

Effects of Background Noise and Stimuli Complexity on Speech Perception

Master's Thesis, Psychology
Aalborg University

October 16th 2023

Louise Lehmann Christensen

Student number: 20183721



Supervisor: Thomas Alrik Sørensen

Co-Supervisor: Stefan Kölsch

Total number of characters: 125,446

Equivalent of normal pages: 52.3

Acknowledgements

I would like to thank my supervisors Thomas Alrik Sørensen and Stefan Koelsch for their guidance and support throughout my work on this thesis. In particular, I am deeply grateful to Stefan Koelsch for agreeing to be my supervisor at the University of Bergen, affording me the opportunity to develop academically while living in the picturesque city of Bergen. I am also profoundly grateful to the Department of Biological and Medical Psychology at the University of Bergen for providing the necessary facilities for conducting this study and for fostering an optimal learning environment.

I am deeply grateful to Vera Tsogli and Ryan Douglas McCardle for their indispensable assistance in programming the experiment. Without their expertise and dedication, this endeavor would not have been possible. I would like to express my gratitude to Anna Maria Matziorinis and Tobias Bashevkin from the ALMUTH group for their willingness to answer my questions and share their knowledge. A special note of appreciation goes to Anna Sofía Castells Nieto for her invaluable assistance in the final phase of testing in the EEG-lab. I want to acknowledge my fellow classmates, Olav Toom Jønsi and Thomas Johnsen Gardasevic, for their valuable support, whether through academic discussions, practical advice in the lab, corrections to my Norwegian writing, or simply maintaining a positive spirit even on the less enjoyable writing days. This research would not have been possible without the participation of all the individuals who volunteered as subjects. Your contribution is deeply appreciated.

To my friends, family, and Christian, I extend my sincere gratitude for your unwavering patience, love, and support, which sustained me throughout this endeavor. Finally, I extend my sincere thanks to all the special people who joined me in running, climbing, and mountain adventures, instilling in me a profound love for Bergen and a special place in my heart for the Norwegian nature - a place I am grateful to have called home. Your collective support and encouragement have been vital in the writing of this thesis, and for that, I am immensely grateful.

Thank you!

Louise Lehmann Christensen

Bergen, October 2023

Abstract

In the realm of speech perception, the presence of background noise poses significant challenges, complicating communication for individuals with both normal and impaired hearing. Expectations play a pivotal role in this soundscape, influencing our ability to detect and interpret auditory input, especially in noisy environments. This thesis delves into the impact of background noise and increased stimuli complexity on speech sound perception, utilizing the Mismatch Negativity (MMN) as a neurophysiological tool. Furthermore, it seeks to understand the impact of these factors within the framework of Predictive Coding (PC).

The primary aim of this thesis is to investigate how background noise and increased stimuli complexity affect the perception of speech sounds and to comprehend the underlying mechanisms within the PC framework. Specifically, this thesis employs the vowel sounds /a/ and /e/ as speech stimuli to investigate the auditory discrimination ability regarding timbre deviants, with unintelligible babble noise used as background noise. Four key hypotheses were proposed and tested, providing insight into the dynamics of this interplay. The results supported the first hypothesis, confirming that the presence of background noise leads to a reduction in MMN amplitude. This outcome underscores the influence of noise on speech sound perception. The second hypothesis was also substantiated, revealing that increased stimuli complexity diminishes MMN amplitude, emphasizing the importance of considering complexity when studying neural processing of speech sounds. While the third hypothesis regarding an interaction between background noise, stimuli complexity, and its influence on the MMN did not yield statistically significant results, a noteworthy trend emerged, suggesting a potential relationship deserving further exploration. Finally, the fourth hypothesis, which predicted an increase in the peak amplitude latency of the MMN in response to the presence of background noise, was supported by the results, highlighting the temporal impact of noise on neural responses. Within the context of predictive coding, these findings can be understood as a reflection of decreased prediction error induced by timbre deviations. The brain forms predictive models of future events, and the reliability of these predictions directly influences neural responses to deviations from them. Increased attentional demands resulting from background noise and stimuli

complexity lead to a decrease in the precision of predictions, thereby affecting the ERP amplitude in response to standard stimuli, which, in turn, results in a reduced MMN signal.

In summary, this study contributes to our understanding of how background noise and increased stimuli complexity impact speech sound perception within a predictive coding framework. While acknowledging the limitations of this research, the findings support previously established results in the field, providing insights for future investigations aimed at enhancing auditory discrimination in noisy environments. This, in turn, holds the potential to ultimately benefit individuals with hearing impairments and advance hearing aid technology.

Table of Contents

Acknowledgements	
Abstract	
Introduction	1
Background	2
Auditory Perception.....	2
Background Noise and Speech Perception.....	3
Vowel Perception	5
Electroencephalography	6
Mismatch Negativity and the Oddball Paradigm	8
MMN as an Index of Sound Discrimination	9
Background Noise and the MMN.....	10
Stimuli Complexity and the MMN.....	12
Predictive Coding	15
Predictive Processes and the MMN	17
Purpose and Hypotheses	20
Materials and Methods	21
Participants	21
Stimuli	21
Background Noise.....	22
Vowel Stimuli	22
Editing of Vowel Stimuli.....	23
Presentation of Vowel Stimuli.....	25
Procedure	26
Sound Pressure Level.....	26
Questionnaire, Informed Consent, and Compensation Procedures	26
EEG-Recording.....	27
Discrimination Test.....	28
Main Experiment	29
Presentation of Blocks.....	31
Eyeblink Corrections	32
Results	33

Behavioral Data	33
Electroencephalographic Data	34
Distribution of Data	34
Statistical Analysis of EEG-Data	36
Main Effects	36
Interactions	37
<i>ERP Waveforms</i>	40
Changes in Latency	41
What Has a Greater Impact on MMN: Background Noise or Increased Stimuli Complexity?	43
Are the MMN Reductions Mainly Derived from Changes in ERPs to Regular or Irregular Stimuli?	44
Discussion	46
Effects on MMN-Amplitude	46
Effects on MMN-Latency	48
Findings Understood in a Predictive Coding Perspective	49
Methodological Considerations and Future Research	52
Conclusion	55
References	57
Appendix	67

Introduction

During a conversation, different types of background noise are often inevitable companions. The presence of background noise can turn speech communication into a difficult task, even for people with normal hearing (Koerner et al., 2016, p. 40). For individuals with hearing impairment, speech perception in noise is even more complicated (Legris et al., 2018, p. 336). But it is not only the background noise that affects what we hear. We tend to interpret auditory input in favor of what we are expecting to hear. So in the context of background noise, we are more likely to detect a sound, if we are expecting and waiting for its emergence (Alain & Winkler, 2012, p. 90).

Half a century after Hans Berger began his study of human electroencephalograms (EEG) in 1920 (Sanei & Chambers, 2007, p. 2), pioneering studies that utilized the mismatch negativity (MMN) to investigate the auditory sensory memory were conducted (Näätänen et al., 1978, p. 313; Tervaniemi, 2022, p. 2). In addition to behavioral data, electrophysiological methods are important supplements for studying speech perception. The MMN-signal is correlated with the behavioral discrimination of auditory stimuli, and the amplitude increases as the ability to discriminate the sounds improves. Moreover, the MMN can be elicited even if the stimuli are not being attended to, which allows for the objective assessment of cortical processing of sound (Lonka et al., 2004, p. 160). Today, cortical auditory evoked potentials (CAEP's), such as the MMN, are routinely used as a clinical tool to demonstrate the audibility of speechlike sounds. In the process of hearing aid fitting, it serves as a valuable validation method, especially in cases where individuals cannot respond behaviorally (Maslin et al., 2023, p. 1). Furthermore, MMN can be employed as an objective measurement to assess the performance of different hearing aid noise reduction systems (Legris et al., 2018, p. 336f), in addition to the subjective methods most frequently used by clinicians to assess reported discomfort (Legris et al., 2018, p. 335).

The experiment conducted in relation to this thesis has been developed with the possibility of using the design in a further study on the performance of different hearing aid systems in the presence of background noise. However, since speech perception relies on and is affected by many components, it is also essential to investigate how background noise affects the perception of speech components among individuals

with normal hearing. One such component is vowel sounds, which can be differentiated based on timbre. Therefore, this thesis aims to investigate the ability to discriminate vowel sounds in the presence of background noise, explore the role of stimuli complexity, and examine how these phenomena can be understood through the lens of predictive coding.

Background

To investigate how background noise affects the perception of speech sounds, it is essential to commence by examining previous research, the current state of knowledge in the field, and the methods used to gather this information. These aspects will be explored in the following section.

Auditory Perception

Sound stimuli travel from the source of origin to the ear through the movement of air molecules. In order for the sound waves to be perceived, the auditory system must convert this mechanical energy into neural signals (Purves et al., 2013, p. 93). The sound waves can be heard when they reach the ear, where they initially induce vibrations in the structures of the outer and middle ear, which, in turn, vibrate the cochlea in the inner ear (Ashley & Timmers, 2017, p. 4). The cochlea is a narrow, spiral-shaped, fluid-filled hearing organ, and within it is the *basilar membrane*. Acoustic vibrations result in pressure variations in the cochlear fluids, causing the basilar membrane to vibrate (Plack, 2004, p. 6). The basilar membrane resonates at different frequencies at various points along the membrane, making it possible to decompose complex sounds into individual frequency components (Ashley & Timmers, 2017, p. 4). At the base of the cochlea, the basilar membrane is thin and stiff, making it most sensitive to high frequencies. Near the apex, at the opposite end of the cochlea, the basilar membrane is thick and loose, and most sensitive to low frequencies (Plack, 2004, p. 7). This arrangement is referred to as *tonotopic organization* or *tonotopy* (Purves et al., 2013, p. 97). In the cochlea, acoustic information is translated into neural activity before being progressively transformed in the auditory brainstem. From the thalamus, neural impulses are mainly projected into the auditory cortex (Koelsch, 2011, p. 1). These neural representations of sounds, are mainly projected over the medial geniculate body, which is the main auditory-responsive portion of the thalamus (Bartlett, 2013, p. 1).

The auditory pathway consists of both bottom-up and top-down projections (Kolesch, 2011, p. 2). Already at the level of the peripheral auditory system, which includes the external, middle, and inner ear (Purves et al., 2013, p. 96), it is evident that cochlear resonance is not a passive process; it cannot be understood merely as a microphone receiving input. Active cochlear processes can tune the system, so to speak, by responding to top-down signals depending on incoming sounds, allowing for soft sounds to be amplified and loud sounds to be compressed (Ashley & Timmers, 2017, p. 4; Eguíluz et al., 2000, p. 5235). When top-down processes enable alterations in the way sounds are processed in the auditory periphery, it not only helps to protect against damage caused by overly loud sounds but also offers the potential to selectively attend to specific signals and to improve the detection of signals in the presence of background noise (Cooper & Guinan, 2006, p. 49).

Background Noise and Speech Perception

When we speak and listen to other people, we often find ourselves in different sorts of noisy environments. In these situations, we are exposed to many sounds that are irrelevant and disturbing when it comes to understanding the speech we are focusing on. Successful speech communication in such scenarios relies on the ability to extract the important acoustic cues from the background noise (Koerner et al., 2016, p. 40).

Even when speech sounds and background noise have the same intensity, individuals with normal hearing can discriminate between speech sounds above chance level (Shetake et al., 2011, p. 1). The ability to extract the relevant information in these realistic and complex listening situations is known as the *cocktail party effect* (Alain & Winkler, 2012, p. 91). As the name implies, it refers to the selective enhancement that occurs when you direct your attention to one person's speech, enabling you to process the speech sounds despite the co-occurring cocktail party noises in the background (Breedlove & Watson, 2017, p. 574). Even though we are able to focus our attention on the speech, several studies have found that background noise diminishes speech intelligibility (Kozou et al., 2005, p. 31). This is mainly explained by how noise affects peripheral hearing mechanisms, where noise exposure results in changes in the hearing threshold. Due to these threshold shifts, the ability to recognize words decreases, partly because of a reduced ability to extract information from speech signals (Kozou et al., 2005, p. 31f).

Since threshold shifts or reduced audibility were considered as one of the primary reasons for speech recognition deficits in noise, this aspect gained more attention than the effects of noise on central auditory processing (Kozou p. 32). It has been observed that different types of background noise have varying effects on speech recognition performance. This suggests that hearing in noise cannot be solely attributed to the audibility of the target sound at the peripheral auditory level and emphasizes the significant role of central processing in the perception of speech in competing background noise (Kozou, p. 32).

The background noise, sometimes referred to as a ‘competitor’ in the literature, can consist of speech, noise, or another signal (different from the target sound (Sperry et al., 1997, p. 71)). It can also be considered a form of masking as the background noise masks the target sound. In this regard, background noise can be divided into two types. One is *energetic masking*, which refers to a listening situation where competing noise overlaps with the speech signal in time and frequency, making the speech inaudible. The other is *informational masking*, which refers to a situation where the listener involuntarily processes irrelevant speech due to an inability to separate the target sounds from distractors that are linguistically meaningful (Harmon et al., 2021, p. 1104). Because an informational masker contains meaningful information that the listener consciously attempts not to be distracted by, it is more cognitively demanding than processing speech signals in the presence of energetic masking (Harmon et al., 2021, p. 1104; Meekings et al., 2016, p. 8). This is supported by the finding that speech competitors disturb speech recognition performance more than non-speech competitors. Additionally, performance declines to a greater extent when the listener is exposed to meaningful speech competitors than when exposed to non-meaningful speech competitors (Kozou et al., 2005, p. 32; Sperry et al., 1997, p. 71). Studies investigating the effect of background noise have used various types of noise and different types of target stimuli, which will be further elaborated upon in the subsection *Background Noise and the MMN*.

In everyday life, we are typically exposed to noise levels ranging from 30–100 dB. For reference, a good night's sleep usually requires noise levels below 35–40 dB. The noise level in a busy open-office environment, with constant background noise, is typically around 60 dB, while an ordinary conversation is usually around 65 dB. If you

walk beside a busy road with heavy traffic, the noise level is estimated to be around 75 dB. When standing on the side of a runway during a jet aircraft departure, noise levels reach approximately 120 dB (Murphy & King, 2022, p. 22). It should be noted that sounds presented at the same dB level may be perceived differently in intensity if presented at different frequencies. For example, if two sounds are presented at 60 dB, one at 1000 Hz (Hertz, cycles per second) and the other at 100 Hz, the sound at 100 Hz is perceived as being almost 20 dB lower in magnitude than the sound at 1000 Hz (Murphy & King, 2022, p. 27; Simpson, 2009, p. 622).

Differences in frequency are also a key factor in distinguishing male and female voices. The fundamental frequency (F0), or frequency of vibration, is closely correlated with the perceived pitch of a voice. On average, male speakers have an F0 of 100-120 Hz, while female speakers typically have an F0 of 200-220 Hz in languages such as German and English (Simpson, 2009, p. 622; Weiss, 2021, p. 153). In certain Chinese dialects, male speakers have been reported to have an average F0 of 170 Hz, while women have an average of 187 Hz. Studies have also identified variations in average male and female F0 within European languages. These differences cannot be solely attributed to anatomical variations but must also be attributed in large part to learned behaviors (Simpson, 2009, p. 625).

Vowel Perception

An important prerequisite for speech understanding is accurate speech perception, which is based on the phonemes of the language in question and how these phonemes, in combination, represent syllables and words (Näätänen, 2001, p. 7). In any language, the basic speech sound stimuli are called *phones*, and the perceptions elicited by phones are called *phonemes* (Purves et al., 2013, p. 396). If one phoneme is replaced by another, it can change the word. For example, the word “bee” consists of the two phonemes /b/ and /i/. If /b/ is replaced by the phoneme /s/, you get the word “see”.

Phones (and phonemes) can be categorized into vowel and consonant speech sound stimuli. Vowel sounds have a tonal quality that elicits the perception of pitch, while consonant sounds are often shorter than vowel sounds and typically involve rapid changes in sound energy, making them acoustically more complex. Even though both vowel and consonant speech sound stimuli are important for accurate speech

perception, consonants have been found to be the main carriers of information when interpreting speech. When presented with a spoken sentence from which either the vowels or consonants have been removed, it is easier to interpret the sentence correctly when the vowels are missing than when the consonants are missing (Purves et al., 2013, p. 396).

Since consonants are considered the primary carriers of information in speech (Purves et al., 2013, p. 396), it is interesting that background noise has been found to affect the processing of consonant and vowel stimuli differently. Neural responses have been shown to be more robust when participants are exposed to vowel stimuli than when exposed to consonant stimuli. An important distinction is that vowels are more steady, whereas consonants are more transient and aperiodic (Koerner et al., 2016, p. 41).

To align with the scope of this thesis and maintain a relatively simple experimental design, the decision was made to exclusively focus on *monophthongs*, which are simple vowel sounds characterized by their lack of detectable changes within them (Colman, 2015, p. 275). Therefore, this thesis employs the vowel sounds /a/ and /e/ as speech stimuli. As briefly mentioned in the *Introduction*, one way to investigate how background noise affects speech perception is by recording the electrical activity of the brain using the method of electroencephalography, which will be described in the following.

Electroencephalography

Electroencephalographic (EEG) recording is a noninvasive method for measuring the electrical neural activity of the brain, which can be used to study brain activity associated with cognitive processes (Purves et al., 2013, p. 31f; Sanei & Chambers, 2007, p. 1). *Electro* refers to the registration of electrical activity, *encephalo* refers to the signals being emitted from the head, and *gram* or *graphy* means drawing or writing (Sanei & Chambers, 2007, p. 1). The method employs surface electrodes, often embedded in a cap, to detect electrical brain waves on the scalp. The number of electrodes can range from a few to 256. It is essential to ensure good contact between the electrode and the skin; for this purpose, conducting gel is applied under each electrode (Purves et al., 2013, p. 32). The voltage of each electrode is compared to a reference

electrode, and the recording from each electrode reflects the electrical activity from the brain area beneath that electrode (Gazzaniga et al., 2014, p. 99).

Activity in the central nervous system involves the flow of currents between synapses, through which nerves transmit information via action potentials (Sanei & Chambers, 2007, p. 4). Unlike single-neuron recordings, the EEG signal does not reflect the firing of individual action potentials. Instead, it derives from the summed dendritic field potentials originating from groups of neurons (Purves et al., 2013, p. 32). As continuous EEG recordings typically reflect the global electrical activity of the brain, the EEG itself reveals little about cognitive processes because it is challenging to link the modulation of activity to any specific stimulus or task (Gazzaniga et al., 2014, p. 100). To analyze the EEG signal concerning specific stimuli, *event-related potentials* (ERPs) are employed, enabling the examination of time-locked brain activity in response to events related to particular stimuli (Ashley & Timmers, 2017, p. 50). Since EEG recordings are taken from the surface of the scalp, it can be difficult to precisely localize where the responses originate. Additionally, the auditory cortex relays its signal to other cortical areas, which can influence the measured responses. Therefore, ERPs are more effective at providing information about the temporal aspects of cognition rather than spatial aspects when investigating components generated in the cortical areas (Gazzaniga et al., 2014, p. 101).

EEG recordings are not only of interest in research but also serve as essential tools in the clinical setting. Since normal EEG patterns are well established, they can be used to detect abnormalities in brain function and are, for example, very useful in the assessment of epilepsy (Gazzaniga et al., 2014, p. 99). ERPs also provide valuable information for diagnosing diseases such as multiple sclerosis. When examining visual evoked potentials, you may observe a delay in the early peaks of the visual evoked response because the electrical signal cannot travel as quickly as usual if there is demyelination in the optic nerve. Similarly, auditory evoked potentials (AEPs) can be employed to locate tumors affecting auditory processing areas based on characteristic wave peaks and our understanding of which anatomical areas in the auditory system give rise to the AEPs of interest (Gazzaniga et al., 2014, p. 101).

One type of event-related potential that has been applied in a wide range of research and is tightly linked to the concepts of memory and attention is the mismatch negativity, which is also the ERP component of interest in this thesis (Winkler, 2007, p. 147).

Mismatch Negativity and the Oddball Paradigm

The mismatch negativity (MMN) is a frontocentrally negative component of the auditory event-related potential (ERP) that typically peaks at 100-250 milliseconds (ms) after stimulus onset (Näätänen, 2001, p. 1, 2002, p. 98). It can be elicited even in the absence of attention, although research has shown that a lack of attention can reduce the MMN response (Fong et al., 2020, p. 4). MMN is generated bilaterally in the auditory cortices and represents an essential biological function: the ability to monitor and detect changes in ongoing auditory stimulation, regardless of the direction of attention (Duncan et al., 2009, p. 1885).

The MMN, as well as its magnetoencephalographic equivalent, MMNm, is elicited by any discriminable change in an otherwise repetitive stream of auditory stimulation (Näätänen, 2001, p. 1, 2002, p. 98). A classic method for investigating the MMN is the employment of an oddball paradigm, in which a train of repetitive standard sounds is interrupted by an infrequent deviant sound (Fong et al., 2020, p. 4). The mismatching stimuli used in the auditory oddball paradigm can differ on any discriminable auditory dimension (Duncan et al., 2009, p. 1885). The MMN is obtained by subtracting the event-related response elicited by the presentation of the standard tone from the response elicited by the oddball or deviant tone (Fong et al., 2020, p. 4).

Recording the MMN signals elicited when participants are exposed to auditory deviant stimuli in the presence of background noise is one method for measuring the neural coding of speech in noisy environments (Koerner et al., 2016, p. 40). The MMN response can be regarded as an objective measure of the participant's ability to extract the target stimuli from the background noise, providing measures of neural timing and the strength of preattentive auditory discrimination based solely on electrophysiological measures (Koerner et al., 2016, p. 40).

Some studies, primarily based on dichotic selective-listening conditions, have shown that strongly focused attention can modulate the MMN amplitude, but the data does not indicate that the withdrawal of attention can completely eliminate the MMN

response (Näätänen, 2002, p. 98). The MMN component is thought to reflect an automatic process in which the difference in an incoming stimulus is detected by comparing it to the sensory memory trace of preceding stimuli (Duncan et al., 2009, p. 1885). Some convincing evidence that the MMN is generated automatically in the absence of attention is provided by MMN recordings in coma patients, where it can serve as a strong predictor for the return of consciousness within a week, and in anesthetized cats, guinea pigs, and rats. Although the MMN amplitude is reduced compared to normal conditions, it is not absent (Näätänen, 2001, p. 1f).

The sensory memory trace upon which the elicitation of the MMN is based represents the repetitive pattern of the standard stimuli (Näätänen, 2001, p. 2). In normal subjects, the duration of the period during which this memory trace remains active, also referred to as echoic memory, has been reported to be at least 5-10 seconds (Garrido et al., 2009, p. 455). However, since the memory trace typically fades within this time window (Sams et al., 1993, p. 363), a deviant stimulus will not elicit an MMN response if the interstimulus interval exceeds this duration (Näätänen, 2001, p. 2). It is also important to note that, for the MMN to be elicited, the standard stimulus element does not necessarily have to be acoustically constant. As long as the standard stimuli share some pattern or rule, the MMN will be elicited when the deviant stimuli violate this pattern or rule (Näätänen, 2001, p. 2). Furthermore, the MMN can be elicited even when there is a very small difference between the standard and the deviant stimuli. However, as this difference increases, the peak latency of the MMN progressively shortens, while the peak amplitude increases (Duncan et al., 2009, p. 1885).

The fact that the MMN can be elicited even when stimuli are not being attended to, combined with its high sensitivity to perceived stimulus changes, provides a unique measure of auditory discrimination accuracy without the need to rely on behavioral measures (Näätänen, 2002, p. 98).

MMN as an Index of Sound Discrimination

MMN can serve as a reliable index of the ability to discriminate linguistic stimuli. In studies investigating such discrimination abilities, phonemes or consonant-vowel syllables are commonly employed. If the participant cannot discriminate between two different phonemes or syllables, no MMN will be elicited. Conversely, the MMN

amplitude is larger when discrimination ability is higher, and when discrimination trials are easier (Näätänen, 2001, p. 7).

Aaltonen et al. (1987) conducted the first study that reported an MMN in response to a change in phoneme stimuli. They used the Finnish phonemes /y/ and /i/. /y/ was used as the standard stimulus, and they found an MMN with a large amplitude and short latency when /i/ was presented as the deviant stimulus. However, when the deviant stimulus was replaced with /yi/, which they referred to as a boundary stimulus because it was less easily discriminable, the MMN signal was smaller and more delayed (Aaltonen et al., 1987, p. 197; Näätänen, 2001, p. 7).

An example of a study using consonant-vowel syllables is one conducted by Kraus et al. (1996), where an MMN was found only in the children who were able to discriminate between /da/ and /ga/ in the behavioral test. For the children who could not behaviorally discriminate, no MMN was found. However, for the /ba/ and /wa/ contrast, both groups of children were able to behaviorally discriminate, and a strong MMN was found (Kraus et al., 1996, p. 971; Näätänen, 2001, p. 7).

As mentioned earlier, background noise can compete with speech signals and thereby interfere with our speech recognition performance (Sperry et al., 1997, p. 71). In situations involving competing background noise, the auditory system's ability to extract relevant acoustic information is therefore crucial for speech perception (Cunningham et al., 2002, p. 97). In the following section, a more detailed description of how the MMN can be used to investigate the effects of background noise on speech perception will be provided.

Background Noise and the MMN

As will be presented here, various studies have explored the MMN signal to investigate how automatic preattentive auditory discrimination is affected by the presence of background noise (Kozou et al., 2005, p. 32).

In a study conducted by Salo et al. (1995), they used white noise at 50 dB as a masking condition and sine tone stimuli at 70 dB as target stimuli. Their investigation focused on the MMN response when subjects were exposed to unilateral stimulation and contralateral masking. This means that when the target stimuli were presented to one ear, white noise was presented to the other ear. When comparing the results from

the condition with contralateral masking to the results of the condition where only target stimuli were presented, they found that the MMN amplitude was significantly attenuated when contralateral masking was present. Additionally, they observed that white noise masking did not affect the MMN latencies (Salo et al., 1995, p. 165f).

In a study conducted by Martin et al. (1999), the researchers investigated how reduced audibility affected the event-related potentials N1 and MMN. They used the speech sounds /ba/ and /da/ as target stimuli at a sound level of 65 dB and employed high-pass masking noise with varying high-pass cutoff frequencies. As the cutoff frequency of the high-pass masker was lowered, more frequencies were included in the noise, resulting in decreased audibility. Their findings revealed that the ERP latencies increased, and amplitudes decreased as the cutoff frequency of the high-pass masker was lowered. Furthermore, they observed that the N1 response was present in all conditions where the stimuli were audible, regardless of whether they were discriminable or not. In contrast, the MMN response was only present in conditions where the stimuli were behaviorally discriminable (Martin et al., 1999, p. 271).

A similar study was conducted by Muller-Gass et al. (2001), in which they used the sounds /ba/ and /da/ as target stimuli and speech noise for masking. They manipulated audibility by presenting the stimuli without masking noise or with masking noise at 65, 70, or 75 dB. In this study, they also observed that the MMN latency increased, and MMN amplitude decreased as audibility decreased. Consistent with the MMN data, the behavioral responses indicated that as audibility decreased, behavioral discrimination also declined, as reflected in the hit rates. From these findings, they suggested that a reduction in audibility leads to deterioration in both preattentive and attentive speech discrimination (Muller-Gass et al., 2001, p. 197f).

Another study conducted by Shtyrov et al. (1998) investigated the interesting phenomenon that noise also alters the hemispheric lateralization of speech sounds (Kozou et al., 2005, p. 32). They measured the magnetic equivalent (MMNm) of mismatch negativity while presenting consonant-vowel syllables during white noise or with no background noise. The MMNm response to speech stimuli, recorded from the auditory cortex, was found to be stronger in the left hemisphere than in the right hemisphere under conditions with no background noise. However, the MMNm in the left

hemisphere decreased, while it increased in the right hemisphere when the speech signals were presented with white noise in the background (Shtyrov et al., 1998, p. 141).

As pointed out by Kozou et al. (2005), many studies make use of broadband noise, which does not reflect the background noise of an everyday listening situation (Kozou et al., 2005, p. 32). In their own study, they decided to investigate how different background conditions affected the processing of speech and non-speech stimuli (Kozou et al., 2005, p. 31). For speech noise, the consonant-vowel syllables [ka] and [ko] were used. For the non-speech stimuli, two complex sounds were created, each with a duration of 157 ms and similar acoustic features (Kozou et al., 2005, p. 34). The five different background conditions were babble noise, industrial noise, traffic noise, wide-band noise, and a silent condition. On the behavioral responses, they did not find any effects of different stimuli or noise types. On the other hand, the MMN results revealed that all noise conditions affected speech processing more than non-speech processing. Furthermore, the MMN amplitudes for both speech and non-speech stimuli were markedly reduced with babble and industrial noise in the background, while only speech stimuli were affected by traffic noise (Kozou et al., 2005, p. 31). While the type of background noise plays a role in the ability to detect the target stimuli, the complexity of the target stimuli itself also affects the processing.

Stimuli Complexity and the MMN

In order to understand speech, the ability to discriminate sounds that vary in different properties, such as frequency and temporal characteristics, is crucial (Wunderlich & Cone-Wesson, 2001, p. 1526). Furthermore, the brain's ability to predict the timing of upcoming events is critical for both music and speech perception (Lumaca et al., 2019, p. 1597). As the complexity of stimuli increases, predicting and discriminating sounds can become more challenging.

One study investigating the effect of temporal complexity was conducted by Lumaca et al. (2019), who tested seventeen right-handed participants (Lumaca et al., 2019, p. 1599). In this study, participants were repeatedly presented with standard rhythmic patterns of five tones. For the deviant stimulus trials, the fourth tone was presented either 100 or 300 ms early. There were three different conditions with a given level of entropy. The stimuli could be 1) isochronous (where all intervals are of

equal duration), 2) have two distinct interval durations (medium entropy), and 3) have four distinct interval durations (high entropy) (Lumaca et al., 2019, p. 1597). They found that the MMN amplitude decreased as rhythmic complexity increased. This is interpreted through a predictive coding framework, where the MMN amplitude reflects the size of the neural prediction error elicited by the timing deviations (Lumaca et al., 2019, p. 1604). Furthermore, the gradual decrease in MMN amplitude was only found for small timing deviants. For large timing deviants, an increase in amplitude was observed in the N1 component, along with increasing stimulus entropy (Lumaca et al., 2019, p. 1597). The N1 component typically occurs 100 ms after the onset of any sound and is known to be sensitive to sudden changes in directed attention. The N1 appears to reflect the detection of feature-specific deviations and is not sensitive to the violations of statistical regularities like the MMN. Therefore, only the MMN can be interpreted as a neural response that indicates error processing (Lumaca et al., 2019, p. 1598). Since the large timing deviations are more salient than the small timing deviations, they can trigger mechanisms of attentional capture. Therefore, the modulation of N1 can be interpreted as a combination of the effects of temporal regularity and attention on the deviant-related responses (Lumaca et al., 2019, p. 1604f).

Wunderlich and Cone-Wesson (2001) investigated the effects of stimulus frequency and complexity in a study with 12 participants (Wunderlich & Cone-Wesson, 2001, p. 1526f). To test the effect of frequency, participants were presented with three test conditions contrasting easily discriminable pure tones. The pure tones were presented at frequencies in different speech-frequency ranges: low (400/440 Hz), mid (1500/1650 Hz), or high (3000/3300 Hz) (Wunderlich & Cone-Wesson, 2001, p. 1528). For both the N1 and MMN components, they found that amplitude decreased as frequency increased (Wunderlich & Cone-Wesson, 2001, p. 1526).

Furthermore, participants were presented with two speech test conditions contrasting the words “bad” /bæd/ with “dad” /dæd/ or the consonant-vowel stimuli /bæ/ and /dæ/, which are more complex than the pure-tone stimuli (Wunderlich & Cone-Wesson, 2001, p. 1528). Even though participants were able to discriminate behaviorally in all conditions, there was a considerable difference in the presence of MMN. They found that the MMN was more frequently present for tones than for speech sounds. For the tones, the MMN was present in 46-71% of the tests. For the speech

sounds, the MMN was present 25-32% of the time (Wunderlich & Cone-Wesson, 2001, p. 1530). Based on existing literature, one would expect to observe MMN responses for discriminable words and consonant-vowel stimuli, as already described in the section *MMN as an Index of Sound Discrimination*. Wunderlich and Cone-Wesson (2001) conducted an additional experiment to investigate the effect of stimulus complexity on the MMN. They used tone complexes consisting of 400 Hz and 3000 Hz, with either a 1500 Hz or 1650 Hz component. Each tone complex had a total duration of 200 ms, but at 80 ms after stimulus onset, the 1500 Hz component was modified to 1650 Hz for the first tone complex. Conversely, for the second tone complex, the 1650 Hz component was changed to 1500 Hz (Wunderlich & Cone-Wesson, 2001, p. 1531f). In the tone complex conditions, they found that the MMN was recorded in only three out of 16 tests (19%), which was similar to the presence of MMN for speech sounds (Wunderlich & Cone-Wesson, 2001, p. 1532). In the behavioral test, all participants were able to discriminate the contrasts with 100% accuracy. In their concluding remarks, Wunderlich and Cone-Wesson (2001) therefore point out that the study indicates that it is not only the participants' discrimination ability that influences the MMN but also the characteristics of the stimulus (Wunderlich & Cone-Wesson, 2001, p. 1535).

In a study conducted by Tsogli et al. (2019), participants were presented with a continuous stream of sound triplets, in which the final tone occasionally deviated either statistically, physically, or in a combination of both (double deviants). The statistical deviation should be understood in terms of transitional probability, representing the probability of a given triplet ending based on the preceding sounds. Meanwhile, the physical deviant was based on the sound location, specifically whether the sound originated from the left or right speaker. The study revealed that the statistical MMN was smaller when it occurred in conjunction with a physical deviant. This finding indicates an interaction between the processing of statistical and physical deviants, suggesting that the physical attributes of a stimulus suppress the processing of extracted statistical properties of the stimulus when it comes to the prediction error processing of sound, as reflected in the MMN (Tsogli et al., 2019, p. 1).

In a subsequent study by Tsogli et al. (2022), they employed a variation of the experimental design used in their previous study. In the analysis, the datasets from 21

participants in the current experiment were compared with datasets from 21 participants in the previous study (Tsogli et al., 2022, p. 4). They tested whether the statistical or physical deviants would still elicit a statistical MMN or a location MMN when temporal predictability was low (Tsogli et al., 2022, p. 3). They manipulated the predictability of stimulus onset by using random stimulus onset asynchronies (SAOs), compared to the prior study where stimuli were presented in an isochronous fashion (Tsogli et al., 2022, p. 1). It was found that the random SOAs resulted in reduced amplitudes for both the statistical MMN and location MMN, which indicates that temporal unpredictability influences the processing of other stimulus attributes. Their results demonstrate an interaction between predictions about different attributes of a stimulus, such that unpredictability in one stimulus attribute hinders the processing of other stimulus aspects (Tsogli et al., 2022, p. 16).

Predictive Coding

The MMN reflects the brain's ability to automatically compare consecutive stimuli and detect differences between them. The underlying mechanisms of the generation of the MMN can be explained within a predictive coding framework (Garrido et al., 2009, p. 453). Through the lens of predictive coding (PC), the mental representations that are a significant part of both perception and action reflect an interplay between descending predictions and ascending sensory signals (Lupyan & Clark, 2015, p. 279). Thus, PC provides a framework for understanding perceptual processes in audition and language comprehension (Koelsch et al., 2019, p. 1).

As explained by May (2021), PC is a proposed general principle for describing how the brain realizes Bayesian inference when interpreting sensory information (May, 2021, p. 1). Bayesian models have been used in various psychological domains, particularly excelling in the fields of visual perception and motor control (Rescorla, 2021, p. 1). It offers a perspective in which perception is not solely driven by bottom-up processes but rather by a combination of sensory stimuli and prior information about the most likely causes of sensations. In this view, perception operates as an inferential process based on probability and follows Bayes' rule. Bayes' rule specifies the optimal update of probable representations of the states of the world in light of current sensory evidence. Therefore, perception is understood as a constructive process in which perceived sensations are used to confirm or disconfirm flexible hypotheses about how

these sensations were generated (Parr et al., 2022, p. 15f). By considering both the prior probability and the likelihood, *Bayesian inference* computes the probability of receiving the current sensory input under the present setting (Aitchison & Lengyel, 2017, p. 2). This probability represents personal expectations based on an individual's prior knowledge and beliefs, rather than the actual objective probability (Parr et al., 2022, p. 38).

In general terms, PC can be described as a hierarchical inference process in the brain (Garrido et al., 2009, p. 459), where an internal generative model of the world is used to predict sensory input based on current and past sensory experiences (Keller & Mrsic-Flogel, 2018, p. 425). Essentially, perception occurs when sensory information is integrated with our predictions based on a model of the origins of that sensory information (Garrido et al., 2009, p. 459). PC is built on the idea that it is often more efficient to represent the difference between a prediction about sensory input and the actual sensory input. This difference is referred to as the *prediction error* (Aitchison & Lengyel, 2017, p. 2). Prediction errors are used to adjust expectations regarding incoming stimuli based on prior stimuli, thereby contributing to the constant generation of updated predictions (Koelsch et al., 2019, p. 18). To enhance the quality of predictions, prediction errors ascend a hierarchy. In the case of auditory perception, this hierarchy could encompass the auditory brainstem and thalamus, primary auditory cortex, auditory association cortex, and frontal cortex (Koelsch et al., 2019, p. 1). In PC, it is emphasized that there is an asymmetry between the forward flow of information that computes residual errors and the backward flow of information that delivers predictions (Lupyan & Clark, 2015, p. 280). In some situations, deploying large amounts of prior knowledge is preferable, while it might be less efficient in others. In all circumstances, the focus is on minimizing the overall prediction error by adjusting the weight assigned to bottom-up inputs versus top-down expectations, thereby constantly re-estimating sensory uncertainty. These estimations of sensory uncertainty represent the brain's best guess at the degree of reliability of the sensory input predictions, and they modify how much impact prediction error signals should have at different levels of processing (Lupyan & Clark, 2015, p. 281).

This precision weighting is a mechanism that tunes how much incoming stimuli are modulated by top-down predictions (Lupyan & Clark, 2015, p. 281), and it implies

that it is not only the upcoming sensory input that needs to be predicted but also the degree of precision of this prediction, meaning that there are two kinds of descending predictions. These are referred to as first-order predictions: predictions of perceptual content, and second-order predictions: predictions about the precision of first-order predictions (Koelsch et al., 2019, p. 5). This serves as a way of assuring that the prediction errors have the right sort of influence when being processed. When prediction errors are precise, they contribute to expectation-updating higher in the hierarchy. On the other hand, if we are sitting in a bar with a lot of background noise and expect the precision errors to be unreliable, the imprecise prediction errors will be ignored (Koelsch et al., 2019, p. 5).

There is a consensus between predictive coding and Bayesian inference regarding the importance of combining both bottom-up and top-down processes in understanding perceptual processes. While Bayesian inference provides an optimal mathematical model for computing predictions as a framework to describe behavior, predictive coding explains the neural responses we are interested in when examining the mismatch negativity response.

Predictive Processes and the MMN

Two main hypotheses for explaining the neural mechanisms underlying the MMN are “adaptation” and “sensory memory/predictive coding” (Fong et al., 2020, p. 4). According to the *adaptation hypothesis*, the MMN rests on neuronal adaptation in the auditory cortex (Garrido et al., 2009, p. 454), and can be explained by observed changes in the N1, which is a negative electrical response that peaks 100 ms after stimulus presentation (Friston, 2005, p. 830). Due to neuronal adaptation, the N1 response to a repeated stimulus will be delayed and attenuated. This gives rise to the MMN, which results from subtracting the N1 response to a deviant stimulus from the N1 response to a repeated stimulus (Garrido et al., 2009, p. 454; Jääskeläinen et al., 2004, p. 6809). Thus, the negative deflection simply represents a non-adapted evoked response, indicating a “release” from suppression (Fong et al., 2020, p. 4). Meanwhile, a challenge to the adaptation hypothesis is that the N1 and MMN apparently originate from different sources (Friston, 2005, p. 830; Jääskeläinen et al., 2004, p. 6809). It also fails to explain why the MMN can be induced by an omission stimulus in cases where

an expected tone is absent, and no adaptation could have occurred (Fong et al., 2020, p. 4).

As already mentioned, the oddball paradigm makes use of standard and deviant stimuli, where a mismatch negativity response is expected to be elicited when the deviant stimulus is presented. In the meantime, no response to the standard stimulus is expected. This is because neural activation is diminished due to the repeated presentation of a stimulus, which is referred to as *repetition suppression* (RS) (Friston, 2005, p. 829; Mayrhauser et al., 2014, p. 1). From a predictive coding perspective, RS occurs because the processed input matches the prediction. Thus, the decreased activation for repeated (or standard) stimuli reflects a smaller prediction error for expected stimuli (Mayrhauser et al., 2014, p. 1). On the other hand, the presentation of a deviant sound that suddenly interrupts the otherwise repetitive sound sequence and, thereby, disrupts the prediction formed by the brain, will produce a considerable amount of prediction error. Thus, the MMN reflects the level of prediction error when an unexpected stimulus is presented (Fong et al., 2020, p. 5). On the other hand, to incorporate the notions of prediction of content and prediction of precision, the MMN elicited by deviant stimuli represents a prediction error where the processed input does not match the prediction of content (first-order prediction), even though that prediction is considered fairly precise (second-order prediction). Therefore, studies investigating MMN when subjects are aware that they will encounter unpredictable auditory input can be used to illustrate the distinction between first-order and second-order predictions (Koelsch et al., 2019, p. 5f).

Something quite interesting is that several studies have suggested that the MMN is not affected when participants are aware that a deviant stimulus is about to occur (Koelsch et al., 2019, p. 6). In a study by Ritter et al. (1999), a visual cue was used to indicate when a duration deviant was about to occur. Although the participants were made aware before the occurrence of the deviant stimuli, it did not influence the amplitude or latency of the MMN (Koelsch et al., 2019, p. 6; Ritter et al., 1999, p. 835). The same findings were reported in a study by Sussman et al. (2003), where a visual cue was used to indicate the occurrence of a pitch deviant (Koelsch et al., 2019, p. 6; Sussman et al., 2003, p. 634). Rinne et al. (2001) obtained the same results, even though the participants themselves controlled when they were exposed to deviant tones

by pressing a button (Koelsch et al., 2019, p. 6; Rinne et al., 2001, p. 2209). These findings indicate that, in relation to the MMN, the processing of prediction errors is not influenced by prior knowledge or beliefs (Koelsch et al., 2019, p. 6). However, the MMN is followed by different positive electrophysiological responses that reflect processes influenced by prior knowledge: the P3a and P3b. The P3a occurs around 250-300 ms after stimulus onset at anterior electrodes when deviant stimuli automatically attract attention. It has been observed to be significant only when the deviants occur without prior warning (Koelsch et al., 2019, p. 6; Sussman et al., 2003, p. 634). The P3b occurs around 300 ms or later at posterior electrodes when deviant stimuli are consciously detected. It has been observed to be significantly reduced due to prior knowledge (Koelsch et al., 2019, p. 6; Ritter et al., 1999, p. 835).

The predictive coding framework can be used to explain why prior knowledge does not modulate the prediction error signals reflected in the MMN (Koelsch et al., 2019, p. 6). One such explanation is presented by Koelsch et al. (2019), who suggest a novel formulation of predictive filtering (i.e. predictions about precision) (Koelsch et al., 2019, p. 5f). Caused by the repetition of a stimulus, the predictions of content and the precision of predictions change due to sensory learning. Neuronal populations responsible for the encoding are released from precision constraints, allowing the small prediction errors induced by standard stimuli to eliminate themselves more easily as the stimuli match the expectation. Hence, the elimination process relies on two components: 1) Sensory learning allows for more accurate predictions of content, and 2) the precision of constraints decreases as the precision of ascending prediction errors increases. This enables the ascending prediction errors to selectively match with representations that validate and further strengthen the predictions. When deviant stimuli are suddenly presented, the prediction errors they elicit “fall on deaf ears” because representations of input that do not match the predictions are attenuated. When the prediction error cannot eliminate itself, it will elicit an MMN (Koelsch et al., 2019, p. 6f). As such, sensory learning, which enables the strengthening of predictions regarding content and the precision of predictions, allows signals from standard stimuli, represented as minor and insignificant prediction errors, to be more effectively eliminated. Consequently, prior knowledge does not affect the prediction error signals reflected in the MMN.

Accordingly, the MMN can be elicited even when we are aware that we will encounter unpredictable auditory input (Koelsch et al., 2019, p. 6). As mentioned earlier, it can also be elicited in the absence of attention, providing us with a tool for testing change-detection abilities regardless of the direction of attention (Duncan et al., 2009, p. 1885; Fong et al., 2020, p. 4).

Purpose and Hypotheses

The purpose of this thesis is to explore the impact of background noise on the perception of speech sounds by examining the cortical auditory evoked potential known as Mismatch Negativity (MMN). Additionally, the role of increased stimuli complexity will be investigated with the aim of exposing participants to stimuli that more closely resemble everyday listening situations, rather than using completely unvarying stimuli. This thesis also seeks to understand the observed effects of background noise and stimuli complexity reflected in the MMN through the lens of predictive coding. Therefore, it addresses the following research question:

How does background noise and increased stimuli complexity affect the perception of speech sounds, and how can the underlying mechanisms be understood within a predictive coding framework?

Specifically, the first hypothesis in this study suggests that the MMN signal will decrease in amplitude when background noise is present compared to when there is no background noise. Another hypothesis posits that the MMN signal will decrease in amplitude as stimuli complexity increases. Furthermore, an interaction is hypothesized between the effects of background noise and increased stimuli complexity. In other words, they interact in a manner such that the reduction in MMN amplitude following background noise becomes even more pronounced with increased stimuli complexity. The final hypothesis anticipates an increase in the peak amplitude latency of the MMN in response to the presence of background noise. No specific hypothesis was formulated regarding changes in latency in response to increased complexity.

Materials and Methods

Participants

Datasets from 27 adults (10 women; mean age = 26.22 years, $SD = 3.23$) were included in the analysis. Exclusion criteria based on self-report included hearing impairment and a history of neurological disease. No participants were excluded due to these criteria. However, during the preprocessing of the EEG data, datasets from four participants were excluded, as described in more detail in the subsection *Eyeblink Corrections*. Three of the included participants were left-handed. 10 participants reported having received musical training beyond music lessons in school, and 13 reported playing a musical instrument. 17 participants reported having practiced their instrument(s), including choir and individual singing lessons, for a certain number of years, although some participants commented that they had not maintained or had lost their skills. Among the 17 participants, the minimum number of years of practice was two, and the maximum was 19 ($M = 7.94$, $SD = 5.74$). The study was conducted in accordance with the guidelines of the Norwegian Agency for Shared Services in Education and Research ('Sikt'). It received approval from the Regional Committee for Medical and Health Research Ethics for Western Norway (Reference Number: 2018/590).

Stimuli

In previous studies, various approaches have been employed to generate stimuli representing speech. Examples of these approaches include the use of the speech synthesizer OVE III B (Aaltonen & Suonpää, 1983, p. 410) or the Klatt synthesizer (Cheour-Luhtanen et al., 1995, p. 54). Another method for producing speech stimuli is semi-synthetic speech generation, where the waveform of the vowel is initially extracted from a real utterance and then used as input to an artificial vocal tract filter (Alku et al., 1999, p. 1329; Kozou et al., 2005, p. 34). As mentioned earlier, various types of noise have been used as background noise, including babble noise, industrial noise, traffic noise, and white noise (Kozou et al., 2005, p. 36). Therefore, the stimuli used in previous studies exhibited varying degrees of resemblance to everyday listening environments. Although the present study employed a limited selection of vowels and one specific type of background noise, the stimuli were chosen with the aim of closely resembling a real-life listening situation.

Background Noise

The background noise in the experiment is babble noise recorded in a café setting, where only the speech sounds are emphasized. The babble noise is re-synthesized so that it sounds human but is entirely unintelligible (Pigeon, 2020). It was created by research engineer and sound designer Dr. Ir. S. Pigeon (Pigeon, n.d.), who has kindly given permission to use it as background noise in this experiment. The sound file can be found on the Mynoise.net website, specifically under the ‘babble’ preset in the ‘Cafe Restaurant’ option (Pigeon, 2014). The babble noise was originally created as a tool to reduce the annoyance of intelligible speech. It is used to mask the noise from distracting conversations, such as those that occur in open office environments or similar settings. Babble noise is more efficient than white noise in masking speech sounds, enabling it to be presented at a lower volume while still effectively masking intelligible speech (Pigeon, 2020).

Vowel Stimuli

The vowel sounds, /a/ and /e/, used in the experiment, were pronounced by Professor of Linguistics Bruce Hayes (Hayes, 2013a) who has kindly given permission to use the sound files in this experiment. A chart showing the International Phonetic Alphabet (IPA) vowels, with the vowels applied in the experiment marked, can be seen in Figure 1 (Hayes, 2013b). The selection of vowel sounds was based on the criterion that they had to be easily discriminable to the participants. In the process of choosing suitable vowel sounds for the experiment, a small pilot study was conducted in which a few fellow students were asked to determine if the sounds were easily discriminable. Since the experiment took place in Norway, it involved individuals whose native language is Norwegian.

The Vowels						
	Front Unrounded	Front Rounded	Central Unrounded	Central Rounded	Back Unrounded	Back Rounded
Upper high	i	y	ɨ	ɥ	ɯ	u
Lower high	ɪ	ʏ				ʊ
Upper mid	e	ø	ɘ	ɵ	ɤ	o
Lower mid	ɛ	œ	ɜ	ɞ	ʌ	ɔ
Upper low	æ		ɐ			
Lower low	a _f	ɶ	a _c		ɑ	ɒ

Figure 1 – A chart displaying the International Phonetic Alphabet (IPA) vowels. The vowels used in the experiment are marked with red.

Editing of Vowel Stimuli

The primary tool for editing the vowel stimuli was *Audacity*, version 3.2.5, an audio editor that can be downloaded for free and contains the necessary tools for editing the sound files for this experiment (Audacity, 2023). Due to the very short and precise time span required for the vowel stimuli, complications arose while using Audacity. These issues were resolved by using the online tool *Audio Trimmer* to speed up the sound files (AudioTrimmer.com, 2023). Originally, the sound file for vowel /e/ had a duration of almost 420 milliseconds, and /a/ was nearly 610 milliseconds. To preserve the waveform of the vowel sounds, Audio Trimmer was used to speed up the sound files, reducing their length to just over 150 milliseconds. Subsequently, Audacity was employed to trim the few silent milliseconds at the beginning and end of each sound file, ensuring that each one had a total duration of 150 milliseconds. Once both sound files were adjusted to the correct length, they were saved and used for further editing.

When pitch is edited through Audacity, an estimated frequency of the original sound is automatically generated. For the original sound file for vowel /a/, the estimated frequency was 132.231 Hz, and for /e/, it was 141.593 Hz. The pitch was edited so that both /a/ and /e/ were presented at pitches A2 (with a frequency of 110 Hz) and C3 (with a frequency of 130.810 Hz). The illustration below presents the different pitches and their corresponding frequencies in Hz.

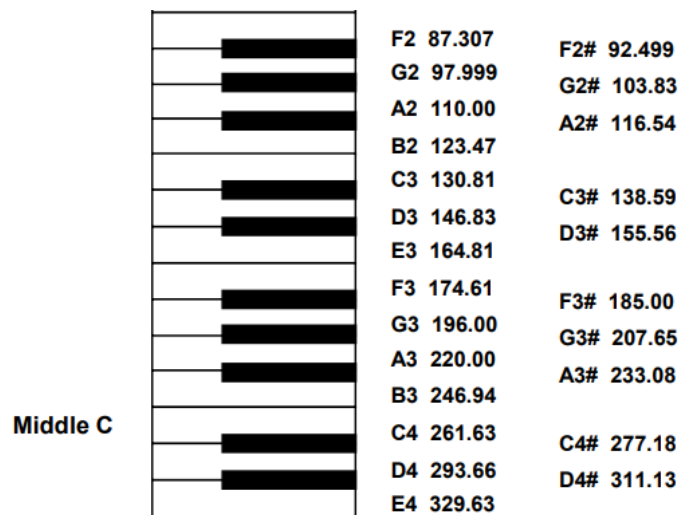


Figure 2 – Some tones, including A2 and C3, presented along with their corresponding frequencies in Hertz, illustrated together with the keys on a piano (Irvine, 2000).

Stimulus onset asynchrony (SOA) was adjusted by adding a silent pause of 300 milliseconds at the end of the sound files, creating versions with and without this additional break. When the sound files are presented in random order, this results in an asynchrony in the stimulus onset. Thus, the short SOA has no pause between sounds (the interstimulus interval¹), while the long SOA consists of the 300-millisecond pause. The location was edited by making half of the sound files play through the right speaker and the other half through the left speaker². Below are examples of how the processed sound files, ready for use in the experiment, are visualized in Audacity. The y-axis represents amplitude, and the x-axis shows the time in milliseconds.

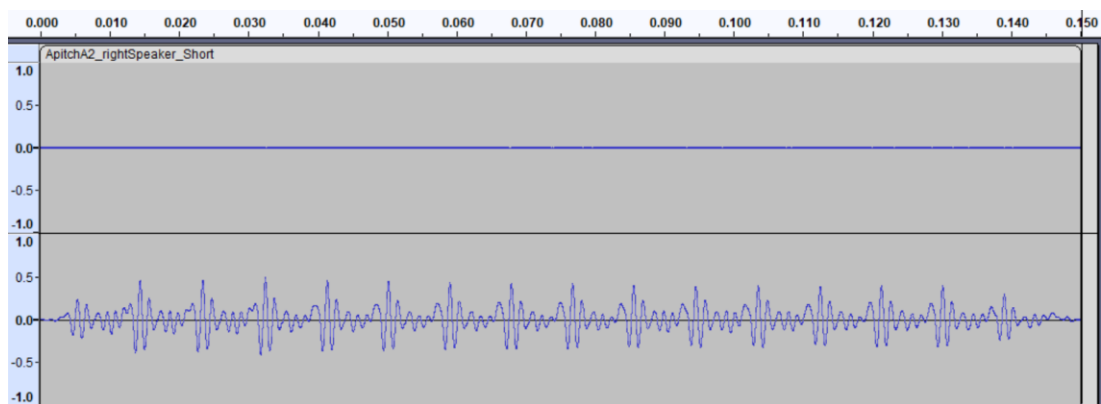


Figure 3 – Visualization of the sound file for timbre /a/, pitch A2, played from the right speaker. The total length of the sound file is 150 milliseconds.

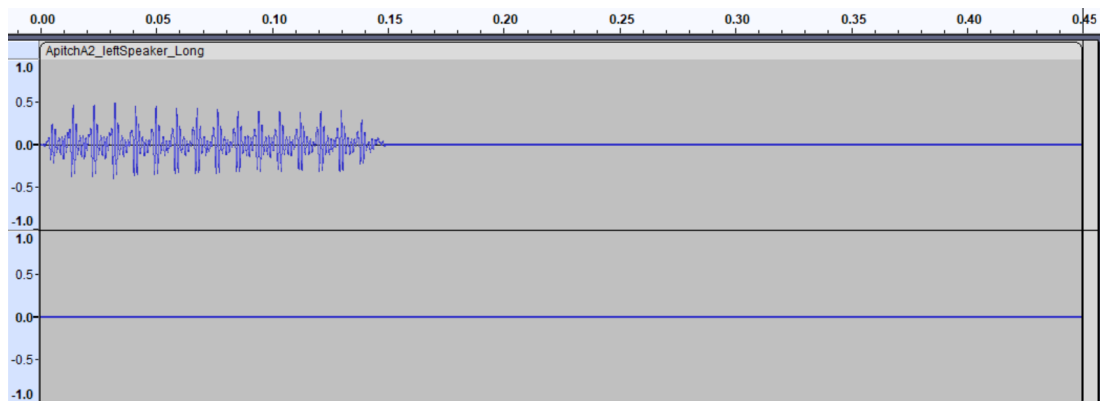


Figure 4 – Visualization of the sound file for timbre /a/, pitch A2, played from the left speaker. The vowel sound is 150 milliseconds with an added silent pause of 300 milliseconds. The total length of the sound file is 450 milliseconds.

¹ To be precise, the interval was set to 0.5 milliseconds in the script for the experiment.

² Even though it has no implications for further analysis in this thesis, it should be noted that the speakers playing the sound files were switched in the laboratory. Thus, the speakers were emitting sound from the opposite direction of what was indicated by the settings of the sound files.

Presentation of Vowel Stimuli

As described earlier, the mismatching stimuli used in the auditory oddball paradigm can differ on any discriminable auditory dimension (Duncan et al., 2009, p. 1885). In this experiment the stimuli have variations in terms of timbre, pitch, location, and SOA. Since we are interested in the ability to discriminate the vowel sounds /a/ and /e/ and, consequently, the MMN elicited by timbre-deviants, all variations of the different dimensions are presented at the same rate, except for timbre. This specifically means that pitch A2 is presented as frequently as pitch C3. The same applies to the left and right locations and the long and short SOAs.

A recommended recording condition for MMN is to present deviant stimuli with a probability between 10% and 20% (Duncan et al., 2009, p. 1886). Text box 1 offers an overview of the distribution of standard and deviant stimuli used in this experiment:

<i>The following distribution applies to both blocks in each permutation in the main experiment:</i>
Total amount of stimuli presentations (including standards, deviants, and castanet sounds): 1207
Castanet sounds: 8
Deviant stimuli: 120
Standard stimuli: $1207 - 8 - 120 = 1079$
Presentation of deviant stimuli in percent: $120/1079 * 100 = 11,12\%$

Text box 1 – Information about the distribution of standard and deviant stimuli. This information is applicable to both the first and second blocks in all eight permutations presented in Table 2 in the subsection *Presentation of Blocks*.

For timbre, the standard occurs frequently ($p = .89$), while the deviant occurs infrequently ($p = .11$). In half of the blocks, the deviant sound is /a/, and in the other half, it is /e/ (as shown in Table 2 in the subsection *Presentation of blocks*). However, each participant is presented with either /a/ or /e/ as the deviant throughout the experiment.

Procedure

The experiment consists of two parts: A discrimination test and the main experiment, both of which are presented with and without background noise. In the following, the different steps of the experiment and associated procedures will be described.

Sound Pressure Level

Before starting the experiment, the volume of the vowel stimuli and the background noise were separately measured to ensure that all participants were exposed to the same intensity. The volume was measured using a sound level meter (REED R8050 Sound Level Meter, Type 2, 30 to 130 dB). When measured at the position of the participant's head, the intensity of the background noise varied between 60 and 65 dB due to dynamic changes in the sound file. When measured with the sound pressure level meter, it was found that the speakers playing the vowel stimuli had slightly different volumes, with the speaker on the left-hand side of the participant playing slightly louder than the speaker on the right-hand side. When measured at the position of the participant's head, the noise from the left speaker was 61.5 dB, and the noise from the right speaker was 63 dB.

Questionnaire, Informed Consent, and Compensation Procedures

Before moving to the Faraday cage and starting the preparation of the EEG measurement by placing electrodes, the participants were welcomed in an adjacent room. Here, they received different documents that had to be read, filled out, and signed before proceeding with the study. They were provided with a document containing information about the study, participants' rights, and data handling, concluding with a declaration of consent that required their signature. Additionally, all participants received compensation of 200 Norwegian kroner for their participation and were required to sign a receipt confirming their payment.

All participants were provided with a questionnaire covering topics such as handedness, hearing loss, neurological and psychological disease, medication use, musical training, and language impairments. Most of the questions required simple 'yes' or 'no' responses, as illustrated in the questionnaire found in the Appendix. The answers were used to ensure that the EEG measurements were obtained from a neurotypical population with normal hearing based on self-report.

EEG-Recording

The experiment was conducted inside a Faraday cage. Participants sat in a chair facing a desk with a computer screen and keyboard. The vowel stimuli and background noise were presented through two distinct sets of speakers positioned on each side of the screen, forming an equal side triangle with the participant's seating position. The vowel stimuli came from a set of speakers positioned in the upper right and left corners of the desk where the participant was seated. The background noise was delivered through another set of speakers located just underneath the first set.

The EEG signal was recorded from 27 passive electrodes mounted in an EEG cap. These electrodes were positioned according to the international 10-10 system, and the recording was done at a sampling rate of 500 Hz using BrainAmps DC (Brain Products GmbH, Munich, Germany). Additionally, an extra electrode was placed on the left mastoid, serving as the reference during recording, while another electrode was placed on the back of the neck as a ground reference. All electrode impedances were kept below 20 k Ω . The placement of electrodes during recording is illustrated in Figure 5 below.

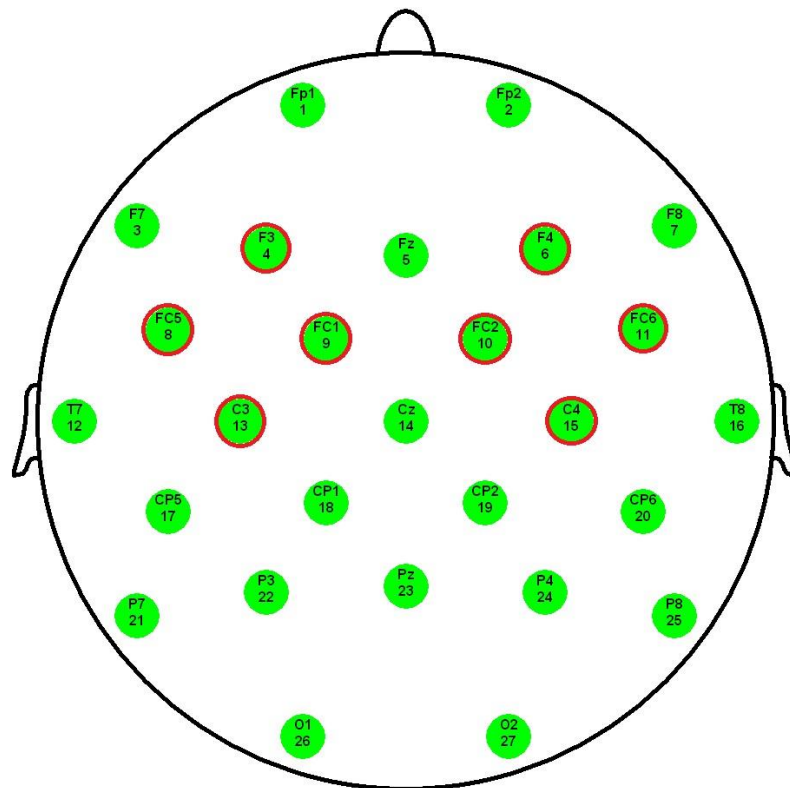


Figure 5 – Illustration of the electrode placement used for EEG recording (excluding the reference electrode at the left mastoid and the ground electrode on the neck). The electrodes marked with a red circle are the ones included in further analysis: F3, F4, FC5, FC1, FC2, FC6, C3, and C4.

Electroencephalographic recordings were obtained throughout the entire main experiment. To minimize artifacts in the EEG recording, participants were instructed to sit as still as possible, with particular attention to avoiding movements of their jaws and eyes (Tsogli et al., 2022, p. 8). Throughout the experiment, the experimenter was located in an adjacent room. Since it was possible to hear through the walls, participants were instructed to knock on the door or speak if they had any questions or felt uncomfortable in any way during testing. Additionally, the experimenter entered the Faraday cage during pauses between tests (three times throughout the entire experiment) to check on the participant's wellbeing or provide instructions for the next block.

Discrimination Test

In addition to the main experiment, the participants completed a discrimination test. In this test, participants were presented with various trials containing the same vowel stimuli used in the main experiment. To minimize the impact of order effects, the order in which participants were presented with different parts of the experiment was counterbalanced. This meant that half of the participants began with the discrimination test, while the other half began with the main experiment. The discrimination test aimed to ensure that the stimuli were acoustically distinguishable, specifically in terms of timbre. It was also employed to assess whether the stimuli could be distinguished acoustically based on the other parameters: pitch, SOA, and location.

The discrimination test consisted of 160 trials. After each trial, participants had to provide their answer and press the spacebar to proceed to the next trial. In each trial, sound pairs were presented with an interstimulus interval of 1 second. The second sound could either be the same as the first sound or different in terms of timbre, pitch, location, or SOA. For each trial, there would be a change in only one of these parameters. In the 40 trials representing the SOA, participants were presented with three sounds. In half of the trials, the pauses between the sounds were the same. In the other half, the pause between the second and third sound was either longer or shorter than the pause between the first and second sound. As described in the subsection *Editing of Vowel Stimuli*, a silent pause of 300 milliseconds was added to the sound files representing long SOA. Consequently, the short pauses consisted solely of the standard 1-second duration, while the additional 300 milliseconds of silence extended the long pauses to a total duration of 1.3 seconds. In addition to written instructions presented

on the screen during the experiment, participants also received oral instructions before the tests. They were asked to press the 'S' key on the keyboard in front of them if they believed both sounds were the same or 'L' if the last sound was different in one of the four parameters. For each of the four parameters, participants were presented with both options (same/different) 20 times, resulting in 40 trials for each parameter. With 40 trials for each of the four parameters, this amounted to a total of 160 trials.

The participants were presented with the 160 trials twice: once with background noise and once without. Here counterbalancing was achieved by alternating whether the background noise was present in the first or last block of 160 trials. At the end of one block of 160 trials, the number of correct responses was automatically displayed on the screen. Some participants reported that this served as extra motivation when completing the discrimination test for the second time.

Main Experiment

In the main experiment, participants encountered two distinct blocks with varying degrees of complexity. The simple block included a single pitch, location, and SOA, with a timbre deviant serving as an oddball. The complex block featured both pitch types, locations, and SOAs presented at the same rate, also with a timbre deviant as the oddball. Each block was presented twice: once with background noise and once without. To minimize order effects, each participant experienced one of eight permutations of the main experiment. An overview of the different permutations and the order in which each participant was presented with the different trials can be seen in the following subsection, *Presentation of Blocks*. The duration of the complex blocks is always 06:07 minutes in total because the long and short SOAs are presented an equal number of times. The simple blocks consist of only one of the two types of SOA and are thus either 09:07 minutes in duration or 03:07.

This means that when the simple block consisted of only the long SOA, the main experiment had a total duration of 30:28 minutes, not counting the small pauses in between blocks. When the simple block consisted only of short SOA, the total duration was 18:28 minutes. Figure 6 illustrates the order of test blocks as they were presented to participant number 1 (who was presented with permutation 1):

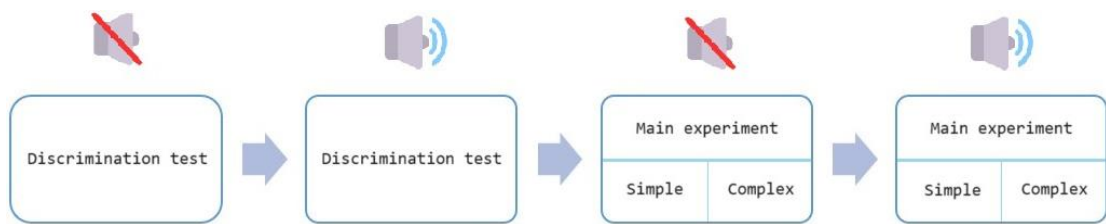


Figure 6 – Illustration of the order in which blocks with and without background noise were presented, based on the sequence applied for participant number 1.

The main experiment began with instructions displayed on the screen. During the EEG recording, participants were asked to watch a silent movie while the auditory stimuli were presented. The silent movie was the same documentary about birds as used in Tsogli et al. (2022). The movie was used to provide some additional stimuli, both to create a lifelike listening situation in which participants were exposed to more than pure auditory input and to prevent them from falling asleep during the rather monotonous experiment. At the same time, the movie was expected not to be too arousing or cognitively demanding, and it did not contain any presentation of letters or speech sounds (Tsogli et al., 2022, p. 8). It has actually been pointed out that the degree to which visual stimuli are being attended to does not seem to influence the MMN. In addition, it has been argued that to observe the MMN, it is optimal to create a condition where attention is directed away from the stimulus to avoid overlap with other ERP components (Garrido et al., 2009, p. 455). During each block of the main experiment, a total of eight castanet sounds were presented, and these were the only stimuli that participants were instructed to actively pay attention to. Participants were asked to press the spacebar every time they heard a castanet sound. The purpose of this task was to keep the participants alert throughout the test blocks without them directing their attention specifically to the vowel stimuli.

Presentation of Blocks

The participants were presented with the discrimination test and the main experiment, both with and without noise, in different orders, which was done for counterbalancing purposes. Below, you will find an overview of the order of experimental blocks for all participants, as well as an overview of the different permutations of the main experiment.

Participant number	Main experiment permutation	Test order (main experiment or discrimination test)		Background noise	
		First	Second	First	Last
1	1	Discr	Main		x
2	2	Discr	Main	x	
3	3	Main	Discr		x
4	4	Main	Discr	x	
5	5	Discr	Main		x
6	6	Discr	Main	x	
7	7	Main	Discr		x
8	8	Main	Discr	x	
9	1	Main	Discr		x
10	2	Main	Discr	x	
11	3	Discr	Main		x
12	4	Discr	Main	x	
13	5	Main	Discr		x
14	6	Main	Discr	x	
15	7	Discr	Main		x
16	8	Discr	Main	x	
17	1	Discr	Main		x
18	2	Discr	Main	x	
19	3	Main	Discr		x
20	4	Main	Discr	x	
21	5	Discr	Main		x
22	6	Discr	Main	x	
23	7	Main	Discr		x
24	8	Main	Discr	x	
25	1	Main	Discr		x
26	2	Main	Discr	x	
27	3	Discr	Main		x
28	4	Discr	Main	x	
29	5	Main	Discr		x
30	6	Main	Discr	x	
31	7	Discr	Main		x

Table 1 – Overview of the order in which each participant was presented with the different blocks of the experiment. For instance, Participant 1 first completed the discrimination test without background noise and then the discrimination test with background noise. Subsequently, permutation 1 of the main experiment was presented without background noise, followed by the presentation with background noise. This scheme includes all tested subjects. Subjects not included in further analysis (numbers 3, 6, 7, and 26) are marked in dark grey.

Permutation 1						
Block	Simple/Complex	Duration (minutes)	Timbre deviant	SOA long/short	Location Left/right	Pitch A2/C3
1 st	Simple	9.07	/e/	Long	Right	A2
2 nd	Complex	6.07	/e/			
Permutation 2						
1 st	Complex	6:07	/a/			
2 nd	Simple	3.07	/a/	Short	Right	A2
Permutation 3						
1 st	Simple	9.07	/e/	Long	Right	C3
2 nd	Complex	6:07	/e/			
Permutation 4						
1 st	Complex	6.07	/a/			
2 nd	Simple	3.07	/a/	Short	Right	C3
Permutation 5						
1 st	Simple	3.07	/e/	Short	Left	C3
2 nd	Complex	6:07	/e/			
Permutation 6						
1 st	Complex	6.07	/a/			
2 nd	Simple	9.07	/a/	Long	Left	C3
Permutation 7						
1 st	Simple	3.07	/e/	Short	Left	A2
2 nd	Complex	6:07	/e/			
Permutation 8						
1 st	Complex	6.07	/a/			
2 nd	Simple	9.07	/a/	Long	Left	A2

Table 2 – Overview of the various permutations of the main experiment. Please note that for the complex conditions, the specific SOA, location, and pitch are not indicated, as both types are presented equally. Depending on the duration of the permutation and the participant's pace in the discrimination test, testing typically lasted for approximately one hour, with a variance of about 10 minutes.

Eyeblink Corrections

The EEG data preprocessing was conducted using the *EEProbe* software package (NITRC, 2011). Visual inspection revealed that the MMN response occurred within different time windows in conditions with background noise (also referred to as noise conditions) and conditions without background noise (also referred to as silent conditions). Consequently, the time window for statistical analysis was set to 160-220 ms after stimulus onset for silent conditions and 220-280 ms for noise conditions.

As mentioned in the subsection *Presentation of Vowel Stimuli*, each block of the main experiment contained 120 deviant triggers. If more than 20 of these triggers were rejected due to artifacts, subjects were excluded. This criterion was chosen to ensure a sufficient level of ERP signals induced by deviant triggers for inclusion in the averaging process. While participants were instructed to minimize movement and blinking, some movement will occur. To prevent these movements from affecting the results,

triggers were automatically excluded if they coincided with blinks or other movements. Given the frequent occurrence of blinks, algorithms have been developed to eliminate them and preserve the data when triggers and blinks overlapped (Parker, 2006). Since no electrodes were applied to record EOG activity, this could not be employed in the classification of eyeblinks (A.N.T. Software BV, 2003). Instead, the process of correcting trial rejections to ensure that all trials rejected due to blinks would be included in the averaging relied on the visual inspection of the waveforms mainly from the frontal electrodes Fp1 and Fp2. This eyeblink correction process allowed us to retain some subjects who would have otherwise been excluded. This means that, for the subjects included in the analysis, at least 100 triggers for the deviant stimuli were averaged in each of the four blocks of the main experiment. Four subjects were excluded because this criterion was not met. On average, for the excluded subjects, the total number of ERP signals that were not rejected due to artifacts was 80.01 ($SD = 12.77$) out of 120 irregular trials, representing 66.67% ($SEM = 5.36\%$) of the irregular trials, and 703 ($SD = 144.86$) out of 1207 regular trials, representing 58.25% ($SEM = 6.0\%$) of the regular trials. On average, for the included subjects, a total of 110.93 ($SD = 5.05$) out of 120 irregular trials was included in the average of the ERPs, corresponding to 92.44% ($SEM = 0.81\%$) of the irregular trials, and 975.01 ($SD = 77.01$) out of 1207 regular trials, corresponding to 80.79% ($SEM = 1.23\%$) of the regular trials.

Results

Behavioral Data

In addition to the main experiment, a discrimination test was conducted to ensure that the stimuli were distinguishable on all parameters that could vary in the main experiment: timbre, pitch, location, and SOA (see subsection *Discrimination Test*). Participants accurately distinguished whether the last sound in a sequence was the same or different, with an average score of 90.3% ($SEM = 1.72\%$) in the absence of background noise and 86.39% ($SEM = 1.9\%$) in the presence of background noise. While average scores were notably high in both conditions, a statistically significant difference was observed between the condition without background noise ($M = 144.48$, $SD = 14.29$, $N = 27$) and the condition with background noise ($M = 138.22$, $SD = 15.81$, $N = 27$) as revealed by a two-sided paired samples t-test ($t(26) = 3.65$, $p = .001$).

Pilot testing indicated that participants had difficulty distinguishing between the sequences designed to test the acoustic distinguishability of stimuli based on SOA. Excluding the SOA-related trials (40 out of 160 trials), participants correctly differentiated sound sequences with an average accuracy of 97.1% ($SEM = 2.49\%$) in the absence of background noise and 91.1% ($SEM = 1.96\%$) in the presence of background noise. A two-sided paired samples t-test revealed a significant difference between the condition without background noise ($M = 116.48$, $SD = 15.51$, $N = 27$) and the condition with background noise ($M = 109.33$, $SD = 12.19$, $N = 27$) ($t(26) = 2.99$, $p = .006$). Thus, despite the impact of background noise on discrimination ability, the discrimination test indicates that participants were able to discriminate the stimuli effectively, with even better results when excluding the SOA-related trials.

Electroencephalographic Data

For the sake of clarity, a brief summary of the data used in the analysis is provided. The data included in the analysis comprises ERPs recorded in response to timbre standards and deviants. These ERPs were obtained from the electrodes F3, F4, FC5, FC1, FC2, FC6, C3, and C4 (see Figure 5). The primary objective of the statistical analysis is to investigate the impact of stimuli complexity and background noise on the MMN. In the following section, the EEG data will be visually presented using violin plots before further presenting the results of the statistical analysis.

Distribution of Data

In the subsection *Interactions* below, a bar graph will be used as visualization for explaining the MMN reductions between different conditions since it shows the relationship between the mean values very clearly (Newman & Scholl, 2012, p. 601). To improve transparency before moving on to the statistical analysis of the EEG data, the data will be presented using violin plots combined with dot plots, which are used to create an informative visualization of the data and to provide a clear presentation of the data distribution (Weissgerber et al., 2019, p. 1507) (see Figure 7 below).

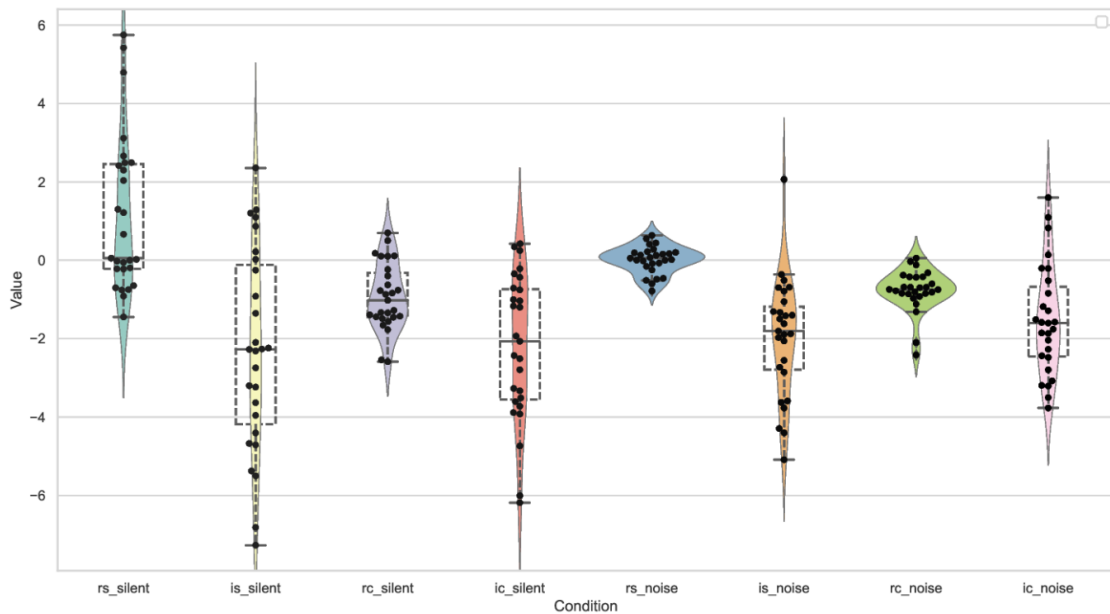


Figure 7 - Violin plots, combined with dot plots, for each condition. The x-axis represents the different conditions, while the y-axis displays the ERP signal values in microvolt. Abbreviations are explained as follows: r (regular), i (irregular), s (simple), and c (complex).

In Figure 7, the black dots depict the individual data points. Symmetric jittering is used to make all the data points visible (Weissgerber et al., 2019, p. 1508). One data point reflects the mean value of the ERP signal recorded across the eight electrodes included in the analysis for one subject in the given condition. Each colored figure depicts the distribution of data using density curves. The very flat curves, most evident in the simple condition for irregular stimuli without background noise (is_silent), reveal that the points are widely spread out, indicating a large standard deviation. Additionally, when examining the complex condition for irregular stimuli without background noise (ic_silent), there is also a slight indication of a bimodal distribution.

The horizontal lines in the box plot represent the 25th, 50th (median), and 75th percentiles. The 50% of the data points lie within the lower and upper bounds indicated by the 25th and 75th percentiles, which is known as the interquartile range. In the simple condition for regular stimuli without background noise (rs_silent), the median is noticeably closer to the lower quartile, suggesting that the data set for this condition is skewed to the right (positively skewed). The whiskers depict the most extreme data points that are not considered outliers (Weissgerber et al., 2019, p. 1508), which reveals two outliers in both the simple and complex conditions for regular stimuli with

background noise (rs_noise and rc_noise), as well as one outlier in the simple condition for irregular stimuli with background noise (is_noise).

Statistical Analysis of EEG-Data

For a statistical analysis of the EEG data, a GLM ANOVA for repeated measurements with the within-subject factors – *presence of background noise* (silence vs. noise), *complexity of stimuli* (simple vs. complex), *regularity of stimuli presentation* (regular vs. irregular), and *lateralization* (left vs. right) – was conducted. For explorative purposes, lateralization was included in the statistical analysis, although it has no direct relevance to the research question. For this purpose, electrodes were divided into two groups, where electrodes F3, FC1, FC5, and C3 correspond to the left side of the head, while electrodes F4, FC2, FC6, and C4 correspond to the right side of the head. This analysis revealed no significant main effect of lateralization ($F(1, 26) = 1.73, p = 0.20$) and no significant interaction effect between the regularity of stimuli presentation and lateralization ($F(1, 26) = 2.48, p = 0.128$), indicating that in this experiment, there is no significant lateralization of the MMN.

In the following sections, the focus will be on the factors *regularity of stimuli presentation* (regular vs. irregular), *complexity of stimuli* (simple vs. complex), and *presence of background noise* (silence vs. noise). These are the factors relevant for assessing whether the current data supports the hypotheses of the thesis. Initially the main effects will be presented followed by a presentation of the interactions and post hoc tests.

Main Effects

In this section, the main effects of regularity of stimuli presentation, complexity of stimuli, and presence of background noise will be presented. This entails examining how each of these independent variables influences the dependent variable, which is the ERP-values.

The first effect we are interested in examining is the effect of the regularity of stimuli presentation (regular vs. irregular), which indicates the mismatch negativity. As explained in the subsection *Mismatch Negativity and the Oddball Paradigm*, the MMN is calculated by subtracting the event-related response elicited by the presentation of regular stimuli from the response elicited by irregular stimuli. In this regard, the GLM

ANOVA for repeated measurements revealed a significant main effect for the regularity of stimuli presentation ($F(1, 26) = 75, p < .001$). It is important to note that this analysis compares all ERPs of regular events with all ERPs of irregular events, irrespective of whether they are part of a simple or complex block or with or without background noise. Therefore, the main effect represents the results obtained when considering all eight electrodes across all blocks and conditions.

In the GLM ANOVA for repeated measurements, when examining the main effect of the complexity of stimuli, a significant difference is observed between conditions with simple and complex stimuli ($F(1, 26) = 14.93, p = .001$). However, the presence of background noise, as compared to no background noise, does not significantly impact the overall ERPs within the MMN time window, as reflected in the main effect for the presence of background noise ($F(1, 26) = .003, p = .955$).

Interactions

In this section, we will first examine whether the MMN is affected by noise and complexity. Subsequently, the results for the three-way interaction between regularity of stimuli presentation, complexity of stimuli, and the presence of background noise will be presented.

First, when assessing the impact of background noise on the MMN, we analyze the interaction between the presence of background noise and the regularity of stimuli presentation. In this analysis, the GLM ANOVA for repeated measurements reveals a significant interaction ($F(1, 26) = 6.27, p = .019$), confirming that noise does indeed affect the MMN. Next, in our investigation of the influence of stimuli complexity on the MMN, we examine the interaction between the complexity of stimuli and the regularity of stimuli presentation, which also shows a significant interaction ($F(1, 26) = 21.07, p < .001$), confirming that stimuli complexity affects the MMN.

Furthermore, we expected that the reduction of the MMN due to noise would be more pronounced in the condition with increased stimuli complexity than in the simple condition. To investigate this, we need to examine the interaction between the presence of background noise, the complexity of stimuli, and the regularity of stimuli presentation. However, this three-way interaction does not show a significant effect when tested as a two-sided hypothesis, as is the case here ($F(1, 26) = 3.27, p = .082$). While

this result does not meet the threshold for statistical significance (considering a statistical threshold of $p = .05$), it does reveal a trend in the data.

Since we hypothesized that the MMN signal would decrease in amplitude when background noise is present compared to when there is no background noise, and that a further decrease in MMN amplitude would be observed as stimuli complexity increases, it is interesting to explore if there are any indications of this being the case, even though the three-way interaction only revealed a trend. To support our expectations, background noise should reduce the MMN reduction between simple and complex conditions. In other words, if the reduced MMN is further reduced, it aligns with the direction that supports our hypothesis.

To investigate if this is the case, we will start by illustrating the MMN of the different conditions through the bar chart in Figure 8 below.

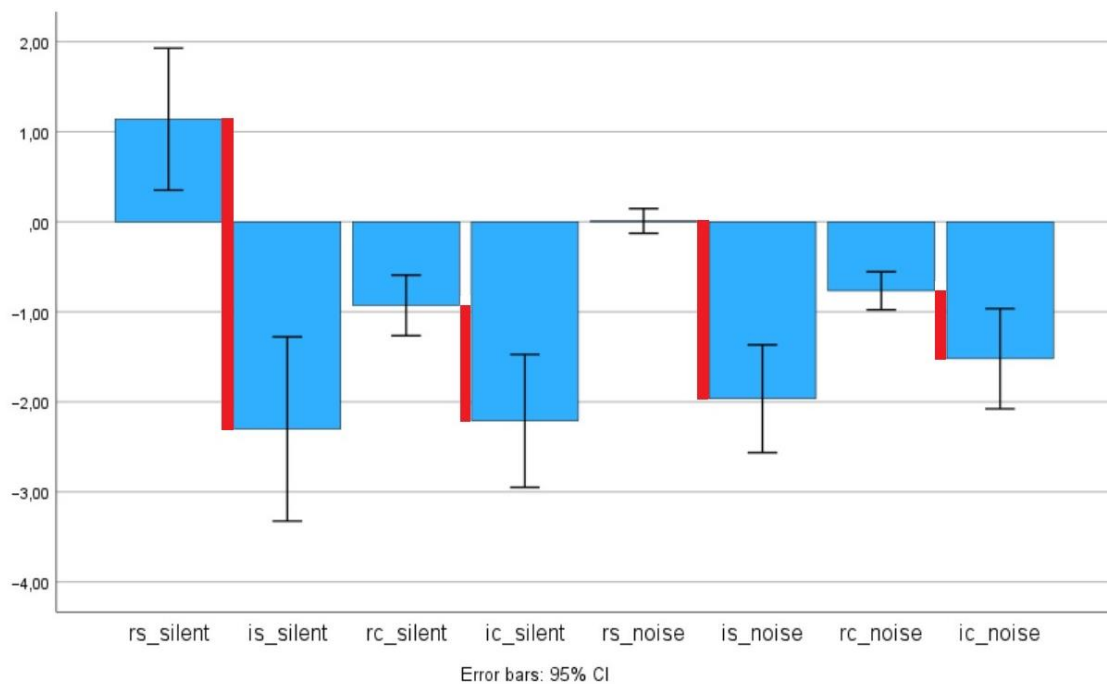


Figure 8 – Bar chart representing the mean of the ERPs across the eight electrodes and all subjects. On the x-axis, you can see the different conditions, while the y-axis displays the ERP signal value in microvolt. The red lines indicate the size of the MMN in the various conditions. Abbreviations are explained as follows: r (regular), i (irregular), s (simple), and c (complex). The presence or absence of background noise is denoted by ‘noise’ or ‘silent’. For instance, ‘rs_silent’ signifies regular stimuli in a simple condition with no background noise. In the silent conditions (without background noise), the values are based on the time window 160-220 milliseconds after stimulus onset, representing the average ERP amplitude in this time frame. In the noise conditions (with background noise), the values are based on the time window 220-280 milliseconds after stimulus onset, as mentioned in subsection *Eyeblink Corrections*.

Since the MMN is derived from the difference between the ERPs of regular and irregular stimuli, the MMN in the different conditions is visualized by the red lines. Therefore, the first red line in Figure 8 indicates the difference between the ERPs of regular and irregular stimuli in the simple condition with no background noise. For a clearer visual comparison, the values represented by the red difference-lines indicating the MMN in Figure 8 are also presented in Figure 9 below.

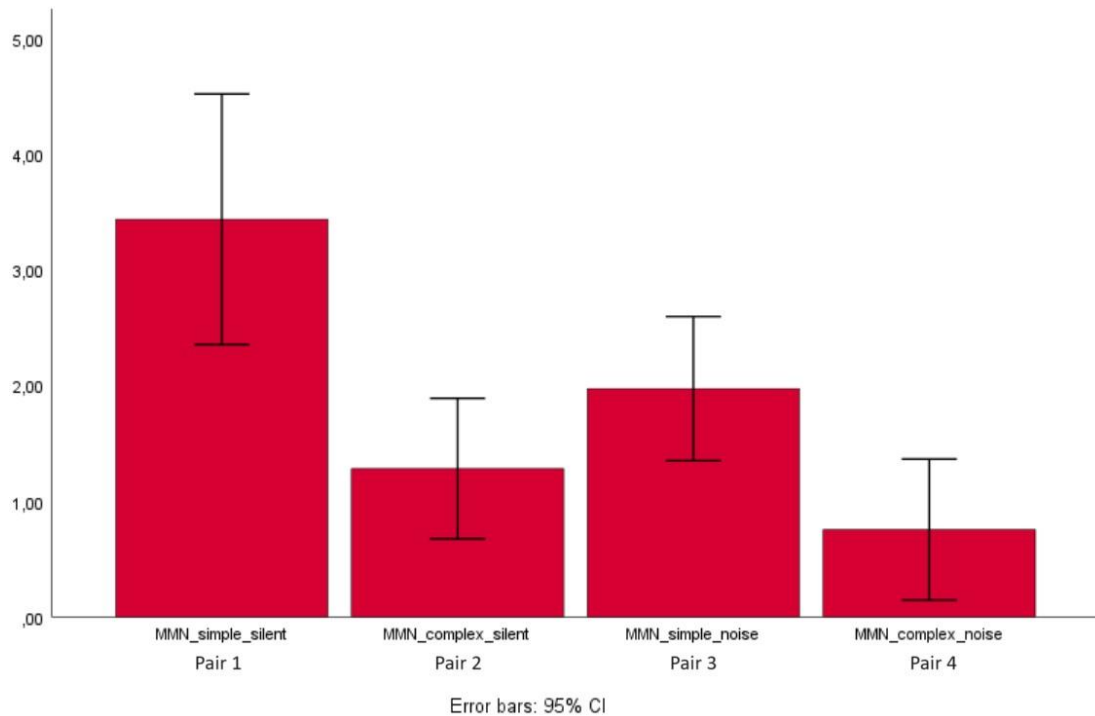


Figure 9 – Red bars indicating the values of the MMN signal, also illustrated as the red difference lines in Figure 8. On the x-axis are the different conditions. The y-axis shows the value of the MMN signal in microvolt. Referring to the labels of the different conditions used in Figure 8, Pair 1 shows the difference between rs_silent and is_silent (simple condition without background noise), Pair 2 shows the difference between rc_silent and ic_silent (complex condition without background noise), Pair 3 shows the difference between rs_noise and is_noise (simple condition with background noise), and Pair 4 shows the difference between rc_noise and ic_noise (complex condition with background noise).

Post-hoc t-tests were conducted to compare the MMN values and investigate the trend indicated by the three-way interaction. The aim was to determine whether the trend aligns with expectations, specifically if background noise leads to a further reduction in the MMN amplitude between simple and complex conditions. When comparing the values for Pair 1 and Pair 2 in a two-sided paired samples t-test, a significant difference was observed ($t(26) = 3.99, p < .001$). Moreover, the MMN amplitude decreased by $-2.16 \mu\text{V}$, indicating that within conditions without background noise, the

MMN amplitude decreased from the simple to complex condition. When comparing the values for Pair 3 and Pair 4, a significant difference is found ($t(26) = 3.64, p = .001$). Additionally, there is a decrease in the MMN amplitude by $-1.22 \mu\text{V}$, which demonstrates that within the conditions with background noise, the MMN also decreases from simple to complex. This indicates that the reduction of the MMN from simple to complex is greater during silence than in noise. To investigate the difference in MMN between the complex conditions with and without background noise, Pair 2 and Pair 4 are compared. In this case, there is a decrease in the MMN amplitude by $-0.53 \mu\text{V}$, but the difference is not statistically significant ($t(26) = 1.60, p = .122$). Thus, the interaction aligns with expectations, showing that the reduction of MMN in the presence of background noise is stronger in the complex condition compared to the simple condition. However, it also indicates that no significant difference in MMN amplitude is detected when adding background noise to complex stimuli.

ERP Waveforms

The results described above, which indicate that the reduced MMN observed when going from the simple to the complex condition is further reduced in the presence of background noise, can also be illustrated by examining the ERP waveforms of the different conditions shown in Figure 10.

The red line represents the average of the ERPs elicited by irregular stimuli, the blue line represents the average of the ERPs elicited by regular stimuli, and the black line represents the MMN, which is the difference between the ERPs of irregular and regular stimuli. While the bars in Figure 8 illustrate the mean values of the ERPs in each condition within the specified time window, Figure 10 displays the waveform from 100 milliseconds before stimulus onset to 300 milliseconds after stimulus onset. When examining the black line representing the MMN (similar to the red lines in Figure 8), it is evident that it exhibits the same reductions as described in relation to the bar chart in Figure 8.

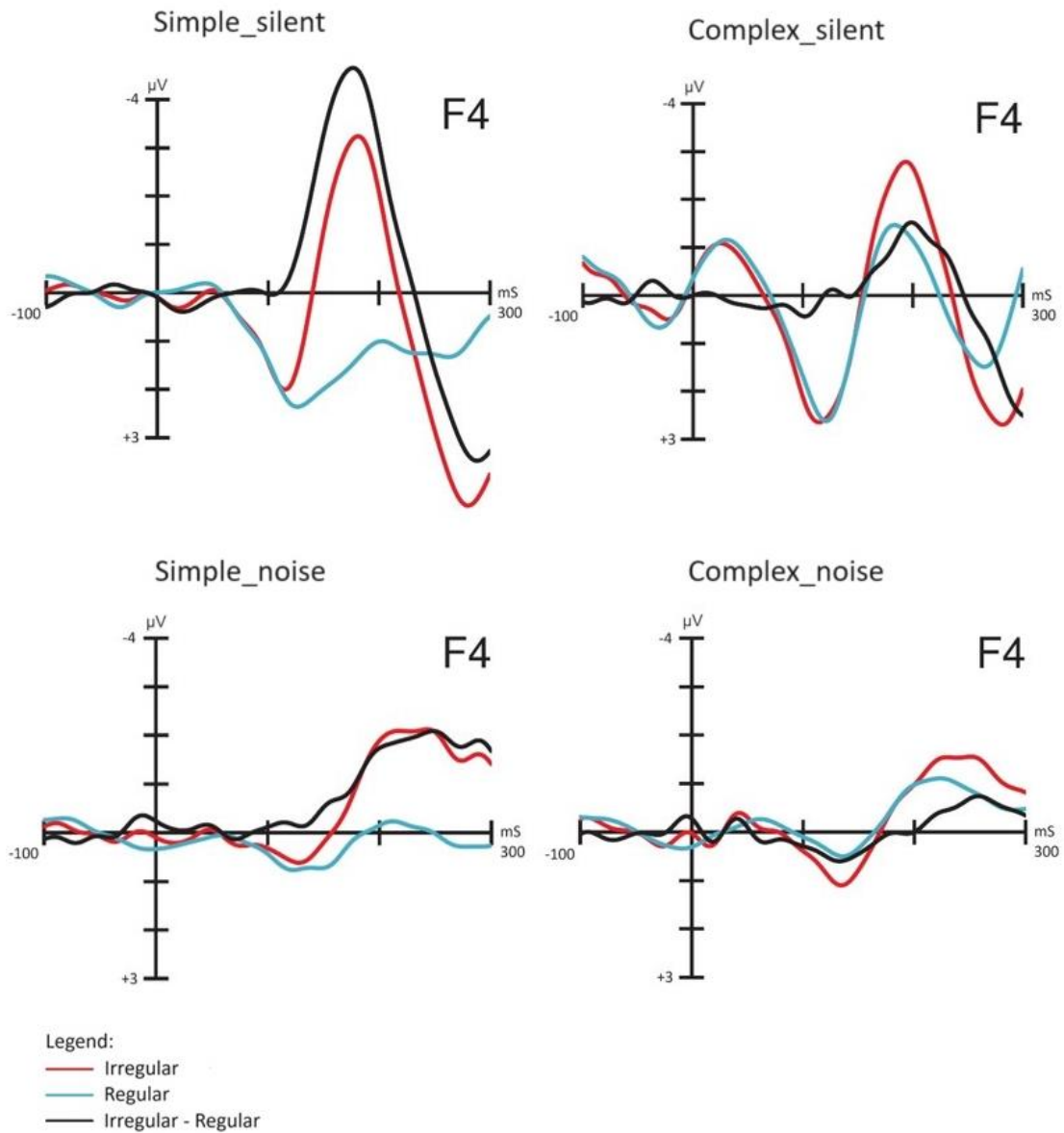


Figure 10 – Illustration of ERP waveforms for electrode F4 across all subjects in the different conditions. As previously mentioned, the time window for statistical analysis was set to 160-220 ms after stimulus onset for the silent conditions and to 220-280 ms for the noise conditions. The black line represents the MMN.

Changes in Latency

As indicated by the waveforms in Figure 10, the peak amplitude latency is longer in the conditions with background noise than in the conditions without background noise. This is the reason two different time windows were chosen for analyzing the MMN with and without background noise (220-280 and 160-220 ms after stimulus onset, respectively). The average peak amplitude latency in the different conditions is provided in Figure 11 below.

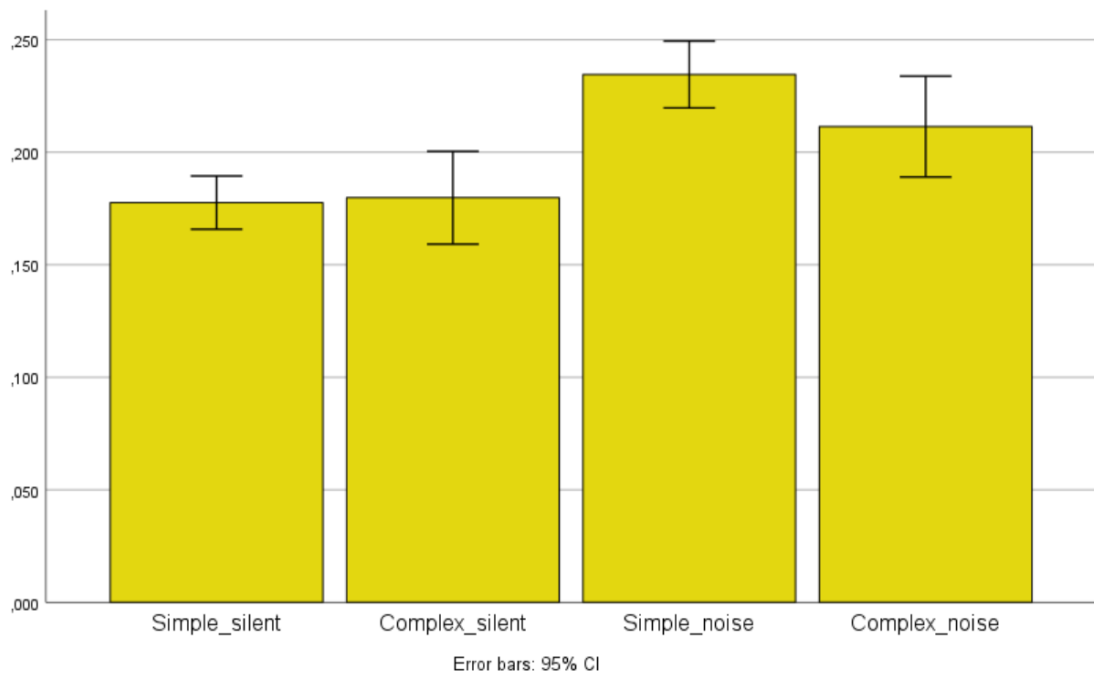


Figure 11 – Yellow bars representing the peak amplitude latency of the MMN in the different conditions. The x-axis represents different conditions, while the y-axis represents time in milliseconds after the stimulus onset. These bars display the mean values across the 21 subjects included. The specific mean values for the different conditions are as follows: Simple_silent ($M = 0.178$), Complex_silent ($M = 0.180$), Simple_noise ($M = 0.235$), and Complex_noise ($M = 0.211$).

Only subjects who exhibited a clear peak within a time window of 100-300 ms after stimulus onset were included in this analysis. This resulted in the inclusion of latency values for 21 subjects. In a two-sided paired samples t-test, no significant difference was found between the MMN peak latency amplitude for simple and complex stimuli within the silent condition ($t(20) = -0.200, p = 0.844$). Similarly, there was no significant difference between simple and complex stimuli within the noise condition ($t(20) = 1.838, p = 0.081$). When comparing the peak latency amplitude for complex stimuli in both the silent and noise conditions, no significant difference was found ($t(20) = -1.971, p = .063$). A significant difference was found when comparing the simple stimuli in both the silent and noise conditions ($t(20) = -6.612, p < .001$). When comparing all latency values obtained in the silent condition with all those obtained in the noise condition, a significant difference was found ($t(20) = -4.940, p < .001$). This suggests that background noise significantly impacts MMN latency, while increased stimuli complexity does not.

What Has a Greater Impact on MMN: Background Noise or Increased Stimuli Complexity?

Additional post hoc t-tests were conducted to investigate whether there was a discernible difference in the impact on the MMN when comparing the presence of background noise and increased stimuli complexity. We first compared the values labeled as 'rs_silent', 'rc_silent', and 'rc_noise' in Figure 8. Subsequently, we compared 'rs_silent', 'rs_noise', and 'rc_noise'.

When performing a two-sided paired samples t-test, a significant difference was found when comparing rs_silent and rc_silent ($t(26) = 5.214, p < .001$), while the difference was not significant when comparing rc_silent and rc_noise ($t(26) = -1.060, p = .299$). This indicates a significant difference as an effect of increased stimuli complexity, while there is no significant effect of background noise being added. When comparing rs_silent to rs_noise, a significant difference was found ($t(26) = 0.004, p = 0.008$), and a significant difference was also found when comparing rs_noise to rc_noise ($t(26) = 7.093, p < .001$). This indicates a significant difference as an effect of the presence of background noise and, in addition, a significant effect when stimuli complexity is increased. When comparing the same conditions but with irregular stimuli, no significant differences were found. Altogether, these differences indicate that, when considering the regular stimulus trials, the increased complexity of stimuli has a greater impact on the ERP signal than the presence of background noise.

To determine if the patterns observed in the regular ERPs were also reflected in the MMN, the values of the MMN for different conditions (as illustrated in Figure 9) were compared using two-sided paired samples t-tests. A significant difference was found when comparing MMN_simple_silent and MMN_complex_silent ($t(26) = 3.989, p < .001$). However, the difference was not significant when comparing MMN_complex_silent and MMN_complex_noise ($t(26) = 1.599, p = .122$). This indicates a consistent pattern with a significant difference attributed to increased stimuli complexity, while the addition of background noise does not have a significant effect. When comparing MMN_simple_silent and MMN_simple_noise, a significant difference was found ($t(26) = 2.506, p = 0.019$), and a significant difference was also found when comparing MMN_simple_noise and MMN_complex_noise ($t(26) = 3.639, p = 0.001$). This also reflects the same pattern with a significant difference due to the presence of

background noise and a significant effect when stimuli complexity is increased in addition. The variations observed in the regular ERPs are mirrored in the MMN, indicating minimal variability in the ERPs to the irregular stimuli.

Are the MMN Reductions Mainly Derived from Changes in ERPs to Regular or Irregular Stimuli?

To investigate whether the MMN reductions mainly derive from changes in the ERPs elicited by regular stimuli, irregular stimuli, or changes in both types of stimuli, additional two-sided paired samples t-tests were conducted. In the following, the term ‘pair’ refers to the same pairs described in relation to Figure 8 and 9.

The data for pair 1 (rs_silent – is_silent, $M = 3.44$) and pair 2 (rc_silent – ic_silent, $M = 1.28$) reveal a marked decrease in MMN when transitioning from the simple to the complex condition without the presence of background noise. It is important to note that this reduction in MMN can primarily be attributed to differences in ERPs elicited by regular stimuli. This is evident because the ERPs in response to irregular stimuli in the simple and complex conditions, without background noise (is_silent, $M = -2.30$ and ic_silent, $M = -2.21$), do not display statistically significant differences when compared in a paired samples t-test ($t(26) = -0.212$, $p = 0.834$). In contrast, the ERPs in response to regular stimuli in the simple and complex conditions, without background noise (rs_silent, $M = -1.14$ and rc_silent, $M = -0.93$), exhibit a significant difference between them ($t(26) = 5.214$, $p < 0.001$).

The values for pair 3 (rs_noise – is_noise, $M = 1.97$) and pair 4 (rc_noise – ic_noise, $M = 0.75$) demonstrate that within conditions where background noise is present, the MMN decreases when transitioning from the simple to the complex condition. This reduction is primarily influenced by changes in the ERPs elicited by regular stimuli. This is evident because the ERPs in response to regular stimuli in the simple and complex conditions with background noise (rs_noise, $M = 0.01$ and rc_noise, $M = -0.77$) differ significantly from each other ($t(26) = 7.093$, $p < 0.001$). In addition, the comparison of the ERPs in response to irregular stimuli in the simple and complex conditions with background noise (is_noise, $M = -1.97$ and ic_noise, $M = -1.52$) does not reveal a statistically significant difference ($t(26) = -1.314$, $p = 0.200$).

The values for pair 1 (rs_silent – is_silent, $M = 3.44$) and pair 3 (rs_noise – is_noise, $M = 1.97$) indicate a reduction in MMN attributed to the presence of background noise. This reduction is primarily influenced by changes in the ERPs elicited by regular stimuli. This is evident because the ERPs in response to regular stimuli in the simple condition, both with and without background noise (rs_silent, $M = 1.14$ and rs_noise, $M = 0.01$), differ significantly from each other ($t(26) = 2.850, p = 0.008$). Additionally, the ERPs in response to irregular stimuli in the simple condition, with and without background noise (is_silent, $M = -2.30$ and is_noise, $M = -1.97$), do not exhibit significant differences ($t(26) = -0.577, p = 0.569$).

The values for pair 2 (rc_silent – ic_silent, $M = 1.28$) and pair 4 (rc_noise – ic_noise, $M = 0.75$) indicate a reduction in MMN attributed to the presence of background noise. This reduction primarily appears to be associated with changes in the ERPs in response to irregular stimuli, as suggested by visual inspection of the blue bars in Figure 8. However, when conducting a two-sided paired samples t-test, no statistically significant differences are found. The ERPs in response to regular stimuli in the complex condition, both with and without background noise (rc_silent, $M = -0.93$ and rc_noise, $M = -0.77$), do not significantly differ from each other ($t(26) = 0.150, p = 0.299$). Similarly, the ERPs in response to irregular stimuli in the complex condition, with and without background noise (ic_silent, $M = -2.21$ and ic_noise, $M = -1.52$), do not reveal a statistically significant difference ($t(26) = -1.929, p = 0.065$).

In the comparison of the following pairs, 1+2, 3+4, and 1+3, significant differences were found within the ERPs elicited by the regular stimuli and not the irregular. When visually inspecting the comparison of pair 2+4, a notable difference was observed within the ERPs elicited by the irregular stimuli, but no significant difference was found. This indicates that variations in MMN amplitude are primarily attributed to changes in the ERPs elicited by the standard stimuli.

Discussion

Effects on MMN-Amplitude

The aim of the current study was to investigate the impact of background noise and stimuli complexity on the perception of speech sounds within an oddball paradigm. Neurophysiological responses were assessed to determine whether the presence of background noise and an increase in stimuli complexity would impede the perception of speech sounds. It was expected that the MMN amplitude would decrease in the presence of background noise compared to conditions with no background noise. The results demonstrated a significant effect of background noise on the perception of speech sounds, as evidenced by a reduction in the MMN signal, supporting previous findings presented in the subsection *Background Noise and the MMN*. The presence of background noise reduces the audibility of the signal. The decrease in MMN amplitude recorded in response to the presence of background noise can be explained by the reduction in audibility, which leads to deterioration in pre-attentive speech discrimination, as also suggested by Muller-Gass et al. (2001). Additionally, it was expected that the MMN signal would decrease in amplitude with increased stimuli complexity. The results revealed that increased stimuli complexity had a significant effect on the perception of speech sounds, reflected in a reduction of the MMN. This is supported by previous studies that reported results related to different components of increased stimuli complexity, such as increased rhythmic complexity (Lumaca et al., 2019, p. 1604), random SOAs (Tsogli et al., 2022, p. 16), and consonant-vowel stimuli as opposed to much simpler pure-tone stimuli (Wunderlich & Cone-Wesson, 2001, p. 1528).

It was expected that the presence of background noise would reduce MMN amplitude, with this effect becoming more pronounced as stimuli complexity increased in addition to the presence of background noise. While the three-way interaction between background noise, complexity of stimuli, and regularity of stimuli presentation did not produce a statistically significant result, a trend was evident in the data, suggesting a potential relationship between background noise and increased stimuli complexity in influencing the MMN. This suggests an interesting objective for future research.

The trend observed in the three-way interaction suggests that MMN is reduced due to the presence of background noise, and this reduction becomes more pronounced

when the complexity of stimuli is increased. For example, when comparing the values for the different conditions depicted in Figure 9, it was evident that MMN amplitudes were larger in conditions with no background noise than in conditions with background noise, and they were also larger in conditions with simple stimuli than with complex stimuli. Post-hoc t-tests were conducted to determine whether the reductions in MMN resulted from changes in the ERPs elicited by regular or irregular stimuli. This analysis aimed to establish whether the differences in MMN were primarily due to reductions in the ERPs for standard sounds or if they were also influenced by a reduction in the ERPs for deviant sounds (also known as error signals). When comparing different pairs of ERPs, significant differences were observed within the ERPs elicited by the regular stimuli, except when comparing MMN signals recorded in complex conditions with and without background noise. This suggests that, in general, error signals were less affected than the ERPs for standard sounds.

Post-hoc tests were conducted to determine whether the presence of background noise or increased stimuli complexity had a greater impact on the ERPs. No significant differences were found when comparing conditions with irregular stimuli, which can be explained by the earlier findings indicating that the ERPs in response to standard stimuli were primarily affected. However, among conditions with regular stimuli, a significant difference was found due to increased stimuli complexity, while the addition of background noise did not have a significant effect. Additionally, a significant difference was observed with the presence of background noise, and a significant effect was also noted with the addition of stimuli complexity. The pattern of these differences suggests that increased stimuli complexity has a more pronounced impact on the ERP signal than the presence of background noise.

Nevertheless, as supported by previously mentioned studies (Kozou et al., 2005; Martin et al., 1999; Muller-Gass et al., 2001; Salo et al., 1995), it is evident from the neurophysiological responses that background noise does impact pre-attentive auditory discrimination reflected in a smaller MMN amplitude. Although the present study did not include behavioral measures to explore whether these results align with previous findings that the MMN is only elicited when stimuli are behaviorally discriminable (Kraus et al., 1996, p. 971; Martin et al., 1999, p. 271; Näätänen, 2001, p. 7), some behavioral data were collected from the discrimination test. It was found that, despite

a high average score for correctly discriminated sound pairs, a significant difference was observed between the conditions with and without background noise, indicating that the presence of background noise significantly reduced the ability to behaviorally discriminate between the sounds. However, it should be noted that the results of the discrimination test do not directly shed light on the relationship between MMN and behavioral discrimination ability, as the discrimination test does not reflect the way stimuli are presented in the main experiment.

Effects on MMN-Latency

Referring to the first study that reported an MMN response to changes in phoneme stimuli conducted by Aaltonen et al. (1987), they employed the phonemes /y/ and /i/, along with the boundary stimulus /yi/. This can be interpreted as a slight increase in stimulus complexity, as the boundary stimulus /yi/ was less easily discriminable. The presentation of this boundary stimulus resulted in a reduction in MMN amplitude and an increase in latency (Aaltonen et al., 1987, p. 197; Näätänen, 2001, p. 7). In the present study, an increase in MMN latency was also observed. However, this prolonged latency did not appear to be primarily attributed to an increase in stimuli complexity but, rather, to the presence of background noise. This is evident as significant differences in latency were found when comparing the simple stimuli in the silent condition to the simple stimuli in the noise condition, or when comparing all values obtained in the silent conditions to all the values obtained in the noise conditions. In comparing our results to the longer latency reported by Aaltonen et al. (1987), one could argue that it is not the slight increase in complexity that should be emphasized in the difference between /i/ and /yi/, but rather how easily discriminable they are compared to /y/. When we observe a significantly longer latency for the conditions with background noise compared to those without, it may simply reflect that the stimuli in the conditions with background noise are less easily discriminable due to reduced audibility.

Furthermore, as described in the subsection *Behavioral Data*, the discrimination test was conducted to ensure that the stimuli were distinguishable across all parameters that could vary in the main experiment: timbre, pitch, location, and SOA. Therefore, increased complexity is not incorporated into the individual stimuli, as in studies where pure-tone stimuli are replaced with consonant-vowel stimuli or short words to

introduce greater complexity (Wunderlich & Cone-Wesson, 2001, p. 1528). Instead, the stimuli used in this experiment are easily distinguishable across all parameters. Complexity is introduced by presenting stimuli in an unpredictable manner, involving either a single change (only the timbre-deviant) or unpredictable alterations in pitch, location, and SOA in addition to the timbre-deviant. While Salo et al. (1995) observed that white noise masking did not affect MMN latencies, the findings of longer latency in response to background noise obtained in this study are supported by studies using other types of masking noise, such as different levels of high-pass-filtered sounds (Martin et al., 1999, p. 271) or speech noise (Muller-Gass et al., 2001, p. 197).

Findings Understood in a Predictive Coding Perspective

The findings in the study indicate that the presence of background noise and increased complexity of stimuli affect the MMN signal, resulting in a lower amplitude and longer latency. As mentioned in the previous subsection *Effects on MMN-Latency*, the increase in stimulus complexity is not implemented into the target stimuli themselves but rather in their presentation. In the simple conditions, the only variation in the pattern of stimuli presentation is the timbre-deviant, which occurs 11% of the time. In the complex conditions, complexity increases by introducing unpredictable variations in pitch, location, and SOA, each with a 50% likelihood of occurrence in addition to the timbre-deviant. This increased complexity, combined with background noise, is expected to make it more challenging to predict the next sound.

In a predictive coding framework, the MMN amplitude reflects the size of neural prediction error elicited by timbre deviations (Lumaca et al., 2019, p. 1604), indicating smaller prediction errors in response to the presence of background noise and increased stimuli complexity, as demonstrated by the reduction in MMN amplitude. As described in the subsection *Predictive Coding*, the brain forms a predictive model of future events and estimates the reliability of predictions within that model. The less reliable the model, the smaller the neural responses to stimuli that deviate from it (Lumaca et al., 2019, p. 1598). This explains the progressive decrease in prediction error observed as background noise is introduced and stimuli complexity increases.

As introduced previously, there are two different kinds of descending predictions related to the reliability of predictions and the predictions themselves: first-order

predictions, which pertain to predictions of perceptual content, and second-order predictions, which involve predictions about the precision of first-order predictions (Koelesch et al., 2019, p. 5). In the simple conditions where all components are highly predictable, and the monotonous pattern is only occasionally interrupted by deviant stimuli, the predictions of perceptual content can be considered highly precise. This results in a significant prediction error in response to deviant stimuli. When background noise is added to this condition, it affects the prediction of the precision of first-order predictions, resulting in less weight being assigned to these predictions. Violations of less precise predictions are more effectively ignored, preventing them from contributing to model updating. Violations of predictions regarded as more precise result in a larger prediction error, which contributes to model updating. This is reflected in the present study by the large MMN amplitude observed in the simple condition without background noise.

While the effects of increased complexity and the presence of background noise can both be explained by reduced predictability of the stimuli pattern and an increase in attentional demand, differences in how they affect predictability can be identified. The conditions with increased complexity of stimuli presentation introduce a less predictable pattern, while background noise affects the audibility of the signal. In this context, background noise might be expected to cause a more pronounced reduction in MMN amplitude than stimuli complexity. This expectation arises because background noise can reduce the audibility of the target stimulus since it is presented simultaneously and can also interfere by adding to the pattern of the speech sound stimuli. A way to explain why background noise appears to have a lesser impact on prediction errors than increased stimuli complexity is through the difference in sound characteristics between the stimuli presented and the background noise. The characteristics of the background noise differ more from the timbre-deviants than the sounds presented in the repetitive stream of standard stimuli, causing it to have a lesser effect on the expected sound pattern in comparison to an increase in complexity of standard stimuli. Consequently, it can be expected that background noise affects prediction errors primarily due to reduced audibility, resulting in less accurate predictions, while an increase in stimuli complexity introduces a less predictable pattern. This explanation aligns with both the expected outcome and observed data, showing the largest

reduction in MMN amplitude when background noise accompanies increased stimuli complexity.

Post-hoc t-tests revealed that variations in MMN amplitude are primarily associated with changes in ERPs evoked by standard stimuli. This association can be attributed to varying attentional demands across conditions. Increased attentional demands, stemming from background noise and stimuli complexity, lead to a decrease in the precision of predictions, subsequently affecting the ERP amplitude evoked by standard stimuli. The impact of noise on the precision of predictions is clearly demonstrated in the ERPs for standard stimuli (rs_silent vs. rs_noise in Figure 8). Although it impedes the precision of predictions, it may still be sufficient to establish a stable model, ensuring that when deviant sounds are detected, the error signal remains consistent (as observed in is_silent vs. is_noise). The model is less precise, yet it remains accurate enough to generate a full error signal from the deviant stimuli. However, the MMN itself is affected by reduced predictability. As stimuli become more complex, attentional demands increase even further, resulting in a less precise model, as evidenced by the diminished MMN values. One limitation of the study is the absence of behavioral data to investigate whether discriminative ability decreases as the error signal weakens. This aligns with previous findings that suggest the MMN signal diminishes when stimuli are less discriminable (Aaltonen et al., 1987, p. 197; Näätänen, 2001, p. 7).

An interesting finding from a predictive coding perspective is that, despite the increased complexity of auditory input when noise was added to the complex conditions, predictability did not significantly differ (rc_silent vs. rc_noise in Figure 8). It is interesting from a PC perspective that the MMN changes despite identical predictability. This warrants discussion since it may appear to challenge the entire PC framework if predictive models with the same degree of stability yield different MMN values. However, this observation can be explained by the fact that noise alters precision, and complexity affects attentional demands. While predictability remains identical, attention is distributed to various stimulus components such as different locations, pitches, SOAs, and timbres, fine-tuning precision and resulting in varying MMN responses. Interestingly, this precision tuning appears to have a greater impact on the ERPs elicited by the standard stimuli compared to the deviant stimuli. In summary, ERPs elicited by the

standard stimuli reflect the stability of the predictive model and seem to be more affected by background noise and increased stimuli complexity than the error signals.

Methodological Considerations and Future Research

While a lot of consideration has gone into creating a noise environment that resembles everyday listening situations, the ecological validity of the observations should be questioned. First and foremost, in terms of speech stimuli, the focus has been exclusively on monophthongs. This means that the implemented vowel sounds /a/ and /e/ are presented in a very simple fashion, characterized by a lack of detectable changes (Colman, 2015, p. 275). This does not closely resemble a typical conversation, which consists of a complex stream of sounds. When examining individual vowel sounds in speech, they are often not presented as monophthongs but can vary in tonal quality, eliciting the perception of pitch and timbre (Purves et al., 2013, p. 396). Furthermore, the babble noise used as background has the advantage of simulating a setting that closely resembles real-life situations where speech perception is interfered with by the voices of other people. However, the babble noise used in this experiment differed from everyday situations in that it was completely unintelligible. Nevertheless, this was prioritized as a necessity to ensure that the results were not affected by participants being disturbed by listening to intelligible speech in a manner that could not be controlled when evaluating the test results. Furthermore, the babble noise had the property of efficiently masking intelligible speech since it was developed for this purpose. While one could argue that it might not significantly impact the results of this experiment, as it involves isolated fragments of speech rather than intelligible speech, it is worth considering whether background noise designed to mask speech might have resulted in a greater reduction of the MMN signal than using a regular, non-resynthesized recording of babble noise. However, this raises the question of using background noise consisting of intelligible speech. It could be expected that background noise with intelligible speech would affect speech signal processing more, as informational masking containing meaningful information is more cognitively demanding than pure energetic masking (Harmon et al., 2021, p. 1104; Meekings et al., 2016, p. 8).

In relation to the background noise developed with the purpose of being a masker, it is also important to consider that in the presence of background noise, neural responses have been shown to be more robust when participants are exposed to vowel

stimuli than when exposed to consonant stimuli. This difference in impact is related to the different properties of consonants and vowels, with consonants appearing more transient and aperiodic, whereas vowels are more steady (Koerner et al., 2016, p. 41). Thus, the chosen background noise might have masked the target stimuli very efficiently if it had been chosen to investigate the impact on the perception of consonant sounds instead of vowel sounds. If the participants had been unable to extract the consonant sounds from the background noise, no MMN would have been elicited. The experimental design employed here might then only permit drawing conclusions about the impact of background noise compared to no background noise and the influence of simple stimuli versus complex stimuli. However, if background noise completely eliminates the MMN response in both simple and complex conditions when background noise is present, no discernible effects or trends would be observed in the three-way interaction involving the combination of background noise and stimuli complexity affecting the MMN. Since consonants are considered the primary carriers of information in speech (Purves et al., 2013, p. 396), it would be interesting to include them in a subsequent study. This would not only contribute to the results already obtained in this area but also allow for an investigation into whether the findings using this experimental setup align with previous research.

Through post-hoc tests, an attempt was made to investigate whether the presence of background noise or increased stimuli complexity had varying impacts on the MMN. The results indicated that increased stimuli complexity had a more substantial effect on reducing MMN amplitude than the presence of background noise. Regarding increased latency, it was suggested that background noise had a more significant impact than increased stimuli complexity. However, these trends provide insights specific to the data obtained under the circumstances and stimuli used in this study. Different types or higher levels of background noise, as well as more complex stimuli, may yield different outcomes. For example, increased noise levels could have a more pronounced effect on the MMN, and depending on the stimuli used, the same reduction in MMN amplitude might be achieved at lower noise levels. These are also factors worth considering in future studies.

Despite the considerations regarding the choice of background noise and target stimuli, it is essential to bear in mind that when examining ERPs, you are investigating

time-locked brain activity in response to specific stimuli (Ashley & Timmers, 2017, p. 50). This requires a controlled environment to obtain valuable data. Thus, it should be emphasized that the results do provide valuable information about processes related to vowel processing in noise, as reflected in the electrical brain response. However, it is important to note that this data is obtained in a controlled laboratory environment using a set of very simple stimuli. These results are valuable because they provide insights into the perception of specific speech components under particular circumstances. This, in turn, contributes to a better understanding of the larger picture.

After implementing potential improvements and conducting additional tests with participants who have normal hearing, it would be interesting to apply a research design similar to the one used in this thesis for clinical purposes. Considering that speech perception in noise is even more challenging for individuals with hearing impairments (Legris et al., 2018, p. 336), it would be valuable to investigate how background noise and increased stimuli complexity affect people with hearing impairments. If the complications in speech perception primarily result from background noise, which reduces audibility and impairs the ability to discriminate speech sounds, one would not expect to observe further reductions in MMN amplitude when increasing stimuli complexity in the presence of certain levels of background noise. Additionally, it is worthwhile to investigate the performance of various hearing aid systems in the presence of background noise to determine whether the use of a hearing aid results in an increase in MMN amplitude in response to deviant sounds, thus providing an objective measure of improved auditory discrimination ability.

A new hearing aid is currently under development, which, among other features, aims to consider the location of the speech signal of interest and the ability to adjust how the hearing aid processes sound before delivering it to the ear based on the specific noise environment (Oculaudio, 2023). Using the terminology applied in this thesis, the primary goal of this emerging hearing aid technology is to enhance the audibility of speech signals in noisy environments. This enhancement reduces the complexity of incoming stimuli and enhances predictability, thus making it easier for people with hearing impairments to follow conversations despite background noise. It would be highly valuable to include this technology in future studies.

Conclusion

This thesis aimed to explore how background noise and increased stimuli complexity affect the perception of speech sounds and how these mechanisms can be understood within a predictive coding framework. To do this, several hypotheses were examined.

The first hypothesis posited that the MMN signal would decrease in amplitude in the presence of background noise, and this hypothesis was supported by the results. The presence of background noise did indeed lead to a reduction in the MMN signal, affirming the impact of noise on the perception of speech sounds. The second hypothesis suggested that the MMN signal would decrease in amplitude as stimuli complexity increased, which was also supported by the results. Increased stimuli complexity had a significant effect on the perception of speech sounds, resulting in a reduction of the MMN signal, which emphasizes the importance of considering stimuli complexity when studying the neural processing of speech sounds. The third hypothesis proposed an interaction between background noise and increased stimuli complexity, with the expectation that the reduction in MMN amplitude following background noise would become more pronounced with complex stimuli. While this interaction did not yield a statistically significant result, a trend was observed in the data, indicating a potential relationship between background noise and increased stimuli complexity. Understanding nuances of this interaction could provide valuable insights into how the brain processes complex auditory information in noisy environments. The final hypothesis anticipated an increase in the peak amplitude latency of the MMN in response to the presence of background noise. The results supported this hypothesis, suggesting that background noise significantly affects the timing of neural responses to speech sounds. No significant effects were found of increased stimuli complexity on the peak amplitude latency of the MMN.

Within the framework of predictive coding, the reduction in MMN amplitude can be understood as a reflection of the decreased prediction error induced by timbre deviations. Essentially, the brain forms predictive models of future events and assesses the reliability of these predictions. The less reliable the model due to factors like background noise and stimuli complexity, the smaller the neural responses to deviations from it. Increased attentional demands arising from background noise and stimuli

complexity led to a decrease in the precision of predictions, impacting the ERP amplitude evoked by standard stimuli. This highlights the role of attention and prediction in shaping neural responses to speech sounds in varying conditions.

Looking ahead, there are exciting avenues for future research. Despite the simplicity of the stimuli used in this experiment, these findings provide insights into the neural processes behind speech perception in noisy environments. A useful direction for future research involves conducting more studies with individuals who have hearing impairments, as speech perception in noisy environments presents a particular challenge for this population. Exploring emerging hearing aid technologies designed to improve audibility in noisy environments could provide valuable insights into enhancing auditory discrimination abilities for individuals with hearing impairments.

References

- Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, 24(3), 197–207. [https://doi.org/10.1016/0301-0511\(87\)90002-0](https://doi.org/10.1016/0301-0511(87)90002-0)
- Aaltonen, O., & Suonpää, J. (1983). Computerized Two-Dimensional Model for Finnish Vowel Identifications. *Audiology*, 22(4), 410–415. <https://doi.org/10.3109/00206098309072801>
- Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46, 219–227. <https://doi.org/10.1016/j.conb.2017.08.010>
- Alain, C., & Winkler, I. (2012). *Recording Event-Related Brain Potentials: Application to Study Auditory Perception* (pp. 69–96). Springer New York. https://doi.org/10.1007/978-1-4614-2314-0_4
- Alku, P., Tiitinen, H., & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology*, 110(8), 1329–1333. [https://doi.org/10.1016/S1388-2457\(99\)00088-7](https://doi.org/10.1016/S1388-2457(99)00088-7)
- A.N.T. Software BV. (2003). *EEProbe Reference Manual*. <http://eeprobe.ant-neuro.com/Pdf/EEProbe-manual.pdf>
- Ashley, R., & Timmers, R. (2017). *The Routledge Companion to Music Cognition*. Taylor and Francis. Chapter 1 & 5. <https://doi.org/10.4324/9781315194738>
- Audacity. (2023). *Audacity*. <https://www.audacityteam.org/download/>
- AudioTrimmer.com. (2023). *Audio Trimmer*. <https://audiotrimmer.com/audio-speed-changer/>

- Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and Language*, *126*(1), 1–41. <https://doi.org/10.1016/j.bandl.2013.03.003>
- Breedlove, S. M., & Watson, N. V. (2017). *Behavioural neuroscience* (Eighth Edition). Sinauer Associates, Inc. Chapter 18
- Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. [https://doi.org/10.1016/0378-5955\(94\)00164-L](https://doi.org/10.1016/0378-5955(94)00164-L)
- Colman, A. M. (2015). *A dictionary of psychology* (Fourth edition.). Oxford University Press.
- Cooper, N. P., & Guinan, J. J. (2006). Efferent-mediated control of basilar membrane motion. *The Journal of Physiology*, *576*(1), 49–54. <https://doi.org/10.1113/jphysiol.2006.114991>
- Cunningham, J., Nicol, T., King, C., Zecker, S. G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing Research*, *169*(1), 97–111. [https://doi.org/10.1016/S0378-5955\(02\)00344-1](https://doi.org/10.1016/S0378-5955(02)00344-1)
- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. <https://doi.org/10.1016/j.clinph.2009.07.045>

- Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., & Magnasco, M. O. (2000). Essential nonlinearities in hearing. *Physical Review Letters*, *84*(22), 5232–5235. <https://doi.org/10.1103/PhysRevLett.84.5232>
- Fong, C. Y., Law, W. H. C., Uka, T., & Koike, S. (2020). Auditory Mismatch Negativity Under Predictive Coding Framework and Its Role in Psychotic Disorders. *Frontiers in Psychiatry*, *11*, 1–14. <https://doi.org/10.3389/fpsy.2020.557932>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>
- Gazzaniga, M. S., Ivry, R., B., & Magnun, G., R. (2014). *The Cognitive Neuroscience: The Biology of the Mind* (4. ed.). W. W. Norton & Company.
- Harmon, T. G., Dromey, C., Nelson, B., & Chapman, K. (2021). Effects of Background Noise on Speech and Language in Young Adults. *Journal of Speech, Language, and Hearing Research*, *64*(4), 1104–1116. https://doi.org/10.1044/2020_JSLHR-20-00376
- Hayes, B. P. (2013a). *Bruce P. Hayes*. <https://linguistics.ucla.edu/people/hayes/>
- Hayes, B. P. (2013b). *Vowel Chart with Sound Files*. <https://linguistics.ucla.edu/people/hayes/103/Charts/VChart/>
- Irvine, T. (2000). *Vibrationdata Piano Page*. <https://www.vibrationdata.com/tutorials2/piano.pdf> and <https://www.vibrationdata.com/piano.htm>

- Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Bellevue, J. W., & Desimone, R. (2004). Human Posterior Auditory Cortex Gates Novel Sounds to Consciousness. *Proceedings of the National Academy of Sciences - PNAS*, *101*(17), 6809–6814. <https://doi.org/10.1073/pnas.0303760101>
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron (Cambridge, Mass.)*, *100*(2), 424–435. <https://doi.org/10.1016/j.neuron.2018.10.003>
- Koelsch, S. (2011). Toward a neural basis of music perception—A review and updated model. *Frontiers in Psychology*, *2*, 1–20. <https://doi.org/10.3389/fpsyg.2011.00110>
- Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. <https://doi.org/10.1016/j.tics.2018.10.006>
- Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, *339*, 40–49. <https://doi.org/10.1016/j.heares.2016.06.001>
- Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, *199*(1), 31–39. <https://doi.org/10.1016/j.heares.2004.07.010>
- Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory Neurophysiologic Responses and Discrimination Deficits in

Children with Learning Problems. *Science (American Association for the Advancement of Science)*, 273(5277), 971–973.

<https://doi.org/10.1126/science.273.5277.971>

Legris, E., Gomot, M., Charpentier, J., Aoustin, J. M., Aussedat, C., & Bakhos, D.

(2018). Assessment of auditory discrimination in hearing-impaired patients.

European Annals of Otorhinolaryngology, Head and Neck Diseases, 135(5),

335–339. <https://doi.org/10.1016/j.anorl.2018.04.004>

Lonka, E., Kujala, T., Lehtokoski, A., Johansson, R., Rimmanen, S., Alho, K., &

Näätänen, R. (2004). Mismatch Negativity Brain Response as an Index of

Speech Perception Recovery in Cochlear-Implant Recipients. *Audiology &*

Neurotology, 9(3), 160–162. <https://doi.org/10.1159/000077265>

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., & Vuust, P. (2019).

Weighting of neural prediction error by rhythmic complexity: A predictive

coding account using mismatch negativity. *The European Journal of Neurosci-*

ence, 49(12), 1597–1609. <https://doi.org/10.1111/ejn.14329>

Lupyan, G., & Clark, A. (2015). Words and the World: Predictive Coding and the

Language-Perception-Cognition Interface. *Current Directions in Psychological*

Science: A Journal of the American Psychological Society, 24(4), 279–284.

<https://doi.org/10.1177/0963721415570732>

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The Effects of Decreased Au-

dibility Produced by High-Pass Noise Masking on N1 and the Mismatch Neg-

ativity to Speech Sounds /ba/ and /da. *Journal of Speech, Language, and Hear-*

ing Research, 42(2), 271–286. <https://doi.org/10.1044/jslhr.4202.271>

- Maslin, M. R. D., Wise, K. J., & Purdy, S. C. (2023). The mismatch response in normal hearing adults: A performance comparison with stimuli relevant for objective validation of hearing aid fittings. *International Journal of Audiology, ahead-of-print*(ahead-of-print), 1–11.
<https://doi.org/10.1080/14992027.2022.2142682>
- May, P. J. C. (2021). The Adaptation Model Offers a Challenge for the Predictive Coding Account of Mismatch Negativity. *Frontiers in Human Neuroscience, 15*, 1–10. <https://doi.org/10.3389/fnhum.2021.721574>
- Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural repetition suppression: Evidence for perceptual expectation in object-selective regions. *Frontiers in Human Neuroscience, 8*(1), 1–8.
<https://doi.org/10.3389/fnhum.2014.00225>
- Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., & Scott, S. K. (2016). Distinct neural systems recruited when speech production is modulated by different masking sounds. *The Journal of the Acoustical Society of America, 140*(1), 8–19. <https://doi.org/10.1121/1.4948587>
- Muller-Gass, A., Marcoux, A., Logan, J., & Campbell, K. B. (2001). The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. *Neuroscience Letters, 299*(3), 197–200. [https://doi.org/10.1016/S0304-3940\(01\)01508-7](https://doi.org/10.1016/S0304-3940(01)01508-7)
- Murphy, E., & King, E. A. (2022). *Environmental Noise Pollution: Noise Mapping, Public Health, and Policy* (2nd ed.). Elsevier. Chapter 2

- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38(1), 1–21. <https://doi.org/10.1017/S0048577201000208>
- Näätänen, R. (2002). The perception of speech sounds by the human brain as reflected by the mismatch negativity brain response. *International Congress Series*, 1232(C), 97–105. [https://doi.org/10.1016/S0531-5131\(01\)00795-6](https://doi.org/10.1016/S0531-5131(01)00795-6)
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329. [https://doi.org/10.1016/0001-6918\(78\)90006-9](https://doi.org/10.1016/0001-6918(78)90006-9)
- Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, 19(4), 601–607. <https://doi.org/10.3758/s13423-012-0247-5>
- NITRC, N. T. & R. C. (2011). *EEProbe*. <https://www.nitrc.org/projects/eeprobe/>
- Oculaudio. (2023). *Oculaudio*. <https://oculaudio.com/product/>
- Parker, D. (2006). *ERP Processing Using EEProbe*. <https://brain.mcmaster.ca/file/EEProbe.procedure.htm>
- Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
- Pigeon, S. (2014). *Cafe Restaurant*. <https://mynoise.net/NoiseMachines/cafeRestaurantNoiseGenerator.php>
- Pigeon, S. (2020). *Babble Noise*. <https://mynoise.net/NoiseMachines/babbleNoiseGenerator.php>
- Pigeon, S. (n.d.). *A Short Biography*. <https://stephanepigeon.com/biography.php>
- Plack, C., J. (2004). Auditory perception. *Psychology Press Ltd. Figures*, 1–19.

- Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., & Woldorff, M. G. (2013). *Principles of cognitive neuroscience*. Sinauer Associates, Inc. Chapter 2 & 12
- Rescorla, M. (2021). Bayesian modeling of the mind: From norms to neurons. *Wiley Interdisciplinary Reviews. Cognitive Science*, *12*(1), 1–15.
<https://doi.org/10.1002/wcs.1540>
- Rinne, T., Antila, S., & Winkler, I. (2001). Mismatch negativity is unaffected by top-down predictive information. *Neuroreport*, *12*(10), 2209–2213.
<https://doi.org/10.1097/00001756-200107200-00033>
- Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H. G. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, *36*(6), 835–838. <https://doi.org/10.1111/1469-8986.3660835>
- Salo, S. K., Lang, A. H., & Salmivalli, A. J. (1995). Effect of Contralateral White Noise Masking on the Mismatch Negativity. *Scandinavian Audiology*, *24*(3), 165–173. <https://doi.org/10.3109/01050399509047531>
- Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. *J Cogn Neurosci*, *Summer*; *5*(3), 363–370. <https://doi.org/10.1162/jocn.1993.5.3.363>
- Sanei, Saeid., & Chambers, Jonathon. (2007). *EEG signal processing*. John Wiley & Sons. Chapter 1
- Shetake, J. A., Wolf, J. T., Cheung, R. J., Engineer, C. T., Ram, S. K., & Kilgard, M. P. (2011). Cortical activity patterns predict robust speech discrimination ability in noise. *The European Journal of Neuroscience*, *34*(11), 1823–1838.
<https://doi.org/10.1111/j.1460-9568.2011.07887.x>

- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätänen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neuroscience Letters*, *251*(2), 141–144. [https://doi.org/10.1016/S0304-3940\(98\)00529-1](https://doi.org/10.1016/S0304-3940(98)00529-1)
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. <https://doi.org/10.1111/j.1749-818X.2009.00125.x>
- Sperry, J., Wiley, T., & Chial, M. (1997). Word recognition performance in various background competitors (Featured Article). *Journal of the American Academy of Audiology*, *8*, 71–80.
- Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637. <https://doi.org/10.3758/BF03196525>
- Tervaniemi, M. (2022). Mismatch negativity–stimulation paradigms in past and in future. *Frontiers in Neuroscience*, *16*, 1–7. <https://doi.org/10.3389/fnins.2022.1025763>
- Tsogli, V., Jentschke, S., Daikoku, T., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports*, *9*(1), 1–12. <https://doi.org/10.1038/s41598-019-42066-4>
- Tsogli, V., Jentschke, S., & Koelsch, S. (2022). Unpredictability of the “when” influences prediction error processing of the “what” and “where.” *PloS One*, *17*(2), 1–19. <https://doi.org/10.1371/journal.pone.0263373>

- Weiss, B. (2021). *Voice attractiveness: Studies on sexy, likable, and charismatic speakers* (1st ed. 2021.). Springer. <https://doi.org/10.1007/978-981-15-6627-1>
- Weissgerber, T. L., Winham, S. J., Heinzen, E. P., Milin-Lazovic, J. S., Garcia-Valencia, O., Bukumiric, Z., Savic, M. D., Garovic, V. D., & Milic, N. M. (2019). Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency. *Circulation (New York, N.Y.)*, *140*(18), 1506–1518. <https://doi.org/10.1161/CIRCULATIONAHA.118.037777>
- Winkler, I. (2007). Interpreting the Mismatch Negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163. <https://doi.org/10.1027/0269-8803.21.34.147>
- Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of Stimulus Frequency and Complexity on the Mismatch Negativity and Other Components of the Cortical Auditory-Evoked Potential. *The Journal of the Acoustical Society of America*, *109*(4), 1526–1537. <https://doi.org/10.1121/1.1349184>



FORESPØRSEL OM DELTAKELSE I FORSKNINGSPROSJEKTET

BAKGRUNNSSTØY OG LYDPROSESSERING

Du forespørres om deltakelse i et forskningsprosjekt der vi undersøker hvordan bakgrunnsstøy påvirker oppfattelsen av tale-lyder. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med prosjektet er å undersøke hvordan bakgrunnsstøy påvirker oppfattelsen av tale-lyder hos personer med normal hørsel og personer med hørselsnedsettelse. I studien vil vi bruke en metode som kalles elektroencefalografi (EEG), en ikke-invasiv metode for måling av elektrisk aktivitet i hjernen. Under eksperimentet vil du høre noen lyder, samt svare på noen spørsmål. Studien gjennomføres som del av en masteroppgave, og opplysningene hentet fra undersøkelsen vil bli brukt til dette formålet. Resultatene fra oppgaven vil bli offentlig tilgjengelig i en prosjektdatabase ved Aalborg Universitet.

Hvem er ansvarlig for prosjektet og hvordan kan de kontaktes?

Prosjektet gjennomføres hos Institutt for Biologisk og Medisinsk Psykologi (IBMP) ved Psykologisk fakultet ved Universitetet i Bergen, som er behandlingsansvarlig institusjon. Selve masteroppgaven er tilknyttet Aalborg Universitet i Danmark.

Ansvarlig for gjennomføringen av prosjektet er:

Professor Stefan Kölsch

Masterstudent:

Louise Lehmann Christensen, e-mail: llch18@student.aau.dk, telefon: +45 53 38 37 87

HVORDAN FOREGÅR EKSPERIMENTET?

Din deltakelse i studien består av to faser som er beskrevet i detalj nedenfor. Vi skal også gi deg flere muntlige instruksjoner om hvert trinn underveis i forsøket, og hvis du har noen spørsmål, er vi her for å svare.

- 1) *Besvare et spørreskjema:* I forbindelse med studien registrerer vi noen få personlige opplysninger om deg, blant annet navn, etternavn og fødselsdato. Vi vil også be deg om å gi oss informasjon om din musikalske bakgrunn og om du har noen hørselshemninger eller nevrologiske sykdommer, eller om du noen gang har blitt diagnostisert med noen form for språkvansker.
- 2) *EEG-eksperiment:* EEG står for elektroencefalografi og måler hjernens elektriske aktivitet via elektroder plassert i en hette. Denne hetten er laget av et svært fleksibelt vev, som en badehette, og vil bli plassert over hodet ditt. For å sikre at den elektriske ledningsevnen mellom hodet og elektrodene er høy nok, må vi påføre en gelé rundt elektrodene. Geléen er dermatologisk godkjent og vannløselig.

Plasseringen av elektrodene kan ta ca. 50 minutter. I løpet av denne tiden kan du slappe av ved å lese en bok eller lukke øynene dine. Når vi er ferdige med plasseringen av elektrodene, starter eksperimentet. I løpet av denne tiden skal du høre noen lyder. Det er svært viktig at du prøver å redusere enhver bevegelse i ansiktet under målingen. Dette vil gjøre det mulig for oss å registrere signal med svært lite støy på grunn av muskelbevegelse. For å gjøre eksperimentet mer underholdende for deg, lar vi deg se på film. På slutten av stimulipresentasjonen blir du bedt om å svare på noen få spørsmål om stimuliene du har hørt.

Etter at forsøket er fullført, fjerner vi hetten fra håret ditt, og hjelper deg å vaske håret for å fjerne geléen. Til dette formålet tilbyr vi sjampo, børste, hårføner og håndkle, slik at du kan gjøre det, eller så kan vi hjelpe deg med det rett etter eksperimentet.

Den totale varigheten av undersøkelsen er ca. 2 timer (5 min til å svare på spørreskjemaet, 50 min til plasseringen av elektroder, 50 min til EEG-målingen, 15 min til fjernelse av elektroder og hårvask).

EEG-metoden som brukes i denne studien er for forskningsrelaterte formål og ikke for kliniske formål. I tillegg er personene som er ansvarlige for å administrere EEG-målingen, ikke kvalifisert for noen form for diagnostisering.

MULIGE FORDELER OG ULEMPER VED Å DELTA I FORSØKET

Din deltakelse har ingen åpenbare, helsemessige fordeler for deg. Din deltakelse er imidlertid viktig fordi den bidrar til forskning på hvordan bakgrunnsstøy påvirker oppfattelsen av tale-lyder hos personer med normal hørsel, og personer med hørselsnedsettelse, som kan gi bedre innblikk i måter å hjelpe personer med hørselsnedsettelse gjennom utvikling av ny teknologi.

Studien innebærer ingen alvorlige ulemper for deg som deltaker, bortsett fra det faktum at vi bruker litt av din tid og at du kanskje føler litt ubehag under påføringen av gelen.

FRIVILLIG DELTAKELSE OG MULIGHET FOR Å TREKKE SITT SAMTYKKE

Det er frivillig å delta i prosjektet. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side. Du kan når som helst, og uten å oppgi noen grunn, trekke ditt samtykke. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Dersom du trekker deg fra prosjektet, kan du kreve å få slettet innsamlede data og opplysninger om deg, med mindre opplysningene allerede er inngått i analyser eller brukt i vitenskapelige publikasjoner. Dersom du senere ønsker å trekke deg eller har spørsmål om prosjektet, kan du kontakte:

Louise Lehmann Christensen, e-post: lch18@student.aau.dk, telefon: +45 53 38 37 87

HVA SKJER MED OPPLYSNINGENE OM DEG?

Informasjonen som registreres om deg skal kun brukes slik som beskrevet i hensikten med studien. Du har rett til innsyn i hvilke opplysninger som er registrert om deg og rett til å få korrigert eventuelle feil i de opplysningene som er registrert.

Alle opplysningene vil bli behandlet uten navn og fødselsnummer eller andre direkte gjenkjenner opplysninger. En kode knytter deg til dine opplysninger gjennom en navneliste.

Prosjektleder har ansvar for den daglige driften av forskningsprosjektet og at opplysninger om deg blir behandlet på en sikker måte. Informasjon om deg vil bli anonymisert eller slettet, inkludert referansennummeret (koblingsnøkkelen) til dataene dine, senest fem år etter prosjektslutt, i 01/10/2028.

FORSIKRING

Alle deltakere i studien er dekket av pasientskadeloven.

KOMPENSASJON FOR DELTAKELSE

Du vil få en kompensasjon på 200 NOK for å delta i studien.

GODKJENNING

Prosjektet er godkjent av Regional komité for medisinsk og helsefaglig forskningsetikk, **2018/2409**.

Etter ny personopplysningslov har behandlingsansvarlig Universitetet I Bergen og prosjektleder Prof. Stefan Kölsch et selvstendig ansvar for å sikre at behandlingen av dine opplysninger har et lovlig grunnlag. Dette prosjektet har rettslig grunnlag i EUs personvernforordning artikkel 6a og artikkel 9 nr. 2 og ditt samtykke.

HVA GIR OSS RETT TIL Å BEHANDLE PERSONOPPLYSINGER OM DEG?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Institutt for Biologisk og Medisinsk psykologi ved Universitetet i Bergen har Sikt – Kunnskapssektorens tjenesteleverandør vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål knyttet til vurderingen som er gjort av personverntjenestene fra Sikt, kan du ta kontakt via:

- Epost: personverntjenester@sikt.no eller telefon: 73 98 40 40.

Med vennlig hilsen

Prosjektansvarlig
Professor Stefan Kölsch

Masterstudent
Louise Lehmann Christensen

SAMTYKKEERKLÆRING

Jeg har mottatt og forstått informasjon om prosjektet *bakgrunnsstøy og lydprosessering*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i EEG-måling
- å delta i spørreskjema
- at opplysninger om meg anonymisert kan anvendes i masteroppgaven

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

Sted og dato

Deltakers signatur

Deltakers navn med trykte bokstaver



SPØRRESKJEMA FOR DELTAKER

1. Navn: _____
2. Fødselsdato: _____
3. Kjønn: Mann Kvinne
4. Høyre- eller venstrehendt: Høyre Venstre
5. Har du hørselstap som du er klar over? Ja Nei
6. Har du noen gang hatt en nevrologisk sykdom eller psykiske lidelser? Ja Nei
7. Har du noen gang hatt psykiatriske lidelser eller sykdommer? Ja Nei
8. Bruker du beroligende midler eller sentralstimulerende medisiner? Ja Nei
9. Har du studert musikk? Ja Nei
10. Spiller du et musikkinstrument? Ja Nei
11. Hvilket instrument spiller du? _____
12. Hvor mange år har du praktisert? _____
13. Har du blitt diagnostisert med noen språkproblemer? _____
14. Hvis ja, hva slags språkproblemer? _____

Takk!

GODKJENNING

Prosjektet er godkjent av Regional komite for medisinsk og helsefaglig forskningsetikk,



QUESTIONNAIRE FOR PARTICIPANT

1. Name: _____
2. DOB: _____
3. Gender: Male Female
4. Right or left-handed: Left Right
5. Do you have any hearing loss that you are aware of? Yes No
6. Have you ever had neurological damage, disorders or disease? Yes No
7. Have you ever had psychiatric disorders or disease? Yes No
8. Do you take any medication that affects the central nervous system (tranquilizers or stimulants)? Yes No
9. Did you have any musical training beyond music lessons in school? Yes No
10. Are you playing any musical instrument? Yes No
11. Which instrument/s are you playing? _____
12. How many years have you been practicing? _____
13. Have you been diagnosed with any kind of language deficit? Yes No
14. If yes, what kind of language deficit? _____

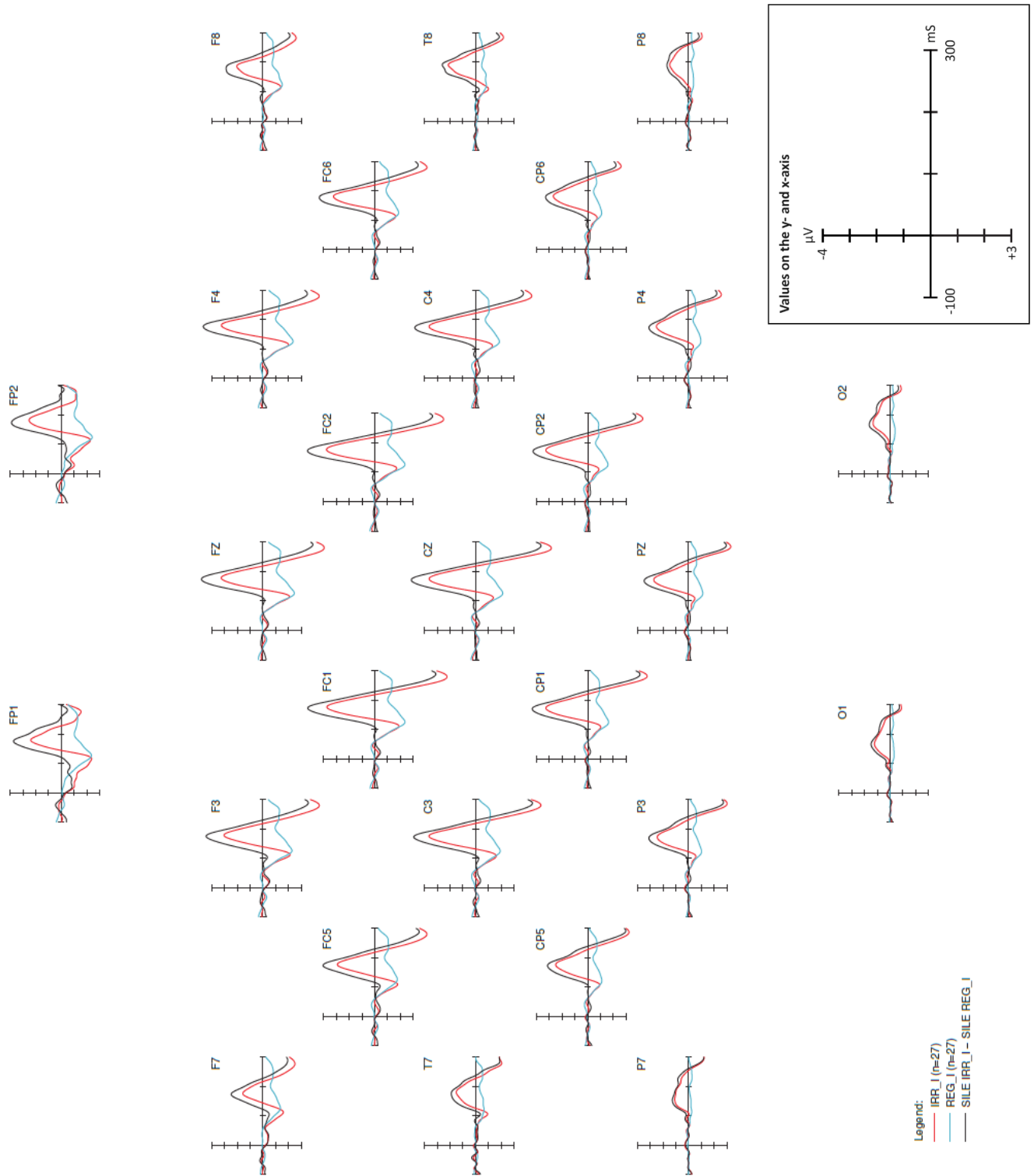
Thank you!

APPROVAL

The Project is approved by the Regional Committee for Medical and Health Research Ethics

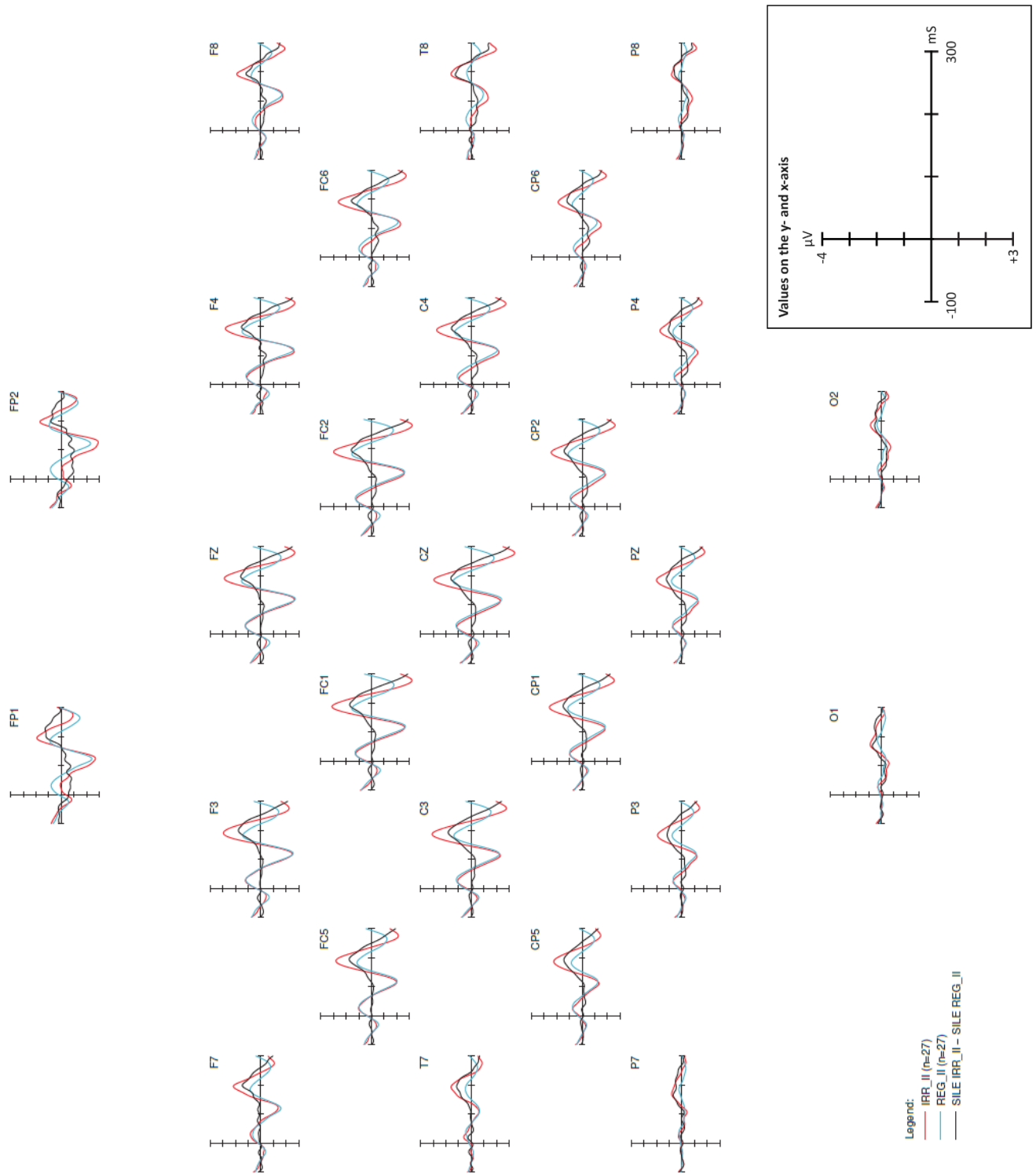
ERP waveforms for the simple condition without background noise

The time window for statistical analysis was set to 160-220 ms after stimulus onset. The red line represents the average of the ERPs elicited by irregular stimuli, the blue line represents the average of the ERPs elicited by regular stimuli, and the black line represents the MMN.



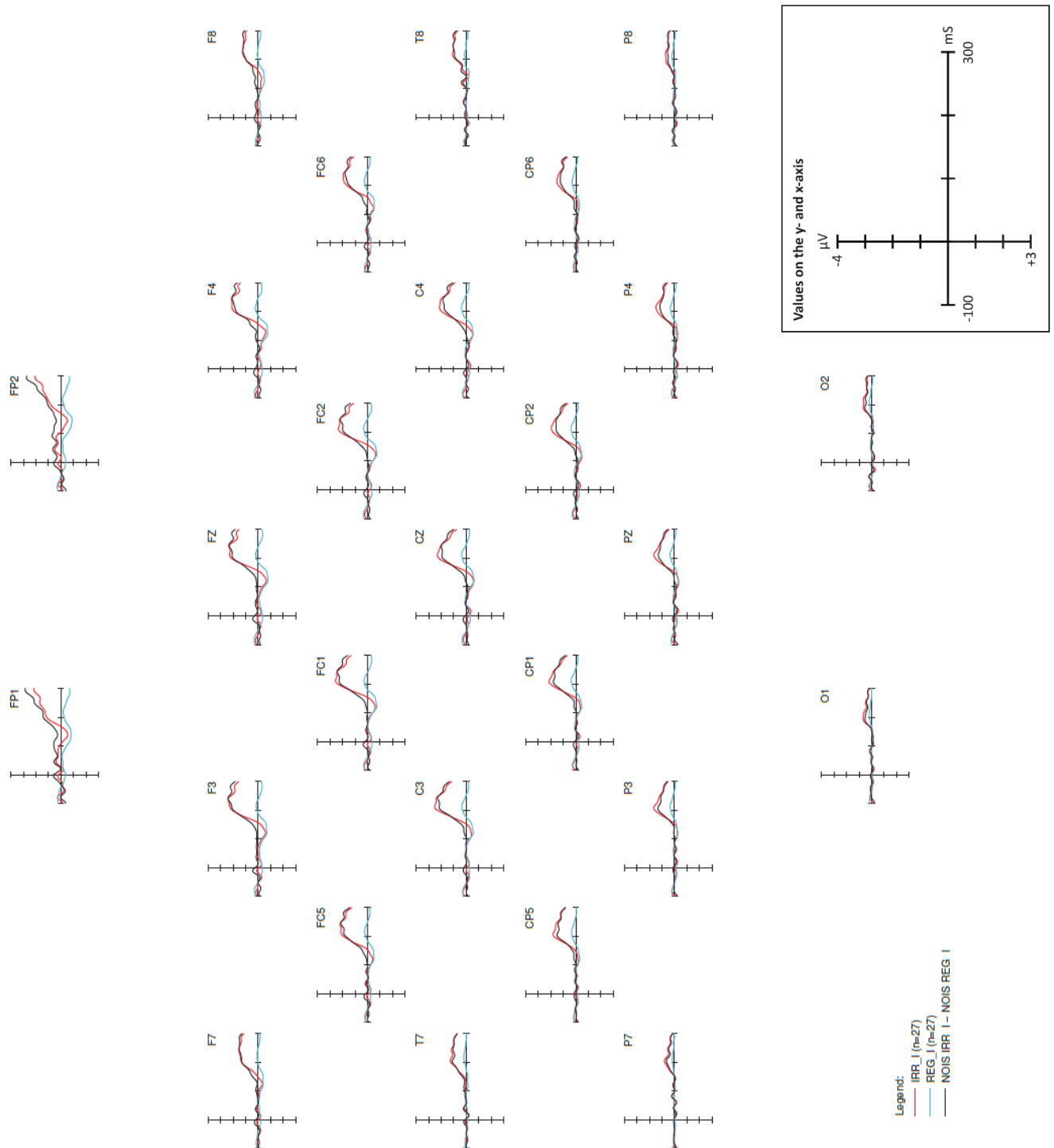
ERP waveforms for the complex condition without background noise

The time window for statistical analysis was set to 160-220 ms after stimulus onset. The red line represents the average of the ERPs elicited by irregular stimuli, the blue line represents the average of the ERPs elicited by regular stimuli, and the black line represents the MMN.



ERP waveforms for the simple condition with background noise

The time window for statistical analysis was set to 220-280 ms after stimulus onset. The red line represents the average of the ERPs elicited by irregular stimuli, the blue line represents the average of the ERPs elicited by regular stimuli, and the black line represents the MMN.



ERP waveforms for the complex condition with background noise

The time window for statistical analysis was set to 220-280 ms after stimulus onset. The red line represents the average of the ERPs elicited by irregular stimuli, the blue line represents the average of the ERPs elicited by regular stimuli, and the black line represents the MMN.

