jääskeläinen, I.P., Jokinen, E., Sams, M., Tiitinen, H., 2017.

Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates. Brain Behav. 7, 1–15. https://doi.org/10.1002/brb3.789

Hallez, H., Vanrumste, B., Van Hese, P., D'Asseler, Y., Lemahieu, I., Van De Walle, R., 2005. A finite difference method with reciprocity used to incorporate anisotropy in electroencephalogram dipole source localization. Phys. Med. Biol. 50. https://doi.org/10.1088/0031-9155/50/16/009

Hämäläinen, M.S., Sarvas, J., 1989. Realistic Conductivity Geometry Model of the Human Head for Interpretation of Neuromagnetic Data. IEEE Trans. Biomed. Eng. 36. https://doi.org/10.1109/10.16463

Han, B., Mostert, P., de Lange, F.P., 2019. Predictable tones elicit stimulus-specific suppression of evoked activity in auditory cortex. Neuroimage 200, 242–249. https://doi.org/10.1016/j.neuroimage.2019.06.033

Haxby, J. V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430. https://doi.org/10.1126/SCIENCE.1063736

Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691. https://doi.org/10.1038/NN1445

Heilbron, M., Chait, M., 2018. Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? Neuroscience 389, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Henry, F.M., Rogers, D.E., 1960. Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. Res. Q. Am. Assoc. Heal. Phys. Educ. Recreat. 31. https://doi.org/10.1080/10671188.1960.10762052

Hogendoorn, H., Burkitt, A.N., 2018. Predictive coding of visual object position ahead of

moving objects revealed by time-resolved EEG decoding. Neuroimage 171, 55–61. https://doi.org/10.1016/j.neuroimage.2017.12.063

Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195, 215–43. https://doi.org/10.1113/jphysiol.1968.sp008455

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Jancke, D., Erlhagen, W., Schöner, G., Dinse, H.R., 2004. Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. J. Physiol. 556, 971–982. https://doi.org/10.1113/JPHYSIOL.2003.058941

JASP Team, 2020. JASP. [Computer software].

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679. https://doi.org/10.1038/NN1444

Kaswan, J., Young, S., 1965. Effect of stimulus variables on choice reaction times and thresholds. J. Exp. Psychol. 69. https://doi.org/10.1037/h0021742

Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive Processing: A Canonical Cortical Computation. Neuron 100, 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kimura, M., Katayama, J., Ohira, H., Schröger, E., 2009. Visual mismatch negativity: New evidence from the equiprobable paradigm. Psychophysiology 46, 402–409. https://doi.org/10.1111/j.1469-8986.2008.00767.x

Koelsch, S., Vuust, P., Friston, K., 2019. Predictive Processes and the Peculiar Case of Music. Trends Cogn. Sci. 23, 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Kok, P., Failing, M. F., & de Lange, F.P., 2014. Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. J. Cogn. Neurosci. 26, 1546–1554. https://doi.org/10.1162/jocn_a_00562

Kok, P., Jehee, J.F.M., de Lange, F.P., 2012a. Less Is More: Expectation Sharpens

Representations in the Primary Visual Cortex. Neuron 75, 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Mostert, P., De Lange, F.P., 2017. Prior expectations induce prestimulus sensory templates. Proc. Natl. Acad. Sci. U. S. A. 114, 10473–10478. https://doi.org/10.1073/pnas.1705652114

Kok, P., Rahnev, D., Jehee, J.F.M., Lau, H.C., De Lange, F.P., 2012b. Attention reverses the effect of prediction in silencing sensory signals. Cereb. Cortex 22, 2197–2206. https://doi.org/10.1093/cercor/bhr310

Larsson, J., Smith, A.T., 2012. FMRI repetition suppression: Neuronal adaptation or stimulus expectation? Cereb. Cortex 22, 567–576. https://doi.org/10.1093/cercor/bhr119

Luthra, S., Li, M.Y.C., You, H., Brodbeck, C., Magnuson, J.S., 2021. Does signal reduction imply predictive coding in models of spoken word recognition? Psychon. Bull. Rev. 28, 1381–1389. https://doi.org/10.3758/s13423-021-01924-x

Makhoul, J., 1975. Linear Prediction: A Tutorial Review. Proc. IEEE 63. https://doi.org/10.1109/PROC.1975.9792

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. Neuroimage 30, 753–767. https://doi.org/10.1016/J.NEUROIMAGE.2005.10.037

Maunsell, J.H.R., Gibson, J.R., 1992. Visual response latencies in striate cortex of the macaque monkey. J. Neurophysiol. 68, 1332–1344. https://doi.org/10.1152/jn.1992.68.4.1332

McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn. Psychol. 18, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Mumford, D., 1992. On the computational architecture of the neocortex - II The role of cortico-cortical loops. Biol. Cybern. 66, 241–251. https://doi.org/10.1007/BF00198477

Mumford, D., Lee, T.S., 2003. Hierarchical Bayesian inference in the visual cortex. JOSA A, Vol. 20, Issue 7, pp. 1434-1448 20, 1434–1448. https://doi.org/10.1364/JOSAA.20.001434

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin. Neurophysiol. https://doi.org/10.1016/j.clinph.2007.04.026

Nobre, A., Correa, A., Coull, J., 2007. The hazards of time. Curr. Opin. Neurobiol. 17, 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011. https://doi.org/10.1155/2011/156869

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., Näätänen, R., 2007. Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). Clin. Neurophysiol. 118. https://doi.org/10.1016/j.clinph.2006.09.001

Phillips, H.N., Blenkmann, A., Hughes, L.E., Bekinschtein, T.A., Rowe, J.B., 2015. Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions. J. Neurosci. 35, 9255–9264. https://doi.org/10.1523/JNEUROSCI.5095-14.2015

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” Speech Commun. 41, 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: A functional

interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87. https://doi.org/10.1038/4580

Recasens, M., Leung, S., Grimm, S., Nowak, R., Escera, C., 2015. Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human brain: An MEG study. Neuroimage 108, 75–86. https://doi.org/10.1016/j.neuroimage.2014.12.031

Richter, C.G., Thompson, W.H., Bosman, C.A., Fries, P., 2017. Top-down beta enhances bottom-up gamma. J. Neurosci. 37, 6698–6711. https://doi.org/10.1523/JNEUROSCI.3771-16.2017

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025. https://doi.org/10.1038/14819

Rohenkohl, G., Cravo, A.M., Wyart, V., Nobre, A.C., 2012. Temporal Expectation Improves the Quality of Sensory Information. J. Neurosci. 32, 8424–8428. https://doi.org/10.1523/JNEUROSCI.0804-12.2012

Rohenkohl, G., Nobre, A.C., 2011. α oscillations related to anticipatory attention follow temporal expectations. J. Neurosci. 31, 14076–14084. https://doi.org/10.1523/JNEUROSCI.3387-11.2011

Rowekamp, R.J., Sharpee, T.O., 2017. Cross-orientation suppression in visual area V2. Nat. Commun. 2017 81 8, 1–9. https://doi.org/10.1038/ncomms15739

Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm. Proc. Int. Conf. 3-D Digit. Imaging Model. 3DIM 145–152. https://doi.org/10.1109/IM.2001.924423

Sherwell, C.S., Garrido, M.I., Cunnington, R., 2017. Timing in predictive coding: The roles of task relevance and global probability. J. Cogn. Neurosci. 29, 780–792. https://doi.org/10.1162/jocn_a_01085

Spratling, M.W., 2017. A review of predictive coding algorithms. Brain Cogn. 112, 92–97.

https://doi.org/10.1016/j.bandc.2015.11.003

Spratling, M.W., De Meyer, K., Kompass, R., 2009. Unsupervised learning of overlapping image components using divisive input modulation. Comput. Intell. Neurosci. 2009. https://doi.org/10.1155/2009/381457

Srinivasan, M. V., Laughlin, S.B., Dubs, A., 1982. Predictive coding: a fresh view of inhibition in the retina. Proc. R. Soc. London. Ser. B. Biol. Sci. 216, 427–459. https://doi.org/10.1098/rspb.1982.0085

St. John-Saaltink, E., Utzerath, C., Kok, P., Lau, H.C., de Lange, F.P., 2015. Expectation Suppression in Early Visual Cortex Depends on Task Set. PLoS One 10, e0131172. https://doi.org/10.1371/journal.pone.0131172

Stefanics, G., Heinzle, J., Horváth, A.A., Stephan, K.E., 2018. Visual mismatch and predictive coding: A computational single-trial ERP study. J. Neurosci. 38. https://doi.org/10.1523/JNEUROSCI.3365-17.2018

Stefanics, G., Kimura, M., Czigler, I., 2011. Visual Mismatch Negativity Reveals Automatic Detection of Sequential Regularity Violation. Front. Hum. Neurosci. 5, 7. https://doi.org/10.3389/FNHUM.2011.00046

Stefanics, G., KremlÃ¡Ä□ek, J., Czigler, I., 2014. Visual mismatch negativity: a predictive coding view. Front. Hum. Neurosci. 8, 1–19. https://doi.org/10.3389/fnhum.2014.00666

Sternberg, S., 1969. Memory-scanning: mental processes revealed by reaction-time experiments. Am. Sci. 57.

Subramaniyan, M., Ecker, A.S., Patel, S.S., Cotton, R.J., Bethge, M., Pitkow, X., Berens, P., Tolias, A.S., 2018. Faster processing of moving compared with flashed bars in awake macaque V1 provides a neural correlate of the flash lag illusion. J. Neurophysiol. 120, 2430–2452. https://doi.org/10.1152/JN.00792.2017

Summerfield, C., De Lange, F.P., 2014. Expectation in perceptual decision making: neural

and computational mechanisms. Nat. Rev. Neurosci. 2014 1511 15, 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends Cogn. Sci. 13, 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008

Thevenet, M., Bertrand, O., Pcrrin, F., Dumont, T., Pemier, J., 1991. The finite element method for a realistic head model of electrical brain activities: Preliminary results. Clin. Phys. Physiol. Meas. 12. https://doi.org/10.1088/0143-0815/12/A/017

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 2003 64 6, 391–398. https://doi.org/10.1038/nn1032

Utzerath, C., St John-Saaltink, E., Buitelaar, J., De Lange, F.P., 2017. Repetition suppression to objects is modulated by stimulus-specific expectations. Sci. Rep. 7, 1–8. https://doi.org/10.1038/s41598-017-09374-z

Veen, B.D. Van, Drongelen, W. Van, Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Transactions on. Biomed. Eng. (NY). 44, 867–880.

Walsh, K.S., Mcgovern, D.P., Clark, A., Connell, R.G.O., 2020. Evaluating the neurophysiological evidence for predictive processing as a model of perception 1–27. https://doi.org/10.1111/nyas.14321