Unknowingly knowing the unknown - exploring the MMN of statistical language learning in a natural linguistic context

Olav Tidemann Garli og Aina Cecilie Moe Klinge

Masteroppgåve

LOGO345, masterprogram i logopedi,

ved

UNIVERSITETET I BERGEN

INSTITUTT FOR BIOLOGISK OG MEDISINSK PSYKOLOGI

DET PSYKOLOGISKE FAKULTET

VÅREN 2023

Antal ord: Kappe: 17298 ord. Artikkel: 7232 ord. Totalt: 24530 ord.
Vegleiar: Arve Egil Asbjørnsen ved Institutt for biologisk og medisinsk psykologi
Bivegleiarar: Sebastian Jentschke ved Institutt for samfunnspsykologi
Sunniva Sørhus Eidsvåg ved Institutt for biologisk og medisinsk psykologi

Forord

Når me fekk tilbod om å gjennomføra ein EEG- undersøking for å kartlegga statistisk språklæring, var dette noko som appellerte til oss begge, både lingvisten og pedagogen. Me hadde aldri vore i nærleiken av en EEG-lab tidlegare, men me hadde ei formeining om at det ville innebera elektrodar og gele. Ein kan seie at det vart ei bratt læringskurve, men me er båe samde i at det har vore både lærerikt og gøy å få gjæra noko helt nytt. Særskilt gildt syns me det var å få vandra rundt på Psykologisk Fakultet ikledd labfrakk.

Me har heldigvis ikkje vore overlate til oss sjølv gjennom prosessen. Me har hatt tett og kunnig oppfylging frå svært kompetente vegleiarar, tre i talet. Ein kan kanskje tenka at det er i overkant mykje med tre vegleiarar på ei masteroppgåve, men dei har alle bidrege med stor fagleg kompetanse og verdifull innspel undervegs. Tusen takk til dykk alle tre; Arve, Sebastian og Sunniva!

Me vil au takka alle dei andre som har vore til stor hjelp og nytte gjennom prosessen: alle vener og kjende som stilte opp som deltakarar; og alle dei som opna Pandoras øskje når dei har spurd om kva vi skriv om.

Ei særskild takk til Odd Arne og Saphira som tolmodig har støtta oss gjennom prosessen og har heia oss heilt fram til mål. Ringraziamo anche Rita e Andrea per l’ufficio e la brioche al pistacchio, il supporto morale e i buoni consigli.

Bergen, 13. mai 2023

Aina og Olav

Content

Monopgraph

Article

Samandrag

Statistisk læring (SL) er ein kjernemekanisme i språklæring. SL har vorte undersøkt i både åtferdsbaserte og nevrofysiologiske eksperiment. I det fylgjande vil me presentera rammene for eit EEG-eksperiment kor tjue deltakarar (ukyndige i russisk eller andre slaviske språk) fekk høyre ein talestraum med russiske ord, kor seks moglege rotord var sett saman med to moglege endingar. Kort fortalt var målet for studien å undersøka om eksponering for naturleg språkstimuli utløyste ein nevrofysiologisk markør for regeltileigning gjennom statistisk læring som kallast statistical mismatch negativity (sMMN). Eit anna mål var å undersøka om deltakarane opplevde læring. Denne kappa inneheld grunnleggjande informasjon som dannar strukturen kring studien. Kappa er strukturert på fylgjande vis:
I fyrste kapittel presenterer me teori knytt problemstillinga, noko som leier fram til andre kapittel kor hypotesane blir presentert. Kapittel tre handlar om metoden og inneheld mellom anna vitskapsteori og metodeteori, samt informasjon om stimuli, deltakar, prosedyrar og sjølve eksperimentet. I kapittel fire vert EEG- og åtferdsdataa presentert, før ein presentasjon av metode- og analysekritikk kjem i kapittel fem. Kapittel seks og sju tek for seg drøfting av validitet og dei etiske aspekta knytt til studien.

Nøkkelord: Språklæring, Statistisk læring, EEG, ERP, Mismatch Negativity, sMMN, phMMN.

Abstract

Statistical learning (SL) is a core learning mechanism in language acquisition. It has been explored in both behavioural as well as in neurophysiological experiments. In this monograph we present the framework for an EEG experiment wherein twenty participants (unfamiliar with Russian or other Slavic languages) listened to a stream of speech of Russian words, consisting of six roots and two possible endings. In short, the goal for this study was to explore whether the exposure to natural language stimuli would elicit a neurophysiological marker for rule acquisition through statistical learning called statistical mismatch negativity (sMMN). Another goal was to determine if the participants exhibited learning. This monograph contains the fundamental information that lays the foundation for the study. The monograph is organised in several chapters. In the first chapter, we present theory connected to our hypotheses, which leads to the second chapter, where these are presented. Chapter three concerns methodology, and contains theory regarding science and method, in addition to information regarding the stimuli, participants, procedure and the experiment itself. In the fourth chapter, the results for the EEG and the behavioural tests are presented, and in chapter five a critique of method and analysis is offered. Chapters six and seven contain a discussion regarding the validity of the study and the relevant ethical aspects.

Key words: Language acquisition, Statistical learning, EEG, ERP, Mismatch Negativity, sMMN, phMMN.

1.0 Theoretical background

1.1 Language Acquisition

Infants begin to learn the words of their native language over their first year (Beech & Swingley, 2023). Normally, they take only a few months to become familiar with the sound categories. These categories include features such as consonants, vowels, and the combination of these sounds. The understanding of words is a skill that is seen between nine and fifteen months of age, and it is connected to the capacity for interpreting and understanding other persons goals and intentions (Bergelson & Swingley, 2012). When it comes to the process of language learning, one important step is word segmentation. This requires the ability to pull words from a continuous stream of speech (Beech & Swingley, 2023). Han et al. (2022) found that children (24-months of age) could learn new words from both adult directed speech (ADS) and infant directed speech (IDS). But the results showed that learning from IDS is predicted by prosodic elements, such as pitch rate, and language acquisition is therefore dependent on both individual and contextual differences in the IDS prosody (Han et al., 2022). Keren-Portnoy et al. (2019) found that infants recognized words that had been presented in isolation, but not words that had been presented as part of a sentence. But despite of this, infants do successfully break sentences into parts, and they show some knowledge of grammatical words that will never appear in isolation (Shi & Lepage, 2008). Researchers have shown that infants use the phonetic characteristics of a language to form generalisations that they apply to their further language acquisition (Mattys & Jusczyk, 2001, Erickson & Thiessen, 2015). For instance, Cardoso (2011) researched the role of coda in relation to a second language acquisition. A definition of coda is that it is made up of the consonants at the end of a syllable (Cardoso, 2011). Cardoso (2011) conducted his experiment on 51 participants, mean age 24, with Brazilian Portuguese as their first language, that were learning English. In Brazilian Portuguese the use of certain elements in the coda is illicit, as only four consonants; [l], [n], [r] and [s] can appear in coda position. In English a greater variety of codas is allowed. Cardoso (2011) hypothesized that the Brazilian Portuguese speakers would process the use of codas as a use of an illusory epenthetic vowel. The results showed that there is a certain degree of correlation between speech perception and production. This indicates that perception precedes the production of speech (Cardoso, 2011). A similar study was made by Hamada and Goya (2014) who were looking at a group of Japanese speaking students trying to learn English. They hypothesized that their participants would learn English words with an open-syllable structure without consonant clusters better than words with these qualities. An open-syllable structure means that the syllable ends with a vowel sound. Their hypothesis was disconfirmed, as they found that the recall accuracy was higher in regard to the words that contained consonant clusters and coda (Hamada & Goya, 2015). Perhaps this can be seen as an expression of the learner's anticipation towards the language he is trying to acquire, based on what it sounds like. Thiessen and Saffran (2003) found that 7-months-old attend more to statistical cues in speech, than to cues connected to stress. On this theoretical basis they argue that there might be a possibility that infants are using their statistical learning (SL) abilities to locate words in a stream of speech. Furthermore, they use these words to discover regularities in stress patterns (Thiessen & Saffran, 2003). This may indicate a connection between SL and the prosodic features within the spoken language. Below we will present SL in relation to language acquisition, thereafter we will present prosody and research on the prosodic elements of language acquisition.

1.2 Statistical Learning

SL in the context of language learning is to use pattern detection and computational skills (Kuhl, 2004). This can be seen as an extraction from linguistic input and an integration into an expanding mental network (Perruchet & Vinter, 1998; Thiessen et al., 2013). SL must be distinguished from habituation, as it refers the processes involved in extraction of information from a stimulus. Habituation is a phenomenon in relation to repeated stimulus, where there is a waning in the elicited response (Valsecchi & Turatto, 2023). Learning regarding the understanding of words and languages is based on the statistical structure of some elements within them, such as frequency, variability, distribution and the probability of a co-occurrence. The statistical structures in all of these elements are connected to SL (Erickson & Thiessen, 2015). The mechanism involved in picking up on these aspects can be explained by our sensitivity to irregularities, and our ability to pick up statistical structures in the world around us (Koelsch et al., 2016). In fact, Saffran et al. (1996) suggests that infants, at only 8 months of age, can accomplice word segmentation due to SL, after being exposed to just two minutes of speech. This may indicate that the mechanisms for the computation of the statistical properties are powerful in this age-group in relation to identifying structures in languages (Saffran et al., 1996). Teinonen et al. (2009) performed a study to explore skills in SL at birth by recording event related potentials (ERP) responses in new-born babies. While they were sleeping, they were exposed to a stream of syllables that contained statistical cues to word boundaries. This means that new-born babies were able to extract the statistical properties from the stream of syllables they were exposed to, and they are able to detect word boundaries. The babies' brain was found to treat syllables differently based on their position within the pseudowords in the stimuli (Teinonen et al., 2009). This shows that the ability to use SL in language acquisition is a capacity we are born with. SL is a mechanism that enables us to learn from our surroundings, and it takes place in every situation every day, from day one. According to Thiessen et al. (2013) learners are sensitive to two aspects of the statistical structure when it comes to language acquisition. These are conditional statistical information and distributional statistical information. These two again include two complementary processes; extraction and integration. The term extraction refers to the process in the working memory, which enables us to hold two elements in our consciousness and at the same time combine them. Integration, on the other hand, refers to the process of combining information, and this will help us identify central tendencies within the information. It is not quite clear how closely these two processes are connected, but the importance of their mutual relationship is undoubtedly significant (Perruchet & Vinter, 1998; Thiessen et al., 2013). Using a behavioral paradigm, Eidsvåg et al. (2015) investigated whether variability in the linguistic input would influence learning. Forty adults, divided into two groups, were familiarized with noun gender subcategories in Russian. One group were presented with a high-variability condition in which they were familiarized with 32 different root-words. The other group were presented with a high-repetition condition, they were familiarized with 16 root-words presented twice to provide the same amount of exposure as in the first group. The results showed that only participants in the high-variability group experienced learning after the initial familiarisation, while participants in the high-repetition condition needed additional exposure in order to learn the endings. This is a demonstration that learner's ability to generalize language input can be influenced by the degree of input variability (Eidsvåg et al., 2015).

SL has been found to play a significant role in several areas that involve learning, not just in relation to language acquisition (Plante & Gómez, 2018). For instance, studies have been conducted to investigate the relationship between SL and reading. One tutorial, based on a review of a range of studies, investigated how an implicit method based on the principles of SL could support children in the process of learning how to read (Arciuli, 2018). Other studies have looked at the relationship between SL and music, and further discussed the coupling between music and language. Jentschke et al. (2008) conducted an ERP study that found that children with specific language impairment also tend to show impairment of music-syntactic processing. Research has shown that there is a connection between developmental language impairment (DLD) and impaired statistical language learning (Arciuli & Conway, 2018; Hsu et al., 2014; Plante et al., 2017), and that there is an overlap in neural resources between music processing and syntax (Koelsch et al, 2005). The association is seen in the light of cognitive mechanisms that are comparable to each other. It also includes shared neural resources that are essential for the processing of musical and linguistic information (Jentschke et al., 2008). The finding is supported by Steinbeis and Koelsch (2008) found that musical meaning is represented in much the same fashion as meaning in language. Schön et al. (2008) found that a consistent mapping of linguistic and musical information might enhance learning compared to speech sequences. This might prove beneficial when in the process of acquiring a new/second language (Schön et al., 2008). However, language and music are not only made up of structures alone. When it comes to speaking and singing, prosody plays an important role. Below we will look at prosody in relation to language acquisition.

1.3 Prosody

Language acquisition entails learning prosodic information (Shukla et al., 2007). Such information might be delivered in stress, pauses, and intonation. According to Kuhl (2004) languages are mostly dominated by trochaic words, where the first syllable is stressed, or iambic words, where the second syllable are stressed. The distinction between trochaic words and iambic words can be used to identify word boundaries. Babies use both statistical and prosodic cues to segment words, and both kinds of cues are related to linguistic stress (Kuhl, 2004). This was supported by Johnson and Seidl (2009) that found evidence that 11 month-old infants tend to weight stress cues in relation to word boundaries more heavily than statistical cues. Prosody might act as a filter that suppresses possible word-like sequences that outspan prosodic constituents in a stream of speech (Shukla et al., 2007). For example, songs have been found to enhance language acquisition in several ways. Songs have an emotional aspect, that might increase the attention we pay them. The use of pitch contours might enhance phonological discrimination. Learning mechanisms might be optimized by the consistent mapping of musical and linguistic structures in songs (Schön et al., 2008).

Most typically, speech is organized into units that are prosodically cohesive, and these are found to range from a single syllable to an entire utterance. Such boundaries are associated with acoustic cues, for example a decline in pitch and final lengthening (Shukla et al., 2007). Kuronen and Tergujeff (2020) studied the development of prosody in a second language. The participants consisted of twenty-five native Finnish-speakers who were learning Swedish. They found that the participants acquired not only phonological, but also phonetic aspects of the second language, and that the process of learning a phonological aspect of a language seems to be supporting the learning of the phonetic-prosodic features in the second language. Based on this, they draw the conclusion that there is a connection between the development of different features in language acquisition (Kuronen & Tergujeff, 2020). Saksida et al. (2021) who conducted a study on Spanish and Italian speaking adults, by checking their ability to learn an artificial language with either non-native or native word order. They found that when the artificial language had a similar word order to their native language the participants were using prosodic information to identify the most common words. When the word order was different from their native language it took three days of exposure of the rhythmic structure for them to learn the most common words. Based on their findings they conclude that familiarity with prosodic cues can facilitate leaning in a second language (Saksida et al., 2021). This means that if the second language resembles the first languages prosodic features, such as rhythmic structure, this can lead to implicit learning in regard to word order and syntactic structure. It has been established that the neural correlates of SL and language perception have been explored using EEG and ERPs. In the following two paragraphs EEG end ERPs will be described in further detail.

1.4 EEG and ERP

EEG is a technique to measure electrical activities in the nerve cells of the brain, mainly in the cortical areas, using electrodes mounted on the scalp (Hugdahl, 1995, pp. 234-235). The electrodes pick up cyclical changes in the membrane potentials of underlying nerve cells, which generates a rhythmic pattern called an EEG wave. The EEG patterns depict the fluctuations in voltage in the neurological structures that they measure. Based on the amplitude and frequency of these fluctuations, one may analyze their behavior (Hugdahl, 1995, pp. 234-235). There are about 86 billion neurons in the average human brain, and the communication between these are what is measured by EEG. The neurons have intrinsic electrical properties, and these properties, from a population of neurons called the pyramid cells, are picked up by the EEG electrodes. The pyramidal cells have polarity like batteries, and are positive or negative depending on two factors. The first is “whether an inhibitory or excitatory stimulus has come to the synaptic junction from the axon of another cell”, and the second is “whether that synapse is proximal or distal to the cell body” (Beres, 2017, p. 248). The electrodes attached to the scalp will then record a negative extracellular potential if this is happening to a large number of pyramidal cells in the same area at the same time (Beres, 2017). Traditionally speech production studies have been viewed as challenging to perform by EEG. Because talking involves many muscles, hence recorded as artifacts, that may lead to results that is hard to interpret. But according to the rise in published language studies in the resent years, this seems to be changing (Ganushchak et al., 2011). This shows some of the breadth of language research that can be done using EEG technology. Ganushchak et al. (2011) highlight that in EEG studies using button-pressing as a response to stimuli, one has to take into consideration the fact that it is unlikely that only the language processes are the contributing factor to the response. There could be error-processing or action slips, where the participant fumbles with the equipment or simply pressing the wrong button (Ganushchak et al., 2011). Given that the EEG experiment in the current study was conducted in a room with dim lighting, these are factors that we need to consider.

From the recording of an EEG, it is possible to detect an ERP. This is a signal elicited by an external or internal stimulus event, and is an "answer" from the brain that takes place within a certain timeframe following the stimulus. It is important to be aware that such signals might also occur in the absence of a stimulus, or they may be caused by motor responses. An example of such motor responses is the blinking of the eye (Hugdahl, 1995, pp. 266-270). The ERPs normally have a small amplitude, and these event related changes are typically obscured by the irregular and arbitrary EEG signals. They appear when we a create an average over repeated recordings (Beres, 2017, pp. 248). ERP-components are categorized with polarity and latency, e.g. N100 describes a negative change in polarity, occurring approximately 100 ms after a releasing stimuli event. The N100 is a relatively large waveform, and it is always elicited when the auditory stimulus is repeated several times (Hugdahl, 1995, pp. 281-282). The N400 is elicited at approximately 400 ms after stimuli. It is often referenced in language studies, but this somewhat slow component is interpreted as a semantics component and show later processing of the stimulus, not detection of a changes in stimulus qualities (Hugdahl, 1995, pp. 303-304). Sanders et al. (2002) studied ERPs in adults when being exposed to three-syllable pseudowords that were hidden within a steam of syllables. They found that the N400 amplitude was enhanced, and for a group of high learners the N100 amplitude was also enhanced for the pseudoword onsets after training. Based on this experiment they suggested implicit learning of transition probabilities (Sanders et al., 2002). Language research using ERPs has proven useful, due to their good temporal resolution. Linguistic variations are often small and fast, and therefore the ERPs are useful in picking up on these small variations, and the neurological changes that occur (Beres, 2017). One drawback when it comes to using EEG in research is the number of trials needed to create valid results. A large number of stimuli have to be presented to a relatively large group of participants to provide useful results. This is because the ERPs, as mentioned above, are a relatively small part of the entire EEG recording per participant (Beres, 2017).

In this current study, we focus mainly on a negative component, the MMN.

1.5 Mismatch Negativity

As the name suggests, Mismatch Negativity or MNN represents a negative waveform. According to Näätänen (1995) the MMN is caused a deviation in the memory trace caused by a sensory input. It has a relatively large amplitude above temporal areas, and it has been suggested that it is taking place in the additory primary and association areas. The MMN is a reflection of specific auditory discrimination of the stimulus (Näätänen, 1995). Situations where MMN responses are normally recorded are when the deviant stimulus is given as an auditory input with a change in either intensity or frequency (Nyman et al., 1990). Näätänen (1995) lists several reasons why the MMN response is a tool that can be usen in both clinical practice as well as for auditory research. First, it provides an objective measure of the individual's ability to discriminate different sound features. Second, attention is not needed to elicit an MMN response. For the MMN to be elicited there is no need for the individual to participate in any specific activities. It has been observed when a person is reading a book, listening to music, watching a movie or sleeping (Beres, 2017). Thirdly, it involves auditory sensory memory. And finally, as auditory short-term memory is crucial for correct speech processing, the MMN provides a means for studying this phenomenon (Näätänen, 1995). The MMN acts at a pre-attentional level, as it has been found in coma patients (Näätänen, 2003). Koelsch et al. (2016) found that violations of statistical regularities that were established on-line, on a moment-to-moment basis, elicited an MMN. The MMN is an abstract feature, that is represented mainly in sensory memory (Koelsch et al., 2016). This corresponds with the suggestion that the MMN reflects the outcome of a neural mismatch process and takes place between the deviant stimulus and a memory trace of the standard stimuli (Winkler et al., 1996).

An EEG-study by Koelsch et al. (2016) involved a novel variant of the SL paradigm. They presented timbres of triplets in isochronous sequences to a group of 18 adult participants. The two first sounds of the triplets had the same amount of probability, but the third sound occurred with a probability of either low (10%), intermediate (30%) or a high probability (60%). They found that endings with low and intermediate probability elicited an early anterior negativity (100 ms to maximum 180 ms), as opposed to the high probability endings. The effect was also higher for the low probability items than the intermediate. They report a statistical MMN (sMMN) that reflects SL of transition probability distribution. These results exceed capabilities of auditory sensory memory (Koelsch et al., 2016).

Tsogli et al. (2019) aimed to compare the physical MMN (phMMN) and the sMMN and to what degree the underlying cognitive processes influence and interaction with each other. Like Koelsch et al. (2016) they were using EEG to record brain responses to deviations in an auditory stream of sound triplets. 21 participants were exposed to a continuous stream of auditory sound triplets where deviations were either statistical or physical, or a combination of the two. Tsogli et al. (2019) found that the statistical deviation was in terms of transition probability, and the physical were due to a change in the sound location. They saw that the changes elicited a difference in the MMN, that the sMMN were smaller than when it co-occurred with a physical change. This means that the processing of prediction errors related to SL is affected by prediction errors related to physical deviance (Tsogli et al., 2019). Interestingly, the lateralisation of the sMMN was maximal over the frontal-midline region, while the lateralisation of the phMMN was maximal over the frontal and central midline regions when exposed to a tonal stimulus (Tsogli et al., 2019).

Other studies (Steinbeis & Koelsch, 2008; Schön et al. 2008) investigated the relationship between SL and music, and further discussed the coupling between music and language. The finding is supported by another study (Steinbeis & Koelsch, 2008), that by conducting an EEG experiment, found that musical meaning is represented in much the same fashion as meaning in language. Schön et al. (2008) found that consistent mapping of linguistic and musical information might enhance learning compared to speech sequences. This might prove beneficial when in the process of acquiring a new/second language.

Tervaniemi et al. (1999) investigated the functional specialization of the human auditory cortex and its role in processing phonetic versus musical sounds. The experiment was recorded using magnetoencephalography (MEG) to detect a MMNm response, which is the MMN counterpart in MEG. MEG involves using a helmet-shaped 122-channel whole head magnetometer. The participants were watching a silent movie and were at the same time presented with sequences of phonemes or chords. These stimuli were either frequent or infrequent. The phonemes were respectively /e/ or /o/, and the chords were A major and A minor. They found that, within the right hemisphere, a stronger MMNm response was elicited by an infrequent cord change, than the MMNm response that was elicited by a phoneme change. As for the left hemisphere, there were no significant differences in the strength of the MMNm response in association with either cord or phoneme change. In addition to these results the researchers found that the MMNm that was elicited in relation to a phoneme change was superior to a chord change. This was true for both hemispheres (Tervaniemi et al., 1999). This showcases the brain's ability to pick up and process linguistic changes, that is an important part when it comes to learning languages. Given that the MEG were used to detect an MMNm response in the participants, it is reasonable to mention the results in the same context as results produced by the use of EEG. It is worth mentioning that Partanen et al. (2011) used naturally produced speech stimuli to elicit an MMN-response. Their stimuli consisted of a pseudoword, for example [tatata], and the changes in the given stimuli were of an acoustic character. They found that all changes to the stimuli elicited an MMN response, but by changing the duration of the vowel this elicited a different response to the other. The changes to the vowel duration also influenced the MMN lateralisation. This showed that it is possible to assess speech sound discrimination in only 30 minutes, using a multi-feature paradigm (Partanen et al., 2011). As mentions the ERPs are a very small part of the EEG recording. To find the ERPs in the overall recording, an individual component analysis was performed.

1.6 EEG data analysis

1.6.1 *Individual component analysis*

Individual component analysis (ICA) is, in its most basic definition, a technique used to separate a mixed signal output across several sensors into the singular emitted signals from the source material (Hyvärinen & Oja, 2000). This is commonly illustrated by the so-called "cocktail party problem", wherein the noise produced in crowded room full of people is recorded with several microphones placed in different locations. ICA will then allow us, based on the recording between the different microphones, to break the collected signal into the various contributing sources. Hyvärinen and Oja (2000) illustrate this, somewhat simplified mathematically, given two sources and two microphones:

$$X_1(t)=a_{11}S_1+a_{12}S_2$$

$$X2(t)=a_{21}S_1+a_{22}S_2$$

Where (t) is time, X_1 and X_2 are the signals received by the microphones, S_1 and S_2 are the signal output from the individual speakers and the a_z are parameters based on the distance from speaker to microphone.

To continue with ICA, one must assume that S_1 and S_2 are statistically independent from each other, i.e., a variation in S_1 will not cause a change in S_2 and vice versa.

Hyvärinen and Oja (2000) explain that ICA generally can be expressed with the formula:

$$\mathbf{x}=\mathbf{A}\mathbf{s}$$

Wherein **x** denotes the vector for the collected output across all sources, **A** is the matrix by which **s**, the sum of the inputs, is mixed. ICA gives us the possibility to estimate **A** and **s** based solely on **x**.

There are a few more assumptions about the components, the S_x, that need to be given in ICA processing (Hyvärinen & Oja, 2000). One is that the components are independent, so that information from S_1 will not tell us anything about S_2. It will tell us something about X, but with the content of the matrix unknown, the quantity and direction of the change is unknown. This is also connected to any potential correlation between variables, and why they must be assumed to be independent, rather than just showing zero correlation: whereas truly independent variables will never have a correlation with each other, two variables showing no correlation for a specific data set can be casual, and not exclude true independence. The main problem in ICA, as stated previously, is to determine the matrix **A,** based on **x.**

One of the foundations in the process of calculating **A**, is the assumption that the independent components must be non-Gaussian in distribution. Hyvärinen and Oja (2000) demonstrate that a Gaussian distribution is symmetrical across outputs, and thus gives no indication of the direction of the columns in the matrix, making both a matrix **A** and $\mathbf{A}^{-1}$ possible.

A step then, in ICA, is finding the maximum non-gaussianity of any given source. Hyvärinen and Oja (2000) describe ICA as taking two steps in the process of estimating the non-gaussianity of the variables. The first step is calculating the maximum non-gaussianity by combining the **s** with a vector that maximizes the non-gaussianity of the data distribution. The second step is to maximize the entropy, both within the singular variable and the joint entropy, as to minimise the joint information.

Hyvärinen and Oja (2000) point out that according to the central limit theorem, any mixture of two independent variables will give a more gaussian distribution than any of its singular components. They demonstrate this in the case of ICA by presenting the following:

$\mathbf{z}=\mathbf{A}^T\mathbf{w}$. $y=\mathbf{w}^T\mathbf{x}=\mathbf{w}^T\mathbf{A}\mathbf{s}=\mathbf{z}^T\mathbf{s}$.

Where **w** is an unknown vector that gives the linear combination of **x,** and **z** is the transpose of the unmixing matrix. *y* is the linear combination of the various elements that make **x**, that is x_i. Hyvärinen and Oja (2000) state that **z** as a vector only has one non-zero element, in the case that **w** is a vector that maximizes the non-gaussianity of $\mathbf{w}^T\mathbf{x}$. This then leads to $\mathbf{w}^T\mathbf{x}$ being equal to one of the independent components, and this process can be repeated for an unlimited number of sources, as long as they are treated as independent, that is, the value of **w** is different for each single source.

The next step in the process is maximizing entropy. Hyvärinen and Oja (2000) explain that through this, the mutual information between variables tends towards zero, and that two variables that share no mutual information are statistically independent.

It is thus then, through these processes that ICA is able to give an approximation of the source signal. Through these three steps, 1) assumption that they are statistically independent from each other, 2) maximizing the non-gaussianity of the signal through manipulation of the central limit theorem and maximization of entropy, and thus 3) constructing a series of unmixing matrices, that each single contributing source is estimated. As will have emerged from this explanation, the mixing matrices are individual to each single input source, and consequently, each participant will have to go through a machine learning process where the ICA algorithm analyses the data for the participant to find the individual signals.

1.6.2 ICA in EEGLAB

Until now, ICA as a general method of source separation has been discussed. In this study, the specific software used for data analysis was EEGLAB (Delorme & Makeig, 2004), a plug-in toolbox for MATLAB (The MathWorks, 2022). EEGLAB uses a specific type of ICA algorithm, called Infomax *runica* (EEGLAB, n.d).

2.0 Main Goals

We now know that the brain reacts in a certain way when exposed to musical stimuli. Our main goal is to discover if the brain reacts in the same way to language. We also want to explore whether language is acquired in relation to this reaction.

2.1 Research Question and Hypotheses

We expect that linguistic stimuli will elicit similar brain / MMN responses as the (non-linguistic) sound/timbre stimuli in the experiments of Koelsch et al. (2016) and Tsogli et al. (2019). More specifically, we expect that physical deviance would elicit a phMMN and that the rule-acquisition process of the transition probabilities between the word roots and endings would elicit a sMMN as in those earlier experiments.

We also expect that both neurophysiological measures of rule acquisition as well as a familiarity test (based upon implicit memory) will be more sensitive to pick up the outcome of the statistical language-based acquisition process than the explicit memory-based measures used by Eidsvåg et al. (2015).

Lastly, we expect that participants will have a better performance in the familiarity test at the end of each block as compared to Tsogli et al. (2019): The experimental paradigm and the stimuli used by Tsogli et al. (2019) required to segment the sound stream while acquiring the underlying rules of arrangement. For the current experiment, the character of the stimuli – words that were separated from one another by a pause and that each had a prosodic structure that indicated begin and end of the word – made the segmentation process unnecessary and therefore the rule-acquisition easier.

3.0 Method

3.1 Theory of science

Before the method employed in this study is described, it is best to give a theoretical foundation that will permit the reader to understand what choices were made and why. Firstly, the theory of science as a concept is explored, before methodology is examined. Then follows a description of the method of this study, and finally a critique of the method.

Methodology is, in its most basic understood form, the methods used in the collection and evaluation of scientific data (Toomela, 2023). In other words, methodology is a “toolbox” of techniques and procedures used to gain scientific knowledge. This leads us to ask the question: if methodology is the tool used “to do” science, what is science? A post-positivistic view of science has been chosen for this study, for a further discussion on various philosophies of science, see Thomassen (2018).

Toomela (2023, emphasis in original) expresses it thus: “Science is *knowledge*.” While this definition is a good start, it is important to further narrow it down, as it is clear that not all knowledge is scientific knowledge. Scientific knowledge is characterised by several factors. The first one is that scientific knowledge is knowledge about causes (Toomela, 2023). This implies that scientific knowledge does not contain knowledge about what is inherently logical, and that can be derived from axioms. Rather, it is knowledge about the connection between phenomena or things, and how they must impact and affect each other.

This leads to a second characteristic of scientific knowledge: that of predictability. Toomela (2023) points out that there are substantially two relevant types of predictability: the one that is due to knowledge of causes, and one that is not. The example brought forth by Toomela is that of birds, which will return to a bird feeder at a specific time once a feeding routine has been established, but that will not do anything to affect the feeding time should the routine be broken. A parallel could be made to this study: we know as a matter of fact that children learn language through SL, and that adults continue to use SL after childhood. What we do not know, and what we hope to find out, are the *causes* of this learning.

As Toomela (2023) points out, there have been several ways to define causality, but the most common understanding is that a relationship exists between cause and effect where one event will cause another event. Thus, causes can exist on their own, but not effects. Toomela (2023, p. 120, emphasis in original) defines science as “A special kind of knowledge, *knowledge about structural-systemic causes*”. Whilst this is more common in the natural sciences, within philosophies of science like positivism and post-positivism (Thomassen, 2018), Toomela (2023) exemplifies this in psychology through the Vygotsky-Luria’s theory of brain-psyche relationships (Luria, 1969, 1973, 2002; Vygotsky, 1960, 1982, quoted in Toomela, 2023), which states that higher psychological functions is the result of activity between physiologically defined centra in the brain, whilst at the same time being dependant on the external, social world which forms these interconnections in the individual.

Now that it is established that science is a specific kind of knowledge, another question emerges; how do we arrive at this very specific kind of knowledge in the first place? Toomela (2023) states two basic principles: firstly, scientific knowledge is constructed. Secondly, scientific knowledge is based on method, and therefore requires methodology.

3.2 Theory of method

The term methodology can generally be understood to be the study of the repertoire of techniques used in research and seems to have overlapping meaning to the word “method” in this context. More specifically, methodology is the overview of the ways in which scientific data is collected and evaluated (Toomela, 2023).

What does it mean that scientific knowledge is constructed? The construct state of scientific knowledge signifies that the knowledge about a specific object or phenomenon is not immediate from how we sense it in the world, which is limited by our physiological predispositions as humans, but rather built, or constructed, from the wider understanding of the contributing phenomena (Toomela, 2023). This also applies to thinking, in that sensory-based thinking only will render the approach to scientific, constructed knowledge impossible; rather, it is with the advent of advanced mental processes, like meta-cognition and language, that allows the qualitatively appropriate method of thinking that will let us construct knowledge about the world beyond the senses (Toomela, 2023).

This then leads to the second point made above, that science is constructed through certain methods. This is well enough, but it poses an interesting problem; how is it possible to know that the knowledge we have is justified? Is the knowledge achieved through method not dependent on the method used? This is especially interesting in regard to the toolbox-metaphor used earlier.

Toomela (2023, p. 125) proposes that science is based on “methodologically justified methods”. By this he means that when “doing” science, that is, when studying an object or phenomenon, the way this is done, and how the results of the study are interpreted, must be justifiable. In this, he lays out two levels in method; the particular way an object is studied, and the way the results are viewed and interpreted, the cognition deployed in the science process.

The particular way an object or phenomenon is studied impacts the way we understand the object or phenomenon itself, because only when “it is understood how the studied thing is going to come into relationships with study conditions, conditions which are created according to the method” (Toomela, 2023, p. 126), can the studied objector phenomenon be understood. That is, a phenomenon can be understood in different ways based on how it is studied. This is an interesting aspect for this study as well, considering how certain paradigms have been chosen as models for the data to be understood. This study operates under certain assumptions about statistical language learning, and how this impacts the brain on a sub-conscious, physiological level. What the study must take into consideration then, is how the various variables employed might fit better in different models, and may be understood differently, given a different theoretical framework.

The other aspect, the cognition of the fact, is of equally, if not greater, importance regarding the approach to methodology and methodological critique. As Toomela (2023) points out, this aspect of methodology takes into consideration the limitation of the scientist himself in the act of interpretating the data from the thing that is studied. It is here that the least obvious, but perhaps most impactful lacunae in the understanding of both the pure studying of the thing and the cognition of the fact can be found; simply because we cannot know what we do not know. That does obviously not mean that we cannot know *that* we do not know, but the individual scientist’s ability nevertheless impacts the interpretation process

(Toomela, 2023). This later aspect of methodology is common for all sciences regardless of quantifiability or other aspects of its data material, whilst the first aspect must adapt to the discipline, since some sciences would be meaningless using the "tools" of others, the general "thinking about" the use of these tools is common for all of them.

This leads us to a synthesis between the cognition of study of the thing and the studying; that of the act of observation. Observation is the way we gain experience of the world, and it is always aimed and directed (Toomela, 2023). Here again, it is crucial to note how the act of observation, and the following analysis, is heavily dependent on the individual who performs these acts (Myers & Hansen, 2002, quoted in Toomela, 2023).

Now that we have seen that the observation is dependent of the scientist, it will be useful to examine what factors influence the scientist. There are mainly two aspects that will be discussed here, in relation to the preceding aspects of "studying the thing" and the "thinking about the studying of the thing", i.e., the specific methods employed in this study, and the models and theories that led to the choice of these "tools", and how this impacts the results and interpretation of these.

It will be useful to split our methods into two subcategories for further discussion: EEG measurements, with ERPs generally, and MMN especially, considering this study's focus on that particular aspect of EEG and psychophysical measurements. In addition to this it will be useful to consider the design of the experiment itself a "tool" from the toolbox, and the way the experiment has impacted what kind of data that has been gathered, and what assumptions are made when this data is analysed.

This study uses EEG in a modified 10-10 system (see fig. 1 for reference) as discussed in detail above. The use of similar electrode placement methods can be found in literature (Koelsch et al, 2016; Tsogli et al, 2019). Interestingly, these studies explore paradigms and reactions are similar to the ones explored in this study, which will be discussed later. It should be stated that while EEG has a range of clinical applications (Niedermeyer & Da Silva, 2004), it would be useful to discuss EEG in a passive-active dichotomy. In the case of epilepsy diagnosis or sleep for example, the EEG readings are passive, in the sense that the brain activity occurs spontaneously, and the EEG recording is a result of the brain function without external sources affecting it.

Whilst this kind of EEG measurements are useful to understand the brain at work on its own, a different kind of research which deals with ERPs seeks to understand how the brain physiologically reacts to the external world, as this affects the brain, through input, auditory, visual or otherwise (Da Silva, 2004). There are several kinds of ERPs, that can be produced by a variety of different stimuli. Examples are visual, called VEP (Celesia & Peachey, 2004) and a range of auditory, like morpho-syntactic error detection (Friederici, 2002) and, like in this study, pattern interruption (Kolsch et al, 2016; Tsogli et al, 2019).

As stated, the way ERPs work is that once a stimulus is administered to the subject, the brain will elicit a series of electrical signals (Da Silva, 2004). Da Silva (2004) present two different theories about the nature of ERPs. One states that once a stimulus is perceived, neural populations become active due to the stimulus, and this activation is time-locked to the stimulus. Thus, by averaging the electric potentials elicited across trials, it is possible to find back to the potentials of activation, and the location where the neural populations are located. The other theory dictated that it is not neural populations that go from being inactive to active, but rather that ERPs are a relocation and rearranging of the already occurring activity, and that the elicited potentials thus are a consequence of focus, rather than activity. This study uses time-averaging, which is the most common way of analysing ERPs (Da Silva, 2004).

This study focuses on a particular kind of ERP, the MMN. The MMN has been seen in relation to studies with linguistic stimuli, and that it plays a part in the discrimination process of speech sounds (Partanen et al., 2011). Whist the study by Partanen et al. (2011) explores auditory discrimination of sounds, a similar paradigm has been used in this study, given how this study also explores discrimination in linguistic stimuli. MMNs however, as stated are not limited to linguistic stimuli, but have been proven to be elicited by several kinds of stimuli, as longs as the stimuli are given as pattern/interruption. The literature on MMNs is extensive, including literature exploring brain reactions to low-probability items occurring instead of high-probability ones in a SL paradigm like the one in this study (for example, see Koelsch et al., 2016; Tsogli et al., 2019; Furl et al., 2011; Paraskevopoulos et al. 2012). These studies, however, elicit their MMNs through tones/timbres rather than linguistic stimuli, and the use of linguistic stimuli in a similar paradigm has yet to be explored.

Our experiment contains two core behavioural measurements used in psychophysical experiments (Wichmann & Jäckel, 2018, p. 267): Response time (RT) and Confidence. It does this through the deployment of two different kinds of familiarity tests, which are described in detail in the “procedure” and “experiment “sections. (see also Appendix C, post-test section).

The first kind of familiarity test consisted, as described previously, in choosing between a root with either a *–telya* or *-teljem* suffix, where the correct answer was the stimulus that had appeared most frequently during the preceding stimulus administration. The accuracy of the responses, together with the response time and the confidence ratings the participants gave their answers were all recorded. Response time, together with accuracy, has historically been employed to measure a plethora of different psychological paradigms, like executive function, categorisation, and, perhaps most interesting for this study, memory (Donkin & Brown, 2018). A common way RT and accuracy have been used in research is through a speeded choice paradigm, where participants are asked to make simple decisions where both the time deployed in answering and the accuracy of the responses are important factors that should be considered jointly (Donkin & Brown, 2018). This is the case in this study, but Donkin and Brown (2018) do state that the participants are often asked to focus on accuracy and response time during the experiment. This was not the case in our study; rather, the participants were not informed that RT would be recorded. This was a step taken to ensure the naturalness in the response time, and avoid the so called “observer’s paradox”, which dictates that one cannot observe a natural phenomenon when the participants are aware that they are being observed (Labov, 1966, cited in Cukor-Avila, 2000). It can also be argued that the evolution in RT within the individual participants, and the participant bodies as a whole, can be seen in relation with the learning of the stimulus material; the better they learned the stimuli, the faster and more accurately they answered. Ratcliff and McKoon (2008) state that empirical RT distributions for simple two-choice decisions are generally positively skewed, and that an increase in difficulty in the choosing process generally leads to longer RTs and lower accuracy. It seems reasonable then to interpret the opposite skewing, as experienced during the experiment, where the RTs generally became shorter and more accurate, to mean that the decision making was easier. This increased facility we believe is due to the participants familiarisation with the stimulus material and the learning of its contents. It is useful to view the experiment within the framework of the diffusion model (Ratcliff 1978; Ratcliff and Mckoon, 2008), where a decision is dependent on two main factors. These are informational content, including aspects such as frequency and accessibility, perceived by the subject, and the time that passes between the introduction of the information and the moment the decision is made. It is understood that the information that affects the decision-making process contains noise that will pull the subject towards either one or the other decision, occasionally erroneously. Ratcliff and Mckoon (2008) show how a paradigm with two different processes will result in an erroneous RT that is slower than the correct RT. This is in keeping with the results from this study, which show a generally faster response time for correct responses, when compared to incorrect ones.

As stated previously, this study also measured the participants’ confidence in their answers. It has been shown that in experimental designs were participants are asked to rate their answers after having made them generally used Bayes optimal computations, where they based their answers on how correct an answer is supposed to be given the sensory data. In cases where the confidence rating had to be given prior to the answer, the participants used heuristic computations, in which it is the magnitude of the sensory data that influences the decision process (Aitchison et al. 2015). It can therefore be inferred that the sensory data, be it permanent or temporary, affect the decision-making process, and that the better fitting the Bayes computational method that a subject deploys, conscious or otherwise, ought to give more correct decisions, and a larger degree of perceived confidence in the accuracy of the answers. This fact impacts this study in two ways: it makes it possible to confidently state that the increased accuracy and greater confidence seen through the various blocks corresponds with the fact that the participants acted upon a growing quantity of sensory data, and as the exposure to the stimulus material increased, the models the participants used to make decisions became more accurate. It has also been found that there is a correlation between neural activity, error detection and confidence ratings, where the level of Error Positivity, that is, neural activity that appears after a conscious awareness of the fact that an error has been made, decreases as the confidence in the response increases (Boldt & Yeung, 2015). This kind of neural response is analogous to the MMN that has been studied here, even if produced under different conditions; it stands to reason that the MMN, in itself an involuntary response to an error, would see some correlation with processes such as confidence ratings and learning.

As this study deals with SL of linguistic stimuli, this learning had to be tested, to verify its presence, or absence. As discussed above, some of the processes employed to do this are RT and confidence ratings, that the literature has shown to be excellent measuring instruments. These two aspects were integrated into two different familiarity tests that were performed during the experiment. A familiarity test is used to determine whether or not a participant has a recollection of any given stimulus item, and how certain, along an established scale, they are of their statement (Evans &Wilding, 2012). Familiarity as a process is often studied through the Remember/Know paradigm (Tulving,1985; Gardiner and Java, 1993, quoted in Evans and Wilding, 2012), where Remember-reactions about stimuli involve situational recollections about a stimulus, whilst Know-reactions occur when a participant expresses a certainty that the stimulus has been presented before, but without the ability to give any further qualitative information. One can generally understand Remember-reactions as strong memories and Know-reactions as weak memories (Wixted, 2007, 2009; quoted in Evans &Wilding, 2012). This understanding of the stimuli on a Remember/Know level is useful when considering how the stimuli were internalised by the participants, and how this internalisation interacts with the MMN that was measured. It becomes obvious that the participants would internalise the stimuli on a Know level, much more than a Remember level, considering the context in which they were tested for retrieval and confidence rating. While it may be possible that the participants remembered the stimuli as a whole after the experiment, given the in-situ exposure context, this would not be applicable to the decision-making situations during the experiment, since there were no particular situations or contexts to connect to the singular lexeme. This is also in keeping with the understanding of the MMN as a subconscious pattern breaking reaction; if the lexemes were established on a higher consciousness level, a Remember level, a break in the expected string of stimuli would lead to a different kind of neural response.

This study inherently deals with complex stimuli, in that the information contained within them is presented through several different modalities, e.g. phonological content, prosody and stress. As will be discussed later, this presents a problem when it comes to the accuracy in the understanding of the causal relationship (cf. Toomela, 2023), as the various modalities contribute towards the same processes. This use of complex stimuli, however, can be grounded in the literature; Wichmann and Jäckel (2018) state that although many psychophysical methodologies use simple stimuli, this is not a strictly necessary limitation, and more complex stimuli may be used, in particular to study processes that are larger and more complex, like learning.

3.3 Stimuli

The stimulus material consisted of a selection of Russian words. These were chosen to explore how the difference in phonetic content and prosody affected the learning results and the EEG recordings.

The software used to produce the stimulus material was Amazon POLLY (Amazon, n.d), a text-to-speech (TTS) generator. The stimuli consist of a series of word roots, to which two possible suffixes are attached. The following six word roots were used:

- Дви́га – dviga- - [ˈdvʲigə]
- Храни – khrani- - [xrɐˈnʲi]
- Мечта – mechta- - [mʲɪt͡ɕˈta]
- Слуши – sluzhi- - [ˈsłuʂə]
- Спаси – spasi- – [spɐˈsʲi]
- Води – void- – [vɐˈdʲi]

These roots have been chosen due to their phonological content and shape. We wanted to use roots that were two syllables in length, and that differed as much as possible from each other in their phonatory profile. One step in this was choosing words that began with different consonants, or, in the case of “sluzhi” and “spasi”, syllable onsets that varied markedly in voicing.

The suffixes that were added to the roots were теля – telya - [tʲɪlʲə], and телем – telyem – [tʲɪlʲəm]. These resulted in twelve different test items. Initially recordings of L1 speakers of Russian were tested as possible stimulus item sources, but rigorous editing and recording selection notwithstanding, a synthetic voice generator was chosen to minimise the effects of recording quality and other error factors in the L1 recordings.

Our wish was to design our study as close as possible to the paradigm of Koelsch et al. (2016), but we are unsure whether three word roots would be sufficient to facilitate learning, given the evidence thar a high repetition exposure to fewer roots gives poorer learning than a high variability context administration of the same roots (Eidsvåg et al., 2015). Based on this we chose to include six word roots.

3.4 Participants

The participants were mainly recruited from the different faculties at the University of Bergen. This was mostly due to the location of the EEG laboratory where the experiments took place, as we considered it to be easier to recruit participants that already were in a general proximity to the laboratory, and thus would not have to burden themselves with extra travels to participate. They were however, compensated monetarily to cover costs of transportation and time used. We were hoping that students would be open to participating in a study of this kind.

Our aim was to have between twenty to thirty people participate in the experiment. This number was chosen as to give the study enough statistical power to safely be able to give the proof-of-concept, all the while being few enough that the study did not require an unrealistic amount of time and resources. We were hoping for the participant pool to have an equal gender distribution, as this would make it possible to discover any potential gender-based differences. However, this was not possible.

The participants were recruited through posters with tear-away phone number slips (See Appendix D). The posters gave a brief introduction to the study, as well as inclusion and exclusion criteria for participation. The inclusion criteria were to not have any self-reported history of psychological, neurological, or psychiatric disorder or conditions, and be between the ages of 18 and 65. The exclusion criteria were that the participants could not have any knowledge of Russian or any other Slavic language, have no known auditory injuries or deficits, and that they must not be neurodivergent. It has for instance been proven that people with various kinds of cognitive impairments have slower or less intense electrical cerebral reactions than neurotypical peers (Hugdahl, 1995). We wished to avoid this potential error factor and found the exclusion of these participants to be the easiest solution. Another reason to exclude potential neurodivergent participants was that the knowledge of this kind of health information would make the management of participant privacy unnecessarily complicated. We later added an exclusion criterion regarding musical training: the participants must not have musical training beyond mandatory schooling or play any instrument.

The data collection occurred between October 6th and December 13th 2022. The selection consisted of 14 women and 6 men, aged 20-50 (mean age, 26.75, SD = 8.47). Even though our recruitment posters were written in Norwegian, we received interest from a few non-Norwegian speaking students as well. We discussed with our supervisor if this could have any impact on our findings, but we decided to let them participate. It could be interesting to see if multilingualism might affect the data in any way. In the end this resulted in eight participants who stated a different mother-tongue than Norwegian. These were Tigrinya, German, Cantonese, Chinese, Turkish, Urdu, Romanian and Arabic. None of the participants reported any knowledge of Russian in advance of the experiment.

An additional exclusion criteria was added after the distribution of the information posters; that the participants could not have any musical training beyond mandatory school lessons. This created some confusion for the participants, and it also had the unfortunate effect that we lost a few potential participants very close to the data collection. Eight of the participants reported to having received musical lessons in the past, but none of the participants were receiving any lessons currently and had not done so for several years. In agreement with our supervisor, we found that this was acceptable.

3.5 Recording

When the participants arrived, they were presented with a general introduction to the experiment. The circumference of the participant’s head was measured, to find out which electrode cap would provide the most beneficial and optimal recording conditions. Then, they were given a copy of the information sheet (see Appendix B) that they had read previously, as a reminder, together with a declaration of consent (see Appendix A), wherein they signed to agree to the storage of their data and the potential anonymised distribution of said data to third parties, for instance in a public repository (such as OpenScienceFramework, https://osf.io/).

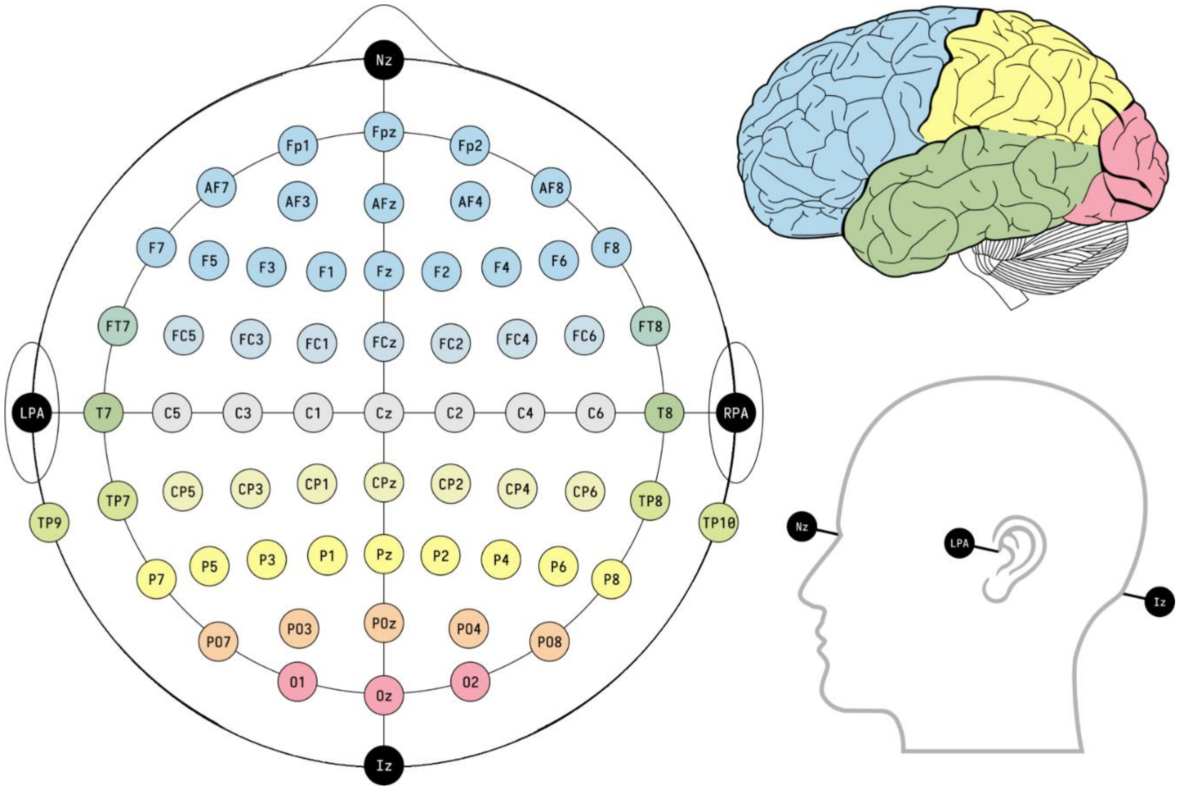
After the participants had signed the declaration of consent, they were provided with a pre-test that explored the linguistic and the musical background of the participant. The participants were asked to declare their level of proficiency in a selection of languages (English, German, French, Spanish, and Polish), in addition to any other unspecified languages they spoke. The spread between the languages spoken by the different participants was considerable, as stated. It was interesting how all participants rated themselves to be close to or outright bilingual, and how the participants whose L1 differed from Norwegian tended to be polyglot, having reached excellent proficiencies both in their L1s, English and Norwegian. Research shows that people who know several languages are more sensitive to a wider range of possibly meaningful input, if that modality of semantic transmission is present in a language they already speak (Potter, Wang & Saffran, 2017). The participants were also asked if they had received any musical training outside of mandatory school lessons, or if they played any instruments. Research shows that musical training has an impact on the ability to perceive variation in pitch in language (Schön, Magne & Besson, 2004), which has been linked to improved SL of language (Erickson & Thiessen, 2015). We wished to exclude trained musicians, as to eliminate the potentially confabulatory factor of improved pitch distinction from the learning process, and to give the study more naturalistic validity, since the majority of the population is not trained musicians.

While one of the authors stayed with the participant during the completion of the declaration of consent and pre-test, the other moved to the testing cabin to begin the assembly of the electrode cap. This entailed attaching the electrodes to the appropriate clip in the cap and preparing the remaining three artifact-capturing electrodes. The electrode that was to be fastened on the cheek bone was prepared with a ring-shaped adhesive which permitted the injection of the electrode gel in a same fashion as the scalp electrodes, whilst the two electrodes that measured artefacts by the eyes were prepared by laying a piece of tape on the non-sensitive side, and filling up the internal space with gel, so that it would create a satisfactory connection.

The positioning system used was based on the 10-10 system. See Figure 1 for precise electrode placement.

Figure 1

Image Shows the Electrode Placement Used in the Experiment



Note: Electrode layout used during the experiments: The electrodes in colours (i.e., not those in black) were recorded during the experiment. TP9 served as reference electrode during the acquisition, Iz as ground electrode. Not shown are the electrodes used to track vertical (on the right chin) and horizontal eye movements (at the outer canthi of the eyes on either side). Image adapted from "EEG 10-10 system with additional information" by L. R. Krohl, 2020. (https://commons.wikimedia.org/wiki/File:EEG_10-10_system_with_additional_information.svg) CC 0

We used a total of 66 electrodes, where 63 were placed on the scalp. In addition, one was positioned on the right cheekbone and two were positioned at the participants' temples to measure the galvanic skin response, and map vertical and horizontal eye-movement. The electrode positioned at TP9 was used as mastoid conduction reference points, and a grounding electrode, not to be counted among the beforementioned 63, was positioned at Iz. The EEG-read outs were recorded on a computer running Windows 10, using Brain Vision Recorder version 1.25 (Brain Products GmbH, 2019). This recording software gave us the possibility to customise our electrode mapping, so as to obtain an accurate and efficient tool.

When the participant had filled out the declaration of consent form and the pre-test, certain key parts of the face and head were scrubbed with an alcohol-wipe. This was done to improve the connectivity in these areas. The areas in question were the application point of the artifact electrode on the cheek bone under the right eye, and on the mastoid process behind each ear, where the referencing TP9 electrode was later placed.

After the participant had been prepared, they were led into the testing cabin and seated in front of a desk fitted with a computer monitor and a keyboard. The complete electrode cap was then placed on the head of the participant, taking care to line the Nz-Iz axis of the cap up with the nasion-inion axis. The Cz-placed electrode was then pressed, and the participant asked to evaluate if the placement of the electrode was in the middle of the skull. When the participant was satisfied with the placement of the electrode, the authors proceeded to attach the artefact electrodes to the cheekbone and by the eyes. The cables from the cap were fastened to the participants chair by tape as to not weight down or pull on the participant's head.

At this point we turned on the monitor to display the resistance overview in the Brain Vision Recorder software. This software, together with lamps on the electrodes themselves, indicated the amount of resistance for the singular electrode.

Once this was done the authors started using the gel-filled syringes to create connections between the singular electrodes and the skin. The first electrode to be worked was the grounding electrode placed over Iz, as the other electrodes could not form a circuit, and thus not be active, without this. Once the resistance of the grounding electrode was at a satisfactory level, we started establishing contact in the other electrodes. The common order was to start with the Cz- position, and then continue along the midline of the brain. We then worked in rows out from the established midline, and the last electrodes to be established were usually the ones above the ears (FT9-T7-TP9 and FT10-T8-TP10). This order of procedure notwithstanding, the establishment of satisfactory connectivity levels required some revisitation of the original injections. The gel either had to seep into the hair of the participant, or the electric bridge was broken by the hair pushing the cap away from the scalp. This reworking of the connections continued until the resistance was acceptably low, as indicated by the light on the screen and on the electrodes turning green.

When we were satisfied with the connectivity, the monitor was made to display the readings from the electrodes. This way, we could verify the quality of the connections. The participants were provided with an explanation on what could and could not be read from the EEG-readout, and asked to perform some tasks to demonstrate artefacts, like eye movements or laughing. The explanation regarding what could be read from the EEG was necessary insofar as several participants expressed worries about both the anonymity and the implications of the wave forms. Some were worried we could measure intelligence, others read thoughts, and yet others that the wave signature would be particular to him or her. These worries were soothed.

Once this was done, the authors changed to input source of the monitor to display the start screen of the experiment. The authors exited the room and set the light levels to a low, consistent level for all participants. The participants were asked to consign any electronics, to be kept outside the cabin. After making sure the recording was running, we started the stimulus administration.

Once the experiment had been completed, the light was turned on, and we entered the cabin to continue to the next step. The EEG cap was removed, and one author started the washing process of the cap and electrodes. The other author administered the post-test, in which the participants were asked to rate if they experienced patterns in the stimuli, and to give confidence ratings of the presence of a selection of possible suffixes (see appendix C). When the participant had finished one author would lead them to washing facilities so they might remove the gel. After this the participants left.

3.5 The experiment

The experiment itself consisted of six blocks. These were further broken down into a) stimulus administration, b) an item differentiation task based on learning of probable root and suffix combinations, with a confidence rating for each answer, and c) rest before the next stimulus administration block.

In section a) stimulus administration, the participants were seated in front of a computer screen whilst wearing headphones. Through the headphones, a randomly generated string of possible root-and-suffix combinations played, where one of the suffixes had been chosen at random to be low probability, whilst the other was high probability. The stimulus items were standardised to ≈800 milliseconds, and so the participants heard approximately 600 stimulus items per block.

Whilst the participants were being subjected to the stimulus stream, a nature documentary was playing on the computer screen. This was done to ensure that participants would remain focused throughout the stimulus administration. As mentioned in the section on the feedback the authors received, the stimulus administration had a soporific effect. Several of the participants presented with alpha wave patterns, which is consistent with their feelings of drowsiness. The nature documentary did contain indirect scenes of death of animals, but only one participant mentioned this after the experiment was over, and we believe that these scenes have not negatively impacted the EEG data.

During the experiment, another measure was implemented to make sure the participants remained focused throughout the stimulus administration. The stimulus material was mostly pronounced by a female voice, but on occasion, approximately six times per block, a male voice would pronounce either a root or a suffix. When they heard the male voice, the participants were supposed to press the space bar on the keyboard. The participants would get feedback on how many male voices that had played, how many times they were successful in pressing the space bar in time, and how many times they had pressed the space bar unwarranted.

In section b) the familiarity test, after the stimulus administration had finished, the nature documentary was paused. The participants were presented with a screen with instructions on how to proceed. The next step, as mentioned earlier, was that the participants were acoustically presented with two equal roots with different suffixes, where one had had a low probability of appearing in the preceding stimulus administration block and the other had had a high probability, and were asked, by pressing either “1” or “2” on the keyboard, which variant they thought sounded most familiar. After they had selected one of the two alternatives, they were asked to judge their confidence in their answer on a scale from 1-5, with the corresponding numerical key on the keyboard. This was repeated ten times, so that participants gave double answers to at least four root-plus-suffix combinations.

Once section b) had finished, the participants were given the possibility to rest, refresh themselves, or make contact with the authors. These paused were sometimes used by the authors to go into the cabin and make adjustments to the EEG cap, as some electrodes would, on occasion, lose contact with the scalp and thus provide unusable or disruptive input.

3.7 Feedback

On the post-test, the participants were invited to give feedback on the experience. This feedback was quite varied. Two of them complained that the lights were too low, and that they almost fell asleep. In the same comment, a wish was expressed for an opportunity to control sound and screen lighting. One participant commented that the room was too cold, which made it a bit uncomfortable. They also mention that the darkness made it hard to locate the correct letter on the keyboard. This could be solved by highlighting the important letters, for instance by using a yellow sticky tape. When it comes to light and sound these are factors that are not to be controlled by the participants. Light and sound are controlled variables, and interference with this these might influence the measurements.

One participant expressed annoyance due to the monotony of the stimulus material, and how they had and urge to “take off the headphones and throw them against the wall”. We are unsure how this kind of emotional response could have affected the participant’s results. However, this is considered a relatively normal reaction according to Beres (2017) who explain that participants can get tired or fed-up with long, monotonous studies. Other participants have also voiced varied negative emotions towards the stimulus material, but this was more akin to frustration for not being able to decipher the pattern and meaning behind the stimulus material.

4.0 Results

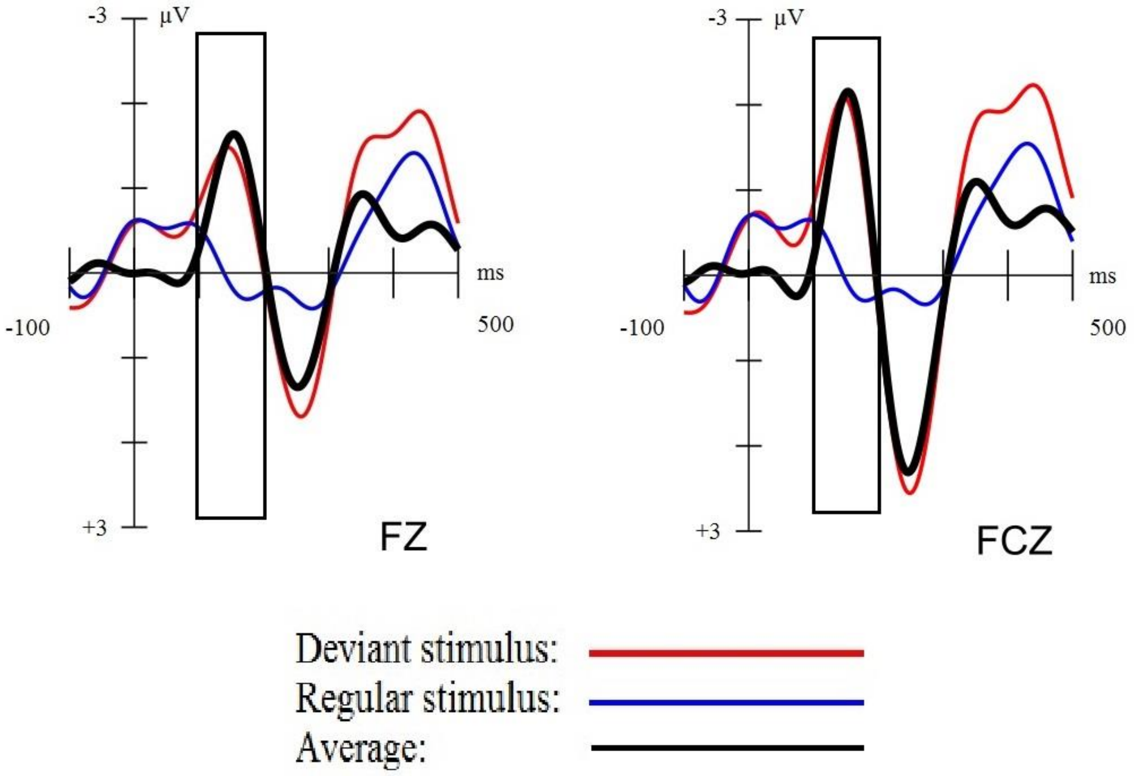
4.1 EEG results

4.1.1 Physical MMN

A phMMN could be observed in a time window between 100 and 200 ms with a maximum amplitude of approximately 2 µV, an amplitude peak around 150 ms (see Figure 2)

Figure 2

Image Shows Voltages of phMMN ERPs

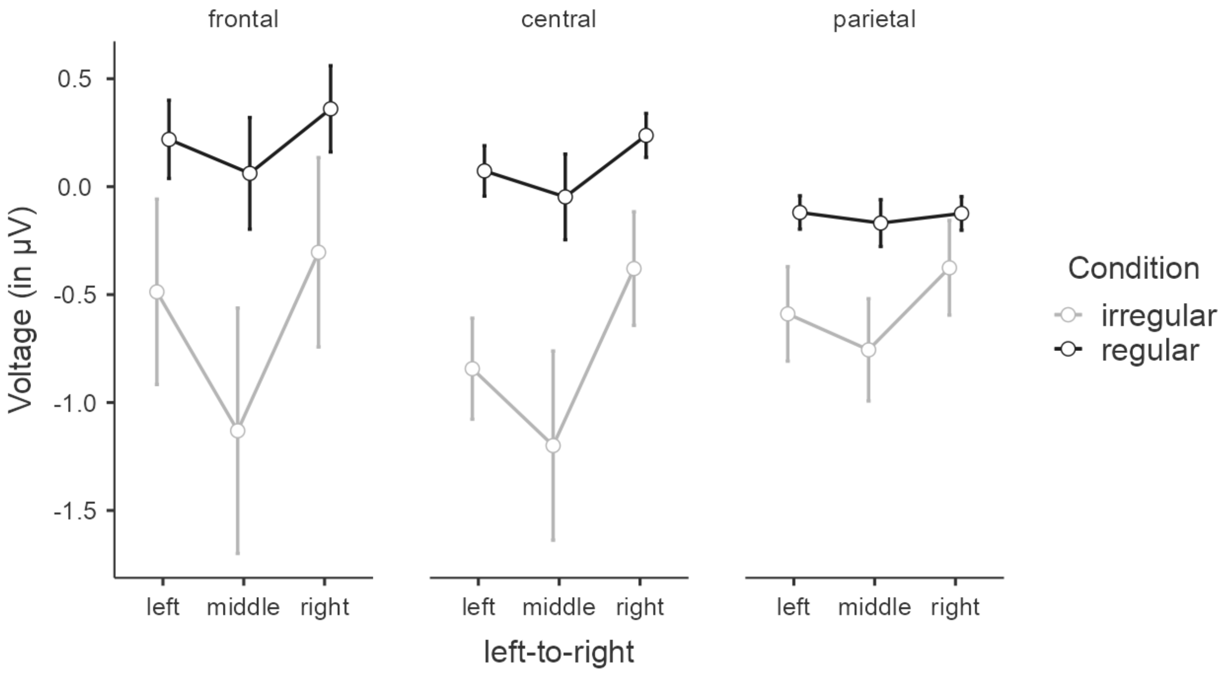


Note: Average ERPs of deviant and regular stimuli for the phMMN, i.e., a change in the physical characteristics of the sound coming either from the standard direction (regular) or the opposite site (deviant). Shown are the electrodes FZ and FCZ electrodes, and over the course of the experiment and indicating the time window of 100 to 200 ms after stimulus onset that was used in the statistical analyses.

The phMMN was evaluated using an ANOVA with the factors condition (with vs. without location change), anterior-to-posterior (frontal, central, parietal), and left-to-right (left, central, right). It revealed a significant effect of physical deviance condition $F(_{1,20})$ = 47.01, p <.001, $\eta^2 = .25$, as well as interactions of condition with anterior-to-posterior, $F(_{1,20})$ = 4.99, $p = .01$, $\eta^2 = .02$, of condition with left-to-right, $F(_{1,20})$ = 12.88, $p = <.001$, $\eta^2 = .02$, and of condition with anterior-to-posterior and left-to-right, $F(_{1,20})$ = 4.15, $p = .004$, $\eta^2 = .002$. This is illustrated in Figure 3.

Figure 3

Image Shows Voltages in phMMNs Across Scalp Locations



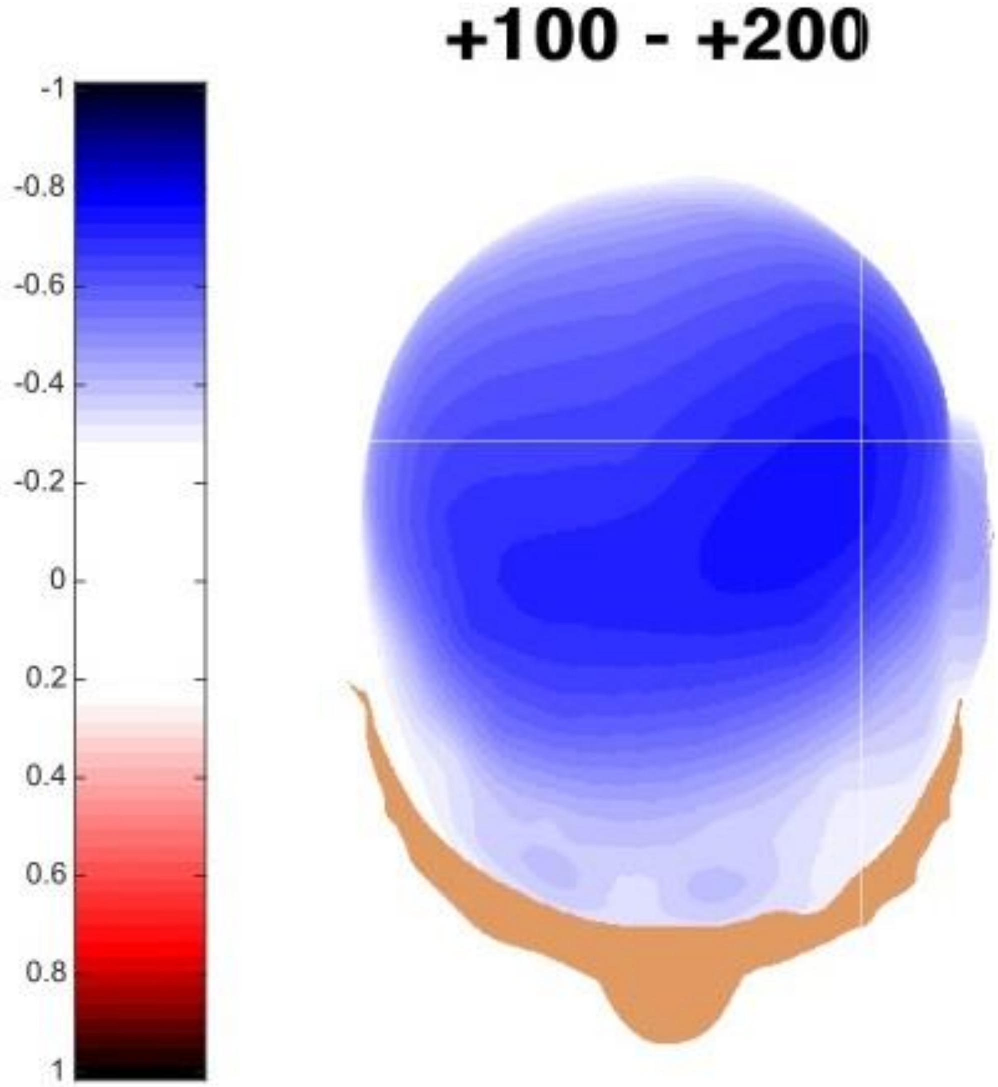
Note: Amplitude differences to a change in the physical characteristics of the presented stimuli, i.e., between the sounds coming from the standard direction (regular) or the opposite direction (irregular), on the anterior-posterior axis, and along the left-to-right axis on the scalp. The greatest activation amplitude difference could be observed at scalp areas around the midline laterally, and in the fronto-central areas along the anterior-posterior axis, resulting in a significant interaction of condition, anterior-to-posterior and left-to-right.

In addition, the ANOVA revealed further significant effects not involving condition, a main effect of left-to-right, $F_{(2,38)} = 24.15$, $p < .001$, $\eta^2 = .06$, and the interaction of anterior-to-posterior and left-to-right, $F_{(2,38)} = 13.77$, $p < .001$, $\eta^2 = .01$, reflecting that also the sounds from the standard direction elicited a brain response that was (although on average having a positive amplitude) lowest over central scalp areas compared to both the left and the right hemisphere (with this pattern more pronounced over frontal and central than parietal sites).

As illustrated in Figure 4, the main effect of condition reflects that there was an amplitude difference between regular and irregular sound direction that generally was present over the whole scalp. The interactions with condition reflect that though the phMMN could be observed at the whole scalp, it had an amplitude maximum at fronto-central scalp sites and along the midline, however, with a slight lateralisation to the left hemisphere.

Figure 4

Image Shows Scalp Potentials for phMMN



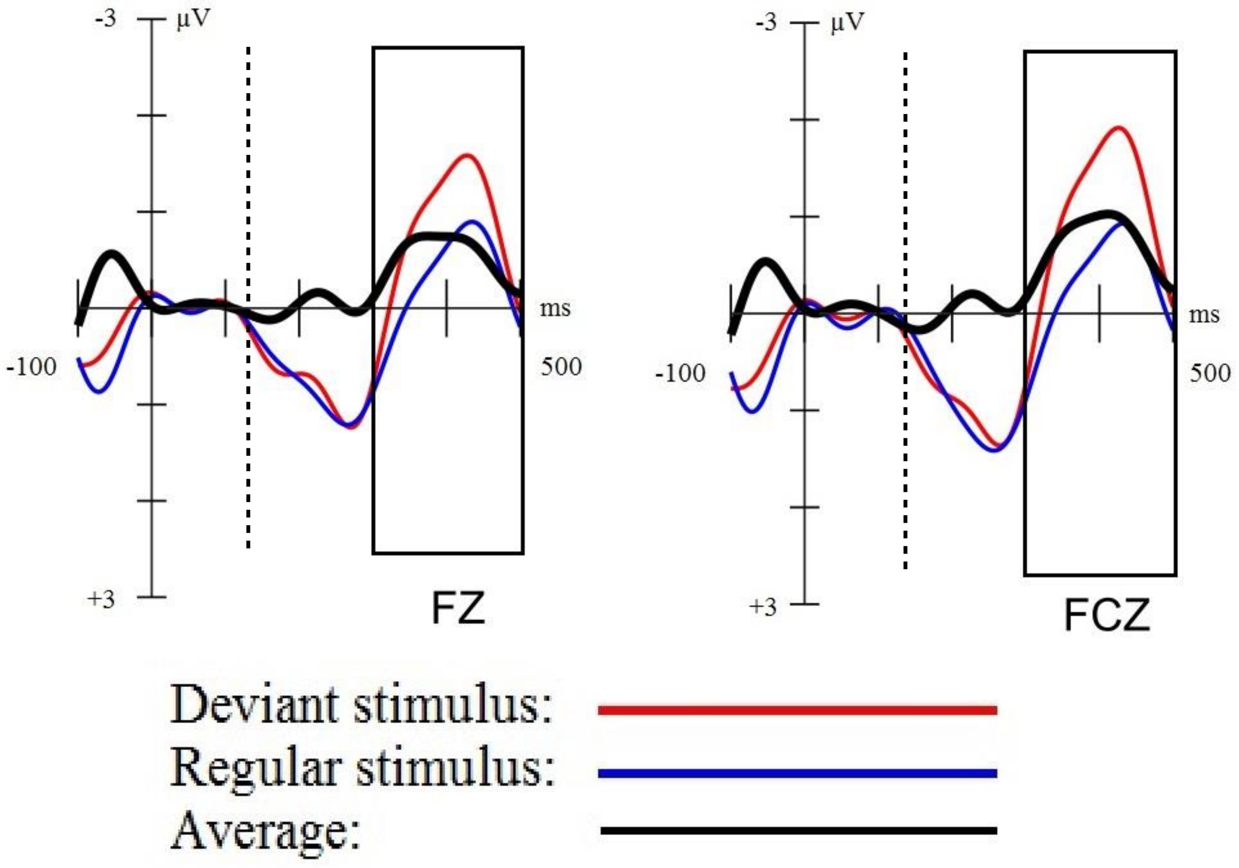
Scalp distribution of the phMMN within the time window 100 to 200 ms. The amplitude difference was largest of fronto-central scalp areas with a slight lateralisation to the hemisphere.

4.1.2 Statistical MMN

An sMMN was elicited as difference between the irregular (i.e., less often – 10%) and the regular (i.e., more often – 90%) word endings. Given that the first part of the ending (-tel) was identical for both endings (-telya or –telyem), the onset of the brain response had to be adjusted for the length of -tel (164 ms): the used time window between 300 to 500 ms is therefore in fact between 136 to 336 ms (relative to the end of -tel). This is illustrated in Figure 5. Within that time window, the brain response to the irregular stimuli has a more negative amplitude than the brain response to the regular stimuli, resulting in a sMMN. The sMMN is most prominent over fronto-central scalp sites, and generally largest over midline electrodes (but with a slight lateralisation to the left) and has an amplitude maximum of approximately 1 µV peaking at around 240 ms (404 ms relative to the onset of –tel). That is, compared to the phMMN, it had a similar scalp distribution, but a smaller amplitude size (2 vs. 1 µV) and a higher latency (150 vs. 240 ms). This is illustrated in Figure 6.

Figure 5

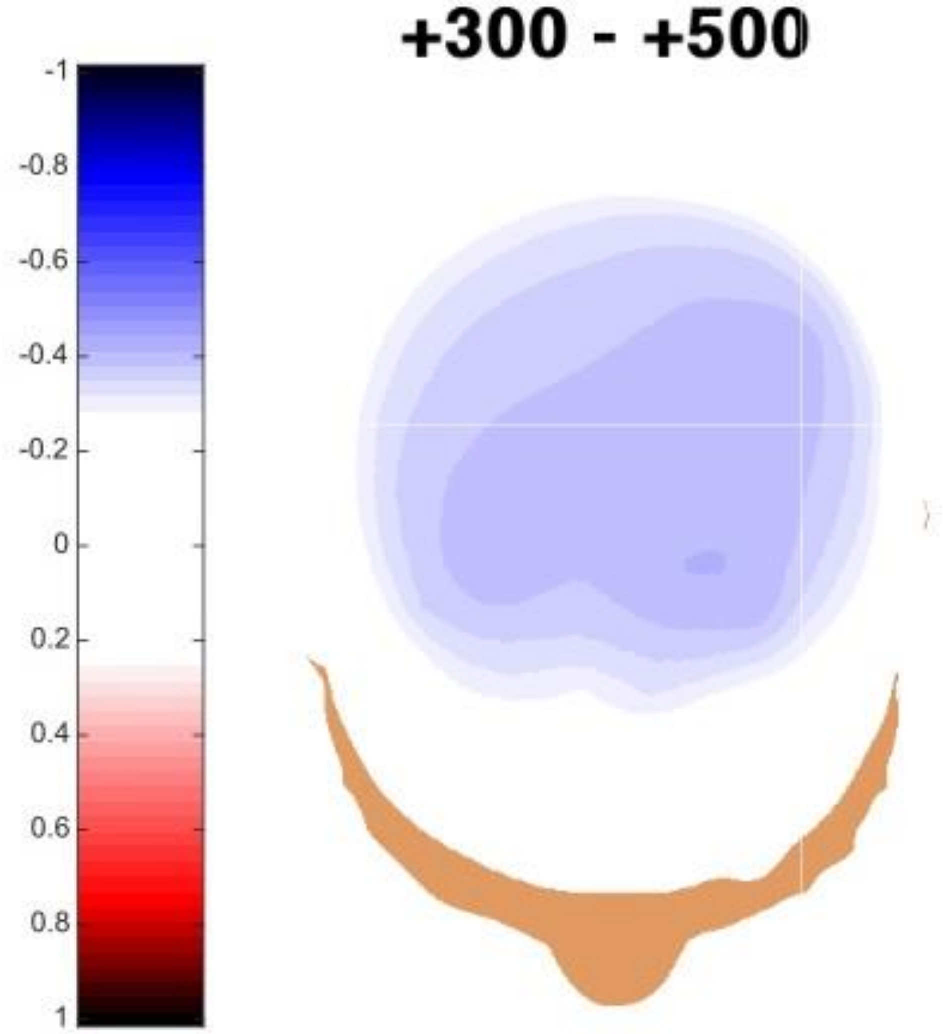
Figure Shows Voltages of sMMN ERPs



Note: Amplitude differences for the sMMN, i.e., between the irregular and the regular endings, and the at the electrodes FZ and FCZ and for the whole experiment. The time window was 300 to 500 ms after the onset of the ending, i.e., 136 to 336 ms after the ending began to differ between the regular and the irregular stimuli (both endings shared their first part -tel, lasting 164 ms). The dotted line marks the ending of -tel-.

Figure 6

Image Shows Scalp Potentials for sMMN

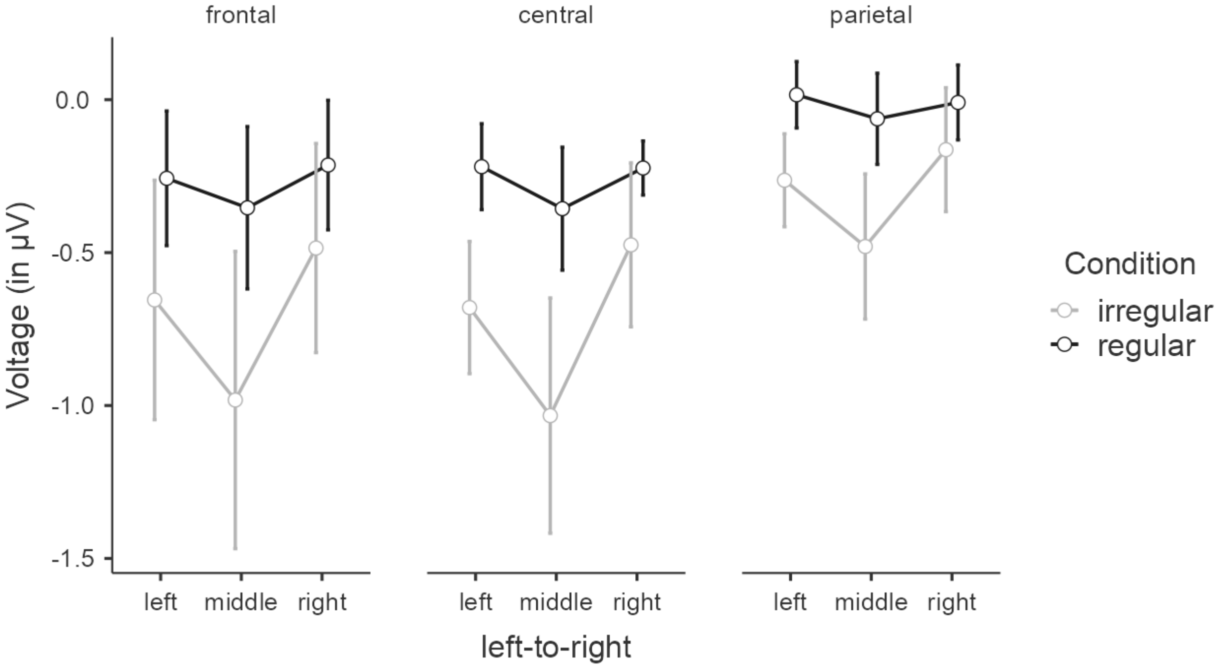


Scalp distribution of the sMMN, within the time window 300 to 500 ms (given that the first part of the ending -tel with a length of 164 ms was the same for both conditions, the time window is 136 to 336 ms after the endings began to differ). The scalp distribution was maximal over fronto-central electrodes and was slightly lateralised to the left hemisphere.

The sMMN was statistically analysed with an ANOVA for a time window of 136 to 336 ms relative to when the two endings became distinguishable. The ANOVA had the factors condition (regular = high transition probability vs. irregular = low transition probability), anterior-to-posterior (anterior, central and posterior) and left-to-right (left, midline and right). It revealed a significant effect of condition, $F_{(1,19)} = 14.40$, $p < .001$, $\eta^2 = .11$, and an interaction of condition and left-to-right, $F_{(2,38)} = 9.16$, p = <.001, $\eta^2 = .01$. This reflects that the amplitude difference in the brain responses to the irregular vs. the regular endings could be observed over the whole scalp with the largest differences to be found around the midline (see figure 7).

Figure 7

Image Shows Voltages in sMMNs across scalp locations



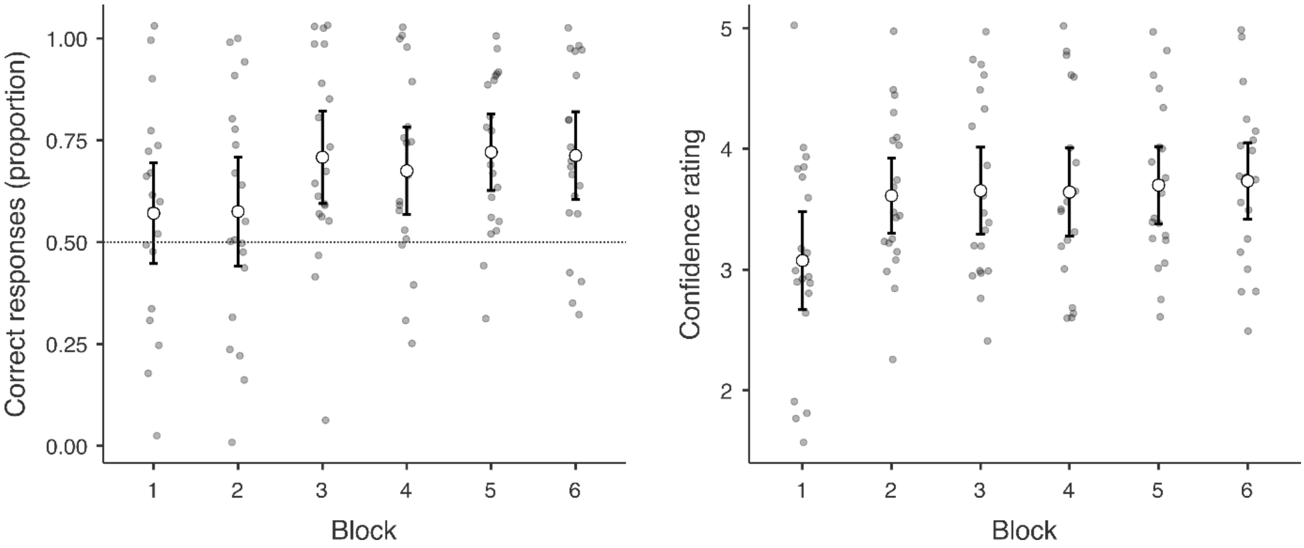
Amplitude differences between the statistically regular (black lines) and the irregular stimuli (grey lines). The three diagrams show the anterior to posterior axis, and within each diagram, the left to right distribution can be found on the x-axis. The greatest amplitude difference could be observed at scalp areas around the midline. The amplitude difference was, in addition, more pronounced over central and frontal sites than over parietal sites along the anterior-posterior axis of the scalp. The greatest activation during irregular stimuli occurred in the frontalo- central areas of the scalp. The interaction of the factors condition (irregular vs. regular) and anterior-to-posterior scalp distribution was not significant.

4.2 Behavioural data and analysis

In addition to assessing the neurophysiological correlates of the acquisition of the transition probabilities between word roots and endings (i.e., the sMMN), a familiarity test was used to assess the behavioural correlates of this acquisition process. In the familiarity test, participants had to decide which of two words (one with the regular and one with the irregular ending) sounded more familiar. Given the decision between two sequences, the chance level was 0.5 and correct recognition significantly above that level would indicate successful acquisition at the behavioural level. (Figure 8, left panel).

Figure 8

Recognition performance and confidence ratings for the implicit-meory-based task



Recognition performance: percentage of how often to regular ending was chosen during the implicit-memory-based familiarity test at the end of each block. Given that the participants had to decide between two sequences, the base rate was at 0.5 and a recognition performance above that level can be regarded to reflect the implicit learning of the transition probabilities between the roots and the endings. Right panel shows confidence ratings across blocks.

The recognition performance (see Figure 6, left panel), was assessed over the 6 blocks of the experiment. Whereas the recognition performance was not above chance level for the first two blocks ($p > 0.123$), it was above chance level for the blocks three to six: t(19) > 3.44, $p \leq 0.001$, d > 0.77, with a correct recognition rate of at least 67.5%. The recognition performance over the whole experiment (i.e., the mean of all six blocks) was 66.0%, which was also significantly above chance level: t(19) = 3.69, $p = .001$, d = 0.83. An ANOVA assessing the recognition performance over the experimental blocks (block 1 to block 6) revealed a main effect of block, $F_{(5,95)} = 3.75$, p = .004, $\eta^2 = .07$.

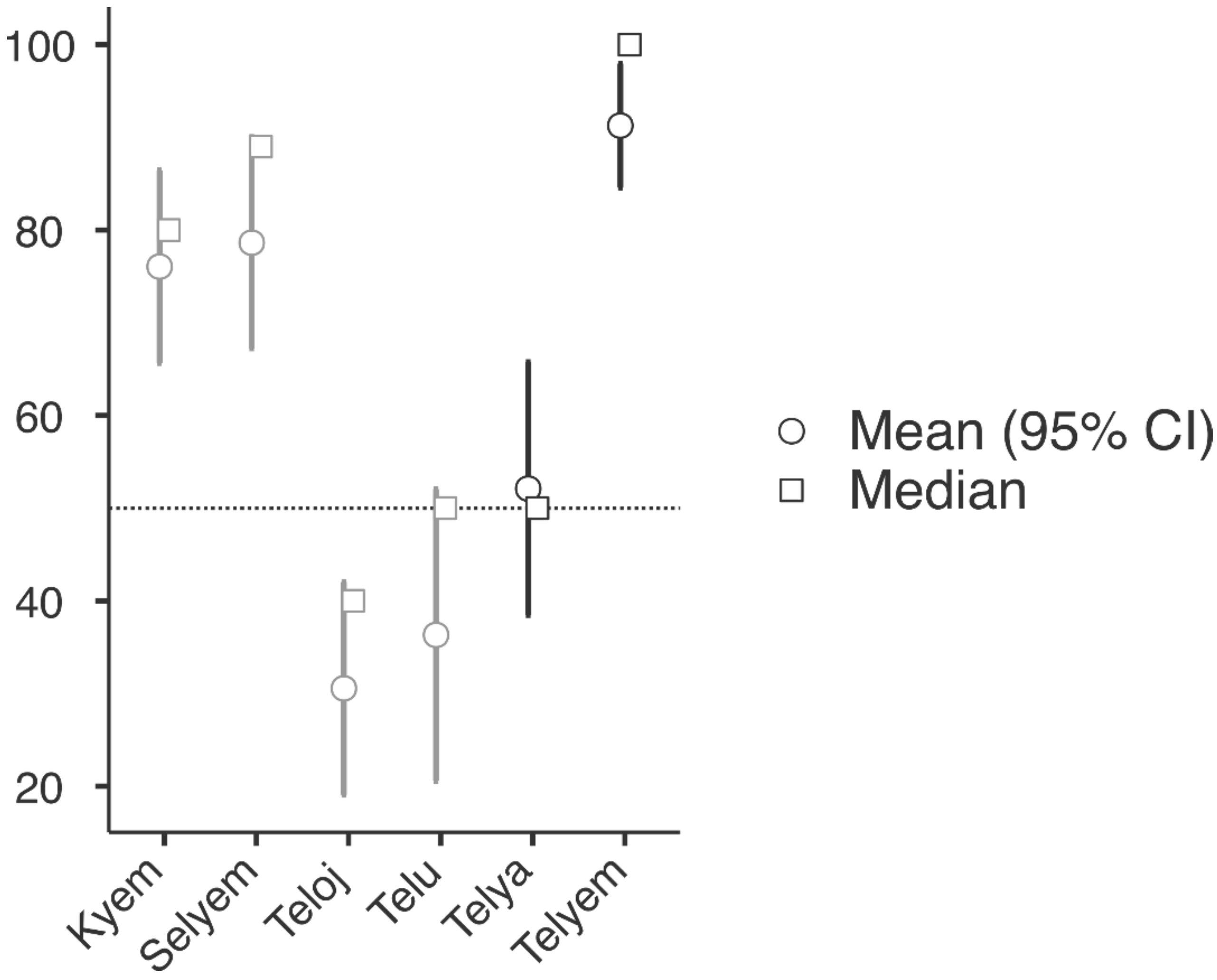
Participant assessed their confidence regarding whether they had chosen the correct sequence on a scale from 1 to 5 (Figure 8, right panel). They rated their confidence at a mean of 3.08 in the first block and between 3.61 and 3.73 for the second to the sixth block. An ANOVA with the factor blocks revealed a main effect of blocks, $F_{(5,95)} = 8.15$, $p < 0.001$, η^2

= .09. It was furthermore assessed whether the confidence ratings (over the whole experiment) differed between trials where the "correct" word (i.e., one containing an ending that had a 90% transition probability) was chosen, compared to those where the "incorrect" word was chosen (i.e., one with a 10% transition probability). The mean confidence ratings were M = 3.11 for "incorrect" and M = 3.60 for "correct" trials, leading to statistically significant difference between those two conditions: $F(_{1,18}) = 14.46$, $p = 0.001$, $\eta^2 = 0.13$.

At the end of the experiment, there was an additional explicit-memory-based familiarity test. Here, the two endings used in the experiment were compared to four endings that were phonologically similar to the presented ending but not used in the experiment. Ratings for each ending could range from 0 to 100, with 0 signifying absolute certainty of the endings's absence, 50 signifying uncertainty, and 100 signifying absolute certainty of the endings's presence. The data are shown in Figure 9. The mean scores for the not presented decoy endings were "Kyem" = 76.1, "Selyem" = 78.6, "Teloj" = 30.6, and "Telu" = 36.3; for the presented endings, they were "Telya" = 52.1, and "Telyem" = 91.3.

Figure 9

Image Shows Confidence Scores for the Explicit-memory-based task



Note: Confidence ratings for the different possible endings, in the explicit-memory-based test after the experiment. The black graphs represent the endings in the experiment, and the lighter graphs represent the decoy alternatives. The scores were 0 = Absolutely certain of absence, 50 = Unsure, and 100 = Absolutely certain of presence.

5.0 Critique of procedure and analysis methods

When reviewing and criticising the methodology in this study, will be useful to consider the various elements that have been discussed so far in the same order as they were presented. Firstly, EEG as a measuring tool and its weaknesses, with ERPs and especially MMNs as focus. Afterwards, how the familiarity tests and experimental design itself may be sources of error. The main focus of the critique will be error sources in the measuring and design, that could potentially lead not only to faulty data material but also to a faulty interpretation of those data, cf. Toomela’s (2023) thoughts on understanding what has actually been measured.

The accuracy of the EEG recording in this study can be victim to two potential error sources: faulty recording, and faulty data treatment after-the-fact. It is the case that methodologically, the EEG recordings will be suboptimal, given aspects of the recording process like participants' hair causing resistance drift in the electrodes. While the more obvious examples of electrode malfunction were removed manually after recording, any slight drift in the singular electrode that escaped manual detection will be an error source. However, given the long tradition of using EEG as a clinical and scientific tool, it can generally be understood to be accurate as a measuring tool.

The potentially greatest error source for the EEG data was the manual treatment of the raw data in preparation for ICA and time-averaging, because this step introduced human error as a part of the equation. Whilst we did operate with criteria for when to remove an entire electrode's reading from the data, i.e., more than 5% of the data from that electrode was erratic, other episodes of erratic behaviour where one would rather remove an entire section worth of data across all electrodes was left to the judgement of the authors. In addition, it must be said that while the ICA was run by the computer, it was left to the authors to select which artifacts that should be removed, and which should be kept.

There are several weaknesses in the ERP as a concept that might render it inaccurate. The first one is the thought, as mentioned above, that an ERP is an activation of neural populations that will become visible once time averaged, when the noise of the spontaneous background noise is removed. This leads to what Da Silva (2004, p. 994) categorises as the "special problems of event potential (EP) analysis": (a) the question of the relation between the EP and the background noise, (b) the problem of detecting and classifying single trial EPs, particularly in order to improve the averaging procedure, (c) a posteriori filtering (Wiener filter) to improve the estimation of the average EP, (d) ways of controlling the averaging procedures, and alternative techniques, and (e) the topology of EPs in relation to the corresponding anatomical sources.

Da Silva (2004) states that for problem (a), the main goal is to reduce the signal-to-noise ratio as to make the signal as clear as possible, but that this process can be affected by the nature of the background activity. If the background noise is strongly rhythmic in nature, the noise quotient in the ratio will be affected. This naturally varied between participants, with some participants expressing that they had closed their eyes to concentrate on the stimuli, or that the experiment had a soporific effect, whilst others remained more alert.

For problem (b), the process towards finding the single ERP in this study has been defined by a series of criteria; firstly, what kind of ERP is being looked for. Since this study is actively looking for MMNs, it does not have to discriminate what kind of ERP is being produced by a stimulus, which is the main difficulty proposed by De Silva (2004). Secondly, given the nature the time averaging procedure, if a different kind of ERP were present or even dominant, the MMN found through time averaging would differ significantly from similar studies (like Jentschke et al, 2016; Tsogli et al., 2019), or a different ERP would become apparent though force of presence. This of course does lead to the potential weakness of looking for a reaction that is not present, or that behaves differently from what is expected; this can affect the analysis of the averaging results.

Da Silva (2004) states that problem (c), and this kind of filtration, is more relevant for data with few ERPs, as this reduces the signal-to-noise-ratio for these kinds of data sets. Given how this study has several thousand stimulus administrations, this problem is not pertinent to this study. Whilst it is true that the study has used frequency filters, these were applied a priori, and thus only affected frequencies that would be known to be erroneous in the recording.

Problem (d) however, is very much pertinent to this study. De Silva (2004) states that the averaging procedures, like the time averaging used in this study, is liable to several sources of error. An additive approach runs the risk of being affected by the background noise (Sayers et al, 1974, quoted in De Silva, 2004). Therefore, the averages of the noise should also be taken into consideration. A step in this direction has been taken in this study through ICA, which seeks to reduce the greater part of variance in the noise.

When considering problem (e), this study has gone to great lengths to ensure accuracy in its measuring the topographical nature of the evoked ERPs. De Silva (2004) states that this is best achieved through a high concentration of electrodes that are closely spaced, and, considering the placement of the electrodes on the scalp in this study (see Figure 1 for reference), this must be said to be achieved. This will also be useful when considering the sMMNs evoked in this study with other, topographically specific negativities like the ELAN (Friederici, 2002) and ERAN (Koelsch, 2009). Whilst it is true that this later ERP has been proven to be related to musical information already present in long term memory, considering the connection and overlap in neural resources between music processing and syntax (Koelsch et al, 2005), and this study’s focus on prosody, keeping these aspects in mind is nevertheless important.

When considering the familiarity tests as measuring devices, it is useful to spend some time considering what these tests are set out to measure: the result of the interaction between the stimuli and the participants. As such, it is proper to start with considering the stimuli itself.

This study uses natural linguistic stimuli, and it thus runs the risk of obtaining a result which, due to the complex nature of these stimuli, it is not possible to be certain which variables affect the provocation of an MMN and cause learning in the participants. This is certainly the case in this study, as the stimuli used contain several aspects that influence our dependent variables.

A way that one can reduce the effect of complexity of the stimuli on the dependent variables is to reduce them to a simple paradigm (Wichmann & Jäckel, 2018, p. 269). Wichmann and Jäckel (2018) illustrate this with a hypothetical paradigm wherein participants must decide if a computer-generated face is male or female. While it is true that a plethora of factors influence this decision, the possible outcomes are reduced to (a) male or (b) female. We have done this in this study, in that it has reduced the composite, complex stimuli, with affecting aspects like prosody, stress, and phonological composition, and reduced it down to a few, simple possible alternatives. Through this approach, this study does include confounding variables like demands on memory, learning, or attention (Wichmann & Jäckel, 2018, p. 269), which are left unaccounted for, but given this study's aim to see the effect of natural linguistic stimuli, the effects of the various different underlying processes of learning are essentially uninteresting; it is the whole act of learning, seen together with the MMN, that is the focus. Thus, the complex underlying process of learning is reduced to a simple dependent variable, and the we were able to vary the stimuli so see how this affected the process.

Another point that Wichmann and Jäckel (2018, p. 269) make, is the so called "know thy stimulus", i.e., is the stimuli valid to measure the potential causal relationship that is being explored. This study is unique in this aspect, given the fact that no study exists with this kind of stimuli. This obviously means that it is not possible to say for certain that the stimuli univocally can be stated to be perfectly valid, and it certainly is not for natural language production, given limited content of the stimuli, but based on the literature that show both that language is (in part) learned by SL (Erickson & Thiessen, 2015), an MMN is elicited when exposed to phonetic stimuli (Partanen et al. 2011), that MMNs are tightly connected to the same neural resources that process music (Koelsch, 2005), and that deficits in linguistic abilities are reflected in deficits in musical abilities (Jentschke et al. 2008), it seems reasonable that a relationship between natural linguistic stimuli and MMN exist. MMN reactions are furthermore proven be elicited during a variety of other activities, like listening to music or sleeping (Beres, 2017), which would indicate that MMNs ought to be elicited while subject to subconscious activities like SL.

This study uses reaction time (RT) and confidence ratings as units of measurement, together with the quantity of correct answers, to determine learning in the participants. It has been chosen to view these aspects in light of the Ratcliff diffusion model, which is the most successful and widely used model for decision-making (Donkin & Brown, 2018). Of course, choosing to adhere to any model brings the risk of not understanding a concept in its nature, but rather in light of our preconception of how it ought to be (Toomela, 2023), but given the models robustness, it seems a reasonable choice. Decision-making is driven based on the quantity of information available to the participants; the more stimuli with more information, that the participants can comprehend, the faster and more accurately a decision can be made.

The participants' confidence in their own accuracy is useful when seen together with the other accompanying aspects of the first familiarity test, RT and accuracy. As stated earlier, increasing confidence, when coupled with increased accuracy and decreased RT, can be seen as a sign of increased sensory data being perceived by the participants, and that they in some way have a conscious relationship with the content of this data. However, what cannot be inferred from confidence ratings is whether or not participants actually have learned above chance level, or if they, notwithstanding the confidence they might have in their answers, actually have the ability to externalise anything accurately from the sensory data. Thus, while confidence can be useful as a supporting piece of evidence to support a theory of learning in the participants, it cannot be used to accurately deem the specificity of this.

This problem is present also in the second, explicit-memory-based familiarity test, i.e., the post-test, where participants were asked to give post-hoc confidence ratings to a list of potential suffixes that they may or may not have heard during the experiment. Whilst this is similar to the test used in Eidsvåg et al. (2015) in that it introduces dummy alternatives to the licit stimuli that were administered during the experiment, it differs in that it is not a yes/no paradigm, but rather a scale. However, it should be said that these confidence ratings can be interpreted along this paradigm, with larger confidence levels entailing a positive response to the prompt. This way, even the second familiarity test can be supported by the evidence of the (a)/(b) paradigm employed in the first familiarity test.

Lastly the problems with the design of the experiment itself should be taken into consideration. Here there are especially three aspects that could affect the validity of the study negatively: the number of participants, wherein the heterogeneity of the participant population is a contributing factor, the quantity of dependent variables within the paradigm, and the human error that has gone into the treatment of the data as part of the analysis process.

The number of participants in this study carries in it both positive and negative aspects. What is most apparent is the fact that with a population *n* = 20, the distribution of their performances is subject to be potentially skewed, given that a gaussian distribution is usually achieved with a higher number of participants. There was also quite a lot of variation between the participants when considering the quantity of languages spoken, as this varied from 2 to 5 per participant. Given the fact polyglotism affects language learning (Papagno and Vallar, 1995) this could have an effect on the results. However, it should be noted that the number of participants was chosen with regards to the fact that it is a proof-of-concept study and the number of participants would allow to prove a potential connection between language learning and MMNs, while at the same time remain within the time and budget restraints the study was subject to. Thus, there has been a trade-off between generalisability and feasibility, but given the aim of the study, this is deemed permissible.

The quantity of dependent factors introduces uncertainty in the validity of the stimuli and the interpretation of the results. Considering the aforementioned maxim “know thy stimulus”, this problem is inherently connected to the fact that this study uses novel kinds of stimuli, and absolute certainty about the affecting factors within them is impossible. This notwithstanding, this study aimed to explore natural linguistic stimuli as a whole as it appears, approximatively, in the real world, and this means that the various variables that make up the stimuli are inherently uninteresting in this context.

Human error affects the study, especially in one step in the analysis process. As mentioned, the data was subject to individual judgement during the refinement process, and whatever elements were included or excluded was decided by the authors. This obviously leads to a certain bias in the data material, but considering the time restraints of the study, it is deemed permissible. One of the authors did preform some guided refinement together with another researcher, to form a training basis for the further data exclusion. In an ideal world, it would have been possible to have both authors evaluate the data and then confront the results, as to achieve a lower degree of bias, but this has not been possible, given the time restraint and the fact that only one computer has been available for analysis purposes.

Lastly, it should be stated that the participant group was linguistically heterogenous, which impacts the results. A downside of this is that we cannot know how the varied linguistic background affected to potential learning in the participants, but an upside is that it gives the study some external validity, as a given population of adults will contain several bi-or-more-lingual individuals.

6.0 Validity

6.1 Internal validity

As mentioned above, one drawback when it comes to using EEG in research is the number of trials that is needed to create valid results in an experiment. A large number of stimuli has to be presented to at least 40 participants to have enough data to draw generalizations from (Beres, 2017). In this current study our group of participants was half the size of this. But if we take into consideration resource and time limitations of this study, the number of participants was deemed adequate.

Another aspect that could affect the internal validity is the use of EEG as a measuring tool. EEGs are known to be sensitive to electrical impulses originating other places the brain (e.g., impulses caused by eye or jaw movements; Hugdahl, 1995). We have been conducting a true experiment, as it takes place in a controlled environment in a laboratory. This has enabled us to secure accurate measurements, control the variables, and given us a fundament on where to compare responses in the process of analyzing the data material (Langdridge, 2006, p. 91). This should also make results reproducible.

6.2 External validity

The external validity of this study is proportionate to its scope. The study is designed to be a proof-of-concept, rather than a definitive explanation of a phenomenon. The generalizability of the study might just be to the population the sample was taken from, if that, and since non-typical neurological conditions and developments have been proven to have an effect of the size and frequency of electrical activity in the brain (Hugdahl, 1995), these populations will have to be examined separately.

Another aspect that affects the external validity is the gender composition of the participant pool. 70% of the participants were women, and this could give a gender biased skew in our data.

7.0 Ethical considerations

In regard to ethical considerations, the first thing we did was to register for the study at RETTE (Risiko og Etterlevelse i forskningsprosjekter) at the University of Bergen (Universitetet i Bergen, 2023). This system is the University in Bergen’s way to create an overview of and control over the processing of personal data in all student and research projects that take place at the institution. Another role of the system is to have an overview of all projects that are related to patient treatment and teaching, and in our case, projects that are related to learning analyses.

In our project we did not handle any personal data that could be used to identify our participants. We were only collecting information regarding age and gender, as well as the information that was asked for in the questionnaire, see appendix C. For instance, one exclusion criterion was that the participants could not be neurodivergent. It has for instance been proven that people with various kinds of cognitive impairments have slower or less intense electrical cerebral reactions than neurotypical peers (Hugdahl, 1995). By excluding them we were hoping to avoid potential error factors. Another reason to exclude potential neurodivergent participants was that the knowledge of this kind of health information would make the management of participant privacy unnecessarily complicated.

All information was treated confidentially and in keeping with current legislation. No personal information was stored digitally, but rather, it was stored on paper, in a locked cabinet in an area with restricted access to the public. Only members of the study handled the information.

Information regarding the experiment, purpose of it and the data collection were sent to all interested in the project before any further arrangements were made. The participants were presented with the same information upon arriving at the laboratory and were asked to sign the document before we proceeded with the experiment. The document can be seen in Appendix B. When each participant had completed the data collection they were debriefed. They also received a printed copy of the debriefing, where they were told what they had participated in.

All participants were compensated with a universal gift card worth 200NOK. Considering that each participant spent approximately two hours in the laboratory, this is a compensation of 100NOK per hour. Given the list of exclusion criterions, the participation in the study were not open for everyone. The participants were not asked for information regarding sensitive topics, and all questions they were to answer were related to the stimuli they had heard. Therefore, the compensation should not be seen as an incentive to recruit participants, but rather as a compensation for their time.

7.1 Consent

The participants were informed of their right to withdraw their consent at any time, even after the data has been collected. This means that we, in our role as researchers, are obliged to delete all data we have collected regarding them, should they demand this to happen (Langdridge, 2006, p. 350). All participants received this information as they entered the study. A fundamental principle in research is that one should treat the participants with respect (Langdridge, 2006, p. 345). We consider information to be a vital part of this principle, and we have strived to provide the participants with the information they need, in order to give us their informed consent on the right premises.

The participants were asked for permission for their anonymised data to be stored in a third-party location, when the data-collection was completed. All participants gave their consent. In accordance with the principle of open and public research, anonymous data may be rendered accessible to the public through a third-party storage location (e.g., Open Science Foundation). Public data will not be stored in a way that individual participants can be traced. See Appendix B for further information.

7.2 Our role as researchers

As researchers, it is important to keep in mind that the aim for the research is to investigate whether a hypothesis is correct, not to confirm a truth one might believe to exist. This can in some cases prove hard to attain, as one may be invested in the project in regard to time, engagement and money. It might be useful to reflect on this prior to conducting the study.

Another important aspect of research is to limit the researchers’ possibilities to influence the participants. Attempts to influence them can happen both deliberately and unconsciously. A good framing of an experiment conducted in a laboratory, can limit the risk of this happening (Warne, 2018). In this study the use of the EEG machine is removing us, as researchers, from the experiment. We do still have to be careful in the process of preparing the stimulus for the experiment, and when analysing, processing and presenting the data subsequently.

With this discussion in mind, we conducted this study and found that our hypotheses were largely confirmed, but the results varied in unexpected ways. We did find that MMNs were elicited when participants were exposed to linguistic stimuli, with distribution patterns that are coherent with previous research. We also found that the participants did implicitly learn the stimulus material to an above-chance level. The second familiarity test however, showed that the explicit learning of the stimuli were less certain, and that the participants generally were more sure of the presence of syllables ending in consonants, regardless of whether or not these endings had been present in the actual material.

8.0 Literature

Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian Analysis of Confidence in Perceptual Decision-Making. *PLoS Computational Biology, 11*(10).https://doi.org/10.1371/journal.pcbi.1004519

Amazon (n.d). *Amazon Polly.* [Text-to-speech generator]. https://aws.amazon.com/polly/.

Arciuli, J. (2018). Reading as Statistical Learning. *Language, Speech & Hearing Services in Schools, 49*(3S), 634–643. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0135

Arciuli, J., & Conway, C. M. (2018). The Promise—and Challenge—of Statistical Learning for Elucidating Atypical Language Development. *Current Directions in Psychological Science*, *27*(6), 492-500. https://doi.org/10.1177/0963721418779977

Beech, C., & Swingley, D. (2023). Consequences of phonological variation for algorithmic word segmentation. *Cognition, 235*, 105401-105401. https://doi.org/10.1016/j.cognition.2023.105401

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology and Biofeedback, 42*(4), 247-255. https://doi.org/10.1007/s10484-017-9371-3

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences - PNAS, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *The Journal of Neuroscience, 35(8)*, 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Brain Products GmbH. (2019). *BrainVision Recorder.* BrainVision Recorder. https://www.brainproducts.com/solutions/recorder

Cardoso, W. (2011). The development of coda Perception in Second, language phonology: A variationist perspective. *Second Language Research, 27(4)*, 433–465. https://doi.org/10.1177/0267658311413540

Celesia, G. G. & Peachey, N., S. (2004). Visual Evoked Potentials and Electroretinograms. | Niedermeyer, E., & Da Silva. F. L. *Electroencephalography: Basic principles, clinical applications, and related fields*. (pp. 1007-1044). Wolters Kluwer.

Cukor-Avila, P. (2000). REVISITING THE OBSERVER'S PARADOX. *American Speech, 75(3*), 253–254. https://doi.org/10.1215/00031283-75-3-253

Da Silva, F. L. (2004). Event-related Potentials: Methodology and Quantification. | Niedermeyer, E., & Da Silva. F. L. *Electroencephalography: Basic principles, clinical applications, and related fields*. (pp. 991-1003). Wolters Kluwer.

Delorme A & Makeig S (2004) EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods 134*:9-21

Donkin, C., Brown, S. D. (2018) Response times and decision making | Wixted, J. T, Wagenmakers, E. J., (Ed.) *Stevens' handbook of experimental psychology and cognitive neuroscience: Volume 5: Methodology* (pp.349-382). (Fourth edition.). Wiley

EEGLAB (n.d) *Background on Independent Component Analysis applied to EEG*. EEGLAB. Retrieved March 10, 2023, from https://eeglab.org/tutorials/ConceptsGuide/ICA_background.html

Eidsvag, S. S., Austad, M., Plante, E., & Asbjornsen, A. E. (2015). Input Variability Facilitates Unguided Subcategory Learning in Adults. *Journal of Speech, Language, and Hearing Research*, *58*(3), 826-839. https://doi.org/10.1044/2015_JSLHR-L-14-0172

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental review*, *37*(Sep), 66-108. https://doi.org/10.1016/j.dr.2015.05.002

Evans, L. H., & Wilding, E. L. (2012). Recollection and Familiarity Make Independent Contributions to Memory Judgments. *The Journal of Neuroscience, 32(21)*, 7253–7257. https://doi.org/10.1523/JNEUROSCI.6396-11.2012

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6(2)*, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *NeuroImage (Orlando, Fla.), 54*(3), 2267–2277. https://doi.org/10.1016/j.neuroimage.2010.10.038

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology*, *2*, 208-208. https://doi.org/10.3389/fpsyg.2011.00208

Hamada, M., Goya, H. (2015). Influence of Syllable Structure on L2 Auditory Word Learning. *Journal Psycholinguistic Research 44*, 141–157 . https://doi.org/10.1007/s10936-014-9284-8

Han, M., De Jong, N. H., & Kager, R. (2022). Prosodic input and children's word learning in infant- and adult-directed speech. *Infant behaviour & development*, *68*, 101728. https://doi.org/10.1016/j.infbeh.2022.101728

Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology, 5*, 175–175. https://doi.org/10.3389/fpsyg.2014.00175

Hugdahl, K. (1995). *Psychophysiology : the mind-body perspective*. Harvard University Press.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks, 13*(4), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with Specific Language Impairment Also Show Impairment of Music-syntactic Processing. *Journal of Cognitive Neuroscience, 20*(11), 1940-1951. https://doi.org/10.1162/jocn.2008.20135

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*(1), 131-141. https://doi.org/10.1111/j.1467-7687.2008.00740.x

Keren-Portnoy, T., Vihman, M., & Fisher, R. L. (2019). Do Infants Learn from Isolated Words? An Ecological Study. *Language learning and development*, *15*(1), 47-63. https://doi.org/10.1080/15475441.2018.1503542

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *Journal of Cognitive Neuroscience, 17(10)*, 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179-190. https://doi.org/10.1111/j.1469-8986.2008.00752.x

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports, 6*(1), 19741–19741. https://doi.org/10.1038/srep19741

Krohl, L. R (2020). *EEG 10-10 system with additional information.* Retrieved from: https://commons.wikimedia.org/wiki/File:EEG_10-10_system_with_additional_information.svg on 26.02.2023

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuronen, M., & Tergujeff, E. (2020). Second language prosody and its development: connection between different aspects. *Language learning journal*, *48*(6), 685-699. https://doi.org/10.1080/09571736.2018.1434228

Langdridge, D. (2006). *Psykologisk forskningsmetode : en innføring i kvalitative og kvantitative tilnærminger*. Tapir.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121. https://doi.org/10.1016/S0010-0277(00)00109-8

The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Näätänen, R. (1995). The Mismatch Negativity: A Powerful Tool for Cognitive Neuroscience. *Ear and Hearing*, *16*(1), 6-18. https://doi.org/10.1097/00003446-199502000-00002

Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *International Journal of Psychophysiology*, *48*(2), 179–188. https://doi.org/10.1016/S0167-8760(03)00053-9.

Niedermeyer, E., & Da Silva. F. L. (2004) *Electroencephalography: Basic principles, clinical applications, and related fields*. Wolters Kluwer

Nyman, G., Alho, K., Laurinen, P., Paavilainen, P., Radil, T., Reinikainen, K., Sams, M., & Näätänen, R. (1990). Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. *Electroencephalography and Clinical Neurophysiology*, *77*(6), 436–444. https://doi.org/10.1016/0168-5597(90)90004-W

Papagno, C., & Vallar, G. (1995). Verbal Short-term Memory and Vocabulary Learning in Polyglots. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *48*(1), 98–107. https://doi.org/10.1080/14640749508401378

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical expertise induces audiovisual integration of abstract congruency rules. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(50), 18196–18203. https://doi.org/10.1523/JNEUROSCI.1947-12.2012

Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. (2011). Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles: Linguistic multifeature MMN paradigm. *Psychophysiology*, *48*(10), 1372-1380. https://doi.org/10.1111/j.1469-8986.2011.01214.x

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of memory and language*, *39*(2), 246-263. https://doi.org/10.1006/jmla.1998.2576

Plante, E., & Gómez, R. L. (2018). Learning Without Trying: The Clinical Relevance of Statistical Learning. *Language, Speech & Hearing Services in Schools*, *49*(3S), 710-722. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjørnsen, A. E. (2017). An fMRI study of implicit language learning in developmental language impairment. *Neuroimage Clinical*, *14*(C), 277-285. https://doi.org/10.1016/j.nicl.2017.01.027

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second Language Experience Facilitates Statistical Learning of Novel Linguistic Materials. *Cognitive science*, *41 Supplement 4*(S4), 913–927. https://doi.org/10.1111/cogs.12473

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59-108.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700-703. https://doi.org/10.1038/nn873

Saksida, A., Flo, A., Guedes, B., Nespor, M., & Peña Garay, M. (2021). Prosody facilitates learning the word order in a new language. *Cognition*, *213*, 104686-104686. https://doi.org/10.1016/j.cognition.2021.104686

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975-983. https://doi.org/10.1016/j.cognition.2007.03.005

Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11*(3), 407-413. https://doi.org/10.1111/j.1467-7687.2008.00685.x

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1-32. https://doi.org/10.1016/j.cogpsych.2006.04.002

Steinbeis, N., & Koelsch, S. (2008). Comparing the Processing of Music and Language Meaning Using EEG and fMRI Provides Evidence for Similar and Distinct Neural Representations. *PloS One, 3*(5), e2226–e2226. https://doi.org/10.1371/journal.pone.0002226

Tal, S. & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition*, *224*, 105055–105055. https://doi.org/10.1016/j.cognition.2022.105055

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21-21. https://doi.org/10.1186/1471-2202-10-21

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional Specialization of the Human Auditory Cortex in Processing Phonetic and Musical Sounds: A Magnetoencephalographic (MEG) Study. *Neuroimage*, *9*(3), 330-336. https://doi.org/10.1006/nimg.1999.0405

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The Extraction and Integration Framework: A Two-Process Account of Statistical Learning. *Psychological Bulletin*, *139*(4), 792-814. https://doi.org/10.1037/a0030801

Thiessen, E. D., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology*, *39*(4), 706-716. https://doi.org/10.1037/0012-1649.39.4.706

Thomassen, M. (2018). *Vitenskap, kunnskap og praksis : innføring i vitenskapsfilosofi for helse- og sosialfag*. (9. Utg.) Gyldendal akademisk.

Toomela, A. (2023). Methodology of Science: Different Kinds of Questions Require Different Methods | Gozli, D., & Valsiner, J. (Edt). *Experimental Psychology : Ambitions and Possibilities*. (pp. 113-151). Springer International Publishing AG.

Tsogli, V., Jentschke, S., Daikoku, T., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports*, *9*(1), 5563–5563. https://doi.org/10.1038/s41598-019-42066-4

Universitetet i Bergen. (2023). *RETTE - UiBs prosjektoversikt*. uib.no. Retrieved 20.04 from https://www.uib.no/personvern/128207/rette-uibs-prosjektoversikt

Valsecchi, M., & Turatto, M. (2023). Habituation to abrupt-onset distractors with different spatial occurrence probability. *Attention, Perception, and Psychophysics*, *85*(3), 649-666. https://doi.org/10.3758/s13414-022-02531-1

Warne, R. T. (2018). *Statistics for the social sciences : a general linear model approach*. Cambridge University Press.

Wichmann, F. A., Jäckel, F. (2018) Methods in Psychophysics | Wixted, J. T, Wagenmakers, E. J., (Ed.) *Stevens' handbook of experimental psychology and cognitive neuroscience: Volume 5: Methodology* (pp.265-307). (Fourth edition.). Wiley

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742*(1-2), 239-252. https://doi.org/10.1016/S0006-8993(96)01008-6

Exploring the mismatch negativity of statistical language learning.

Olav Tidemann Garli and Aina Cecilie Moe Klinge
Department of Biological and Medical Psychology at The University of Bergen

Authors note

This article is part of the master's thesis in Speech and Language Therapy
at the University of Bergen.

Abstract

Statistical learning (SL) is a core learning mechanism in language acquisition. It has previously been explored in both behavioural as well as in neurophysiological experiments. One neurophysiological marker of rule acquisition using statistical learning is denoted as the statistical mismatch negativity (sMMN). Earlier neurophysiological studies exploring the sMMN used sound timbres as stimuli, the current experiment aimed at exploring whether this brain response could also be elicited using naturalistic language stimuli from Russian. Twenty participants (unfamiliar with Russian or other Slavic languages) were exposed to a stream of Russian words, where six possible word roots were combined with two possible endings. The transition probabilities between the word roots and these two suffixes varied, with one ending occurring more often (90%; statistically regular) than the other ending (10%; statistically irregular). The difference between the brain responses to those represent the sMMN. In addition, sound direction was manipulated to elicit a physical MMN (phMMN). The neurophysiological data revealed both a physical and a statistical MMN, with maximum amplitudes around 150 ms and 240 ms respectively. Implicit learning was documented both neurophysiologically from the first block of the experiment (i.e., around 10 mins stimulus exposure) and behaviourally as above chance level recognition from the third block. At the end of the experiment, explicit learning was assessed ,however without consistent results: these were ambiguous, and the participants could not discriminate real endings from decoys.

Key words: Language acquisition, Statistical learning, EEG, ERP, Mismatch Negativity, sMMN, phMMN.

Introduction

Statistical Learning

Statistical learning (SL) is a core learning mechanism in language acquisition. One neurophysiological marker of rule acquisition using statistical learning is denoted as statistical mismatch negativity (sMMN). In this current experiment study electroencephalography (EEG) was used to explore whether the sMMN response could be elicited using naturalistic language stimuli from Russian, and whether the participants showed evidence of learning.

Language is a highly complex cognitive skill that is a cornerstone of the human condition. Compared to communication systems in other species, human language possesses the properties of productivity, enabling us to create an infinite number of utterances from a limited number of basic elements (phonemes), as well as displacement, denoting the ability to refer to objects, events, and ideas that are not immediately present. Given the complexity of language, it is puzzling how comparably effortless most children acquire it. This acquisition process would not have been possible without the help of powerful learning mechanisms. Statistical learning (SL) is one such mechanism. It permits to extract statistical regularities from the environment, e.g., from language (Koelsch et al. 2016). Language is characterized by a natural flow of speech that contains no constant cues to where a word begins and where it ends (Johnson & Seidl, 2009). It is hard to separate individual words out of running speech when exposed to unfamiliar languages. However, there exist certain regularities, specific to each language, guiding which phonemes can be combined to form syllables and how these syllables can be further combined into words. SL permits to extract these regularities and thereby enables the segmentation of the stream of continuous language sounds into meaningful units (words, phrases, and sentences). In fact, research suggests that infants, at only 8 months of age, can accomplish word segmentation based on SL, after being exposed to just two minutes of speech (Saffran et al., 1996). This may be an indication that the mechanisms for the computation of the statistical properties are powerful in this age-group in relation to identifying structures in languages (Saffran et al., 1996). On a higher level, words and grammatical structures are also combined according to such regularities, which can be extracted using the same mechanism of SL (Plante et al., 2014).

During language acquisition, learners are sensitive to two aspects of the statistical structure, namely conditional statistical information and distributional statistical information. They again include two complementary processes, extraction and integration (Perruchet & Vinter, 1998; Thiessen et al., 2013). The term extraction refers to processes in working memory that enable us to hold two elements in our mind at once, and to combine them at the same time. Integration, on the other hand, refers to the process of combining information, and identifying central tendencies within the information (Perruchet & Vinter, 1998; Thiessen et al., 2013). It is not quite clear how closely these two processes are connected, but the importance of their mutual relationship is undoubtedly significant (Erickson & Thiessen, 2015). Learning of the reoccurring regularities within language is based on its statistical structures, such as frequency, variability, distribution and the probability of a co-occurrence, of its elements. The statistical structures underlying the arrangement of these elements are acquired using SL (Erickson & Thiessen, 2015). The mechanism involved in picking up on these aspects can be explained by our sensitivity to regularities, and our ability to pick up statistical structures from the world around us (Koelsch et al., 2016).

The neural correlates of SL have been explored using either electroencephalogrphy (EEG) and event-related potentials (ERPs) (further discussed below) – focusing on the time course of the cognitive processes involved in the processing of regularities acquired by SL – or functional Magnetic Resonance Imaging (fMRI) – focusing on the brain regions supporting those cognitive processes. Such cognitive processes include keeping track of the sequential order and memory encoding. Findings from fMRI studies indicate sinistrad lateralisation in relation to acquisition through repeated exposure and afterwards processing the regularities within natural language stimuli (Plante et al., 2015; Plante et al., 2014; Plante et al., 2015; Plante et al., 2017).

Linguistic redundancy in SL

While SL is a powerful learning mechanism, it might not be enough to account for all aspects of language acquisition. Additional information, e.g., prosody, may therefore be necessary to support rule extraction via SL (Shukla et al., 2007). Such information might be delivered in stress, pauses, and intonation. According to Kuhl (2004) languages are mostly dominated by trochaic words, where the first syllable is stressed, or iambic words, where the second syllable are stressed. The distinction between trochaic words and iambic words can be used to identify word boundaries. This was supported by Johnson and Seidl (2009) who found evidence that 11 months-olds tend to weight stress cues in relation to word boundaries more heavily than statistical cues. Prosody might act as a filter that suppresses possible word-like sequences that outspan prosodic constituents in a stream of speech (Shukla et al., 2007).

Another important factor regarding language acquisition is the phono-tactical structure in the different languages, and how this relates to the phono-tactics of a speaker's L1. Cardoso

(2011) researched the role of coda in relation to second language acquisition. A definition of coda is that it is made up of the consonants at the end of a syllable. The results showed that there is a certain degree of correlation between speech perception and production. This indicates that perception precedes the production of speech (Cardoso, 2011). These results may be interpreted as the participants struggle to implement an established language structure from their first language to the second language they are exposed to. A similar study was made by Hamada and Goya (2015). They were looking at two groups of college students, a Japanese group and an English group. They hypothesized that the Japanese participants would learn English words with an open-syllable structure without consonant clusters better than words with these qualities. An open-syllable structure means that the syllable ends with a vowel. Their hypothesis was disconfirmed, as they found that the recall accuracy was higher for the words containing consonant clusters and coda (Hamada & Goya, 2015). Cutler (1997) argues that it is language dependent which syllable structure an individual will find the easiest to understand. Since this study explores the neurophysiological aspects of language acquisition, EEG and ERPs will be described in in further detail below.

EEG and ERPs

EEG is a technique to measure electrical activity in the nerve cells of the brain, mainly in the cortical areas, using electrodes mounted on the scalp (Hugdahl, 1995, pp. 234-235). The electrodes pick up cyclical changes in the membrane potentials of underlying nerve cells, which generates a rhythmic pattern in an EEG wave. The EEG patterns depict the fluctuations in voltage in the neurological structures that they measure. Based on the amplitude and frequency of these fluctuations, one may analyze their behaviour (Hugdahl, 1995, pp. 234-235). There are about 86 billion neurons in the average human brain, and the communication between these are what is measured using EEG. The neurons have intrinsic electrical properties, and these properties, from a population of neurons called the pyramid cells, are picked up by the EEG electrodes (Beres, 2017). The electrodes attached to the scalp will then record a negative extracellular potential if this is happening to a large number of pyramidal cells in the same area at the same time (Beres, 2017).

From the recording of an EEG, it is possible to detect an ERP. The ERP signal is related to the occurrence of an external or internal stimulus. The signal is an “answer” from the brain, and it takes place within a certain timeframe following the stimulus (Hugdahl, 1995, pp. 266-270). Language research using ERPs has proven useful, due to their good temporal resolution. Linguistic variations are often small and fast, and therefore, the ERPs are useful in picking up on these small variations, and the neurological changes that occur (Beres, 2017). This study focuses on a particular kind of ERP, the Mismatch Negativity (MMN).

The MMN

As the name suggests, the MNN represents a negative waveform. According to Näätänen (1995) the MMN is a result of a deviation in the memory trace caused by a sensory input. It has a relatively large amplitude above temporal areas, and it has been suggested that it is taking place in the additory primary and association areas. The MMN is a reflection of specific auditory discrimination of stimulus (Näätänen, 1995). MMN responses are normally recorded when a deviant stimulus is given as an auditory input with a change in either intensity or frequency (Nyman et al., 1990). Näätänen (1995) lists several reasons why the MMN response is a tool that can be used for auditory research. First, it provides an objective measure of an individual's ability to discriminate different sound features. Second, attention is not needed to elicit an MMN response. For the MMN to be elicited there is no need for the individual to participate in any specific activities. It has been observed when an individual is reading a book, listening to music, watching a movie, or sleeping (Beres, 2017). Thirdly, it involves auditory short-term sensory memory. And finally, as auditory short-term memory is crucial for correct speech processing, the MMN provides a means for studying this phenomenon (Näätänen, 1995). The MMN acts at a pre-attentional level demonstrated by that a MMN could be elicited in coma patients (Näätänen, 2003). The MMN as described so far is elicited by changes in physical properties, a physical MMN (phMMN). The phMMN is dependent on operations in the auditory sensory memory and sensory-memory-representations that are updated instantly. The phMMN will, for example, be elicited in connection to a deviance in sound location (Tsogli et al., 2019). For language, a phMMN was demonstrated in Partanen et al. (2011) who used naturally produced speech stimuli to elicit a MMN response. Their stimuli consisted of a pseudoword, for example [tatata], and the changes in the given stimuli were of an acoustic character. They found that all changes to the stimuli elicited a MMN response, but by changing the duration of the vowel this elicited a different response to the other. This showed that it is possible to assess speech sound discrimination in only 30 minutes, using a multi-feature paradigm (Partanen et al., 2011). The experiment was conducted by the use of EEG, and this technique will be described further detail below. It can also be elicited by changes in more complex acoustic features, such as abstract-feature MMN described by Saarinen et al. (1992), and even by a violation of expectations regarding the arrangement of elements in sound sequences that follow certain rules. Koelsch et al. (2016)

found that violations of such statistical regularities were established through learning those regularities over time eliciting a so-called statistical MMN (sMMN) (Koelsch, 2009). The sMMN is dependent on memory representations formed in learning processes. These lead to representations in the memory that go beyond the capacities of the sensory memory. For instance, a sMMN will be elicited if one experience deviations from learned transition probabilities.

The fact that different types of expectancy violations can elicit MMN-type-responses, corresponds with the suggestion that the MMN reflects the outcome of a neural mismatch process and takes place between the deviant stimulus and a memory trace of the standard stimuli (Winkler et al., 1996).

Background studies

A study by Koelsch et al. (2016) involved a novel variant of the SL paradigm. Koelsch et al. (2016) presented timbres of triplets in isochronous sequences to a group of 18 adult participants. The two first sounds of the triplets had the same amount of probability, but the third sound occurred with a probability of either low (10%), intermediate (30%) or a high probability (60%). They found that endings with low and intermediate probability compared to the high probability endings elicited an early anterior negativity (with an onset around 100 ms, and a maximum around 180 ms), which was denoted as sMMN. The amplitude difference (with the high probability items) was higher for the low than for the intermediate probability items. The reported sMMN can therefore be taken to reflect the acquisition of the distribution of transition probabilities through SL. Notably, the acquisition process exceeds capabilities of auditory sensory memory (Koelsch et al., 2016). It, thus, likely reflects that long-term memory representations of the regularities underlying the different transition probabilities were established using SL

Tsogli et al. (2019) extended that paradigm, comparing the phMMN and the sMMN to explore to what degree the underlying cognitive processes influence and interact with each other. Like Koelsch et al. (2016), they were using EEG to record brain responses to deviations in an auditory stream of sound triplets. 21 participants were exposed to a continuous stream of auditory sound triplets where deviations were either statistical (low compared to high transition probability between sequence root and endings), physical (a change in the sound location), or a combination of the two deviations. Tsogli et al. (2019) observed an interaction effect when both deviations were combined, leading to a smaller amplitude size of the sMMN when it co-occurred with a physical deviation. This means that the processing of prediction errors related to statistical learning is affected by prediction errors related to physical deviance (Tsogli et al., 2019).

Using a behavioural paradigm, Eidsvåg et al. (2015) investigated whether variability in the linguistic input would influence language learning. Forty adults, divided into two groups, were familiarized with noun gender subcategories in Russian. One group were presented with a high-variability condition in which they were familiarized with 32 different root-words. The other group were presented with a high-repetition condition, they were familiarized with 16 root-words presented twice to provide the same amount of exposure as in the first group. The results showed that only participants in the high-variability group experienced learning after the initial familiarization, while participants in the high-repetition condition needed additional exposure in order to learn the endings. This is a demonstration that learners' ability to generalize language input can be influenced by the degree of input variability (Eidsvåg et al., 2015).

The aim of the current study was to combine the experimental paradigms of Koelsch et al. (2016) and Tsogli et al. (2019), on one hand, and of Eidsvåg et al. (2015) on the other hand. The experiments by Koelsch et al. (2016) and Tsogli et al. (2019) employed EEG measurements to trace the rule acquisition process but used timbre stimuli (i.e., non-linguistic). Eidsvåg et al. (2015) used behavioural measurements but with linguistic stimuli. Thus, the current experiment used the language stimuli of Eidsvåg et al. (2015) while tracing the rule-acquisition process using an EEG experiment.

Hypotheses

We expect that linguistic stimuli will elicit similar MMN responses as the (non-linguistic) sound/timbre stimuli that were used in the experiments of Koelsch et al. (2016) and Tsogli et al. (2019). More specifically, we expected that physical deviance would elicit a phMMN and that the rule-acquisition process of acquiring the regularities in the transition probabilities between the word roots and endings would establish the elicitation of a sMMN as in those earlier experiments.

We also expect that both neurophysiological measures of rule acquisition as well as a familiarity test (based upon implicit memory) will be more sensitive to pick up the outcome of the statistical language-based acquisition process than the explicit memory-based measures used by Eidsvåg et al. (2015).

Lastly, we expect that participants will have a better performance in the familiarity test at the end of each block as compared to Tsogli et al. (2019): The experimental paradigm and the stimuli used by Tsogli et al. (2019) required to segment the sound stream while acquiring the underlying rules of arrangement. For the current experiment, the character of the stimuli – words that were separated from one another by a brief pause and that each had a prosodic structure that indicated begin and end of the word – made the segmentation process unnecessary and therefore the rule-acquisition easier.

Method

Participants

We recruited 14 women and six men as voluntary participants in the study, aged 20-50 (mean age, 26.75, SD = 8.47). All participants provided written, informed consent. All participants were right-handed, had self-reported normal hearing, had no musical education beyond mandatory school lessons. 12 of the participants had Norwegian as their mother tongue, while the remaining eight spoke Tigrinya, German, Cantonese, Chinese, Turkish, Urdu, Romanian or Arabic. All participants spoke English with a high degree of fluency, and some of the participants with a different mother tongue than Norwegian also spoke Norwegian well.

Materials

The stimulus material consisted of a selection of six Russian word roots, which would be combined with two possible endings. The word roots were: <Дви́га> - [ˈdvʲigə], <Храни> - [xrɐˈnʲi], <Мечта> - [mʲɪt͡ɕˈta], <Слуши> - [ˈsłuʂə], <Спаси> - [spɐˈsʲi], and <Води> - [vɐˈdʲi]. These word roots were chosen due to their phonological content and shape. All word roots were two syllables in length and differed as much as possible in their phonatory profile from each other. One step in this was choosing words that began with different consonants, or, in the case of <Слуши> - [ˈsłuʂə] and <Спаси> - [spɐˈsʲi], syllable onsets that varied markedly in voicing. The endings that were added to the word roots were теля – telya - [tʲɪlʲə], and телем – telyem – [tʲɪlʲəm]. This resulted in twelve different test items.

Procedure

Participants were recruited using posters, providing the authors e-mail addresses. Once a participant contacted one of the authors, they were given an information sheet to introduce the experiment and how it would be conducted. The information sheet was, however, formulated in a way to avoid that the participants become aware that the aim of the experiment was to assess SL of the rules underlying the arrangement of word roots and endings (i.e., the acquisition process was still based upon implicit memory)

The experiments took place at the EEG-lab at the Department of Psychosocial Science, Faculty of Psychology at the University of Bergen. Once the participants arrived, they were asked to read the information sheet again, ask questions if aspects of the description were unclear to them, and then sign the declaration of consent.

The experiment consisted of three parts. First, the provided information regarding their language background and linguistical knowledge, as well as regarding their musical education (including whether they had played any instruments or not) on a paper questionnaire.

After the first part was completed, the participants were led into an electrically shielded chamber, designed to conduct EEG-experiments. The participants were asked to sit in front of a computer monitor. Once seated the EEG-cap was laced onto their head, with the electrodes already attached to the correct positions in the cap. Afterwards, the electrodes were filled with electrode gel to minimize resistance between the electrodes and the surface of the scalp (the maximum impedances were kept below 5 kΩ). Participants were asked to adjust their seating position, to ensure their comfort in order to reduce possible muscle artifacts in the recording. We furthermore explained the basics of EEG measurement to the participants, including e.g., what impact eye movements had on the EEG recording (combined with a request to minimise such movements).

After all preparations for the recording were finished, the light in the EEG-room was dimmed to a pre-defined level, (quite dark but with enough light to still see the keys on the keyboard). Then EEG recording was started, and the experiment began with instructions that were presented to the participants on the monitor in front of them.

During the experiment, participants could watch a nature documentary, without text or sound (to make the experiment more enjoyable for the participants). To ensure that they were following the stimulus presentation attentively, a cover task was devised. The cover task required the participants to detect a change from the female standard voice the stimuli were spoken with, to an occasionally occurring stimuli spoken by a male (such change occurred approximately once per minute). When such a change happened, the participants had to react by pressing the space bar on a provided keyboard.

The experiment was split into six blocks of 10 minutes each. Within each block, approximately 600 words (each consisting of a root and an ending) were presented. Each word was lasted approximately 800 ms, and whether a word contained a physical or a statistical deviant or were standard stimuli was randomised.

After a block with stimulus presentation was finished, the participants were given a implicit-memory-based familiarity test (similar to the one used in Tsogli et al., 2019). They heard each word root with the two different possible endings, and they had to decide which alternative sounded more familiar by pressing either “1” or “2” on the keyboard. The order of these presentations was randomised, and each ending (standard and statistical deviant) appeared once within the first and once within the second sequence (i.e., 12 sequences in total). After the decision about which sound sequence was more familiar, they had to rate how confident they were in their answers on a scale from “1” to “5”.

Once the participants had finished the familiarity test of the last block, the light was returned to normal levels, the EEG cap was removed from the participants’ head, and the participants left the shielded chamber. They were then asked to fill in a debriefing questionnaire and an explicit-memory-based post-test was administered. In the debriefing, the participants were asked if they thought any pattern was present, and to judge their own performance in searching for patterns in the stimulus material. The post-test presented them with a list of six endings, two of these were used during the experiment, and an additional four were decoy endings. For each ending, the participants rated, for how confident they were in having heard that specific ending on a scale from 0-100. After finishing the post-test, the participants received a universal gift card worth 200 NOK (≈ 20$ US).

EEG data recording

The EEG-read outs were recorded from 66 electrodes (illustrated in Figure 1) with a sampling rate of 500 Hz and without online filtering using Brain Vision Recorder version 1.25 (Brain Products GmbH, 2019). 63 electrodes were placed on the scalp in accordance with the 10-10 system. Additional electrodes were placed on the right cheekbone (for vertical eye movements) and two at the outer canthi of the participants’ eyes (for horizontal eye movements). The electrodes positioned at TP9 served as reference electrode, and the ground electrode, was positioned at Iz.

[Insert Figure 1 here]

The EEG data were analysed using EEGLAB version 2022.1 (Delorme & Makeig, 2004) in MATLAB 2022b (The Mathworks, 2022). First, the data were checked and manually rejected. If there were any longer periods with faulty data in a single electrode (approx. 5% of the recording; 180 seconds), the electrode was removed from the recording (and later on interpolated using the adjacent electrodes).[1] Periods with artifacts affecting several channels (e.g., electrode drifts or continuous artifacts caused by muscular activity) were rejected. After manual rejection, EEGLAB's ICA r*unica* was used to exclude independent components (IC) that represented artifacts, such as eye movements (blinks or horizontal eye movements) and carotid-artery activity (EKG). Data containing artifacts were rejected by removing sampling points whenever the standard deviation within a 200 or 800 ms gliding window exceeded 25 µV at any electrode channel (including the EOG channels. Afterwards, epochs (with the onset at the word endings) were generated, and epochs that occurred within 3 seconds after a voice change we discarded (rejecting activity related to the cover task). The remaining epochs were averaged using four conditions: physical standards, physical deviants, statistical standards, and statistical deviants. From those, both the physical MMN (physical deviants minus physical standards) and the sMMNs (statistical deviants minus statistical standards) was calculated. The physical MMN started at approximately 100 ms with an onset of 0 ms, while the sMMN began with an onset of 164 ms, (the first part both of both the low and high transition probability endings was the same (-tel-), and only at the onset of at *–ja* or /*-jem*, it became possible to distinguish whether the ending was regular or irregular)

Results

Physical MMN

A phMMN could be observed in a time window between 100 and 200 ms with a maximum amplitude of approximately 2 µV, and an amplitude peak around 150 ms (see Figure 2)

[Insert Figure 2 here]

The phMMN was evaluated using an ANOVA with the factors condition (with vs. without location change), anterior-to-posterior (frontal, central, parietal), and left-to-right (left, central, right). It revealed a significant effect of physical deviance condition $F(_{1,20})$ = 47.01, p <.001, $\eta^2 = .25$, as well as interactions of condition with anterior-to-posterior, $F(_{1,20})$ = 4.99, $p = .01$, $\eta^2 = .02$, of condition with left-to-right, $F(_{1,20})$ = 12.88, p = <.001, $\eta^2 = .02$, and of condition with anterior-to-posterior and left-to-right, $F(_{1,20})$ = 4.15, $p = .004$, $\eta^2 = .002$.

[Insert Figure 3 here]

As illustrated in Figure 3, the main effect of condition reflects that there was an amplitude difference between regular and irregular sound direction that generally was present over the whole scalp. The interactions with condition reflect that though the phMMN could be observed at the whole scalp, it had an amplitude maximum at fronto-central scalp sites and along the midline, however, with a slight lateralisation to the left hemisphere.

In addition, the ANOVA revealed further significant effects not involving condition, a main effect of left-to-right, $F(_{2,38}) = 24.15$, $p < .001$, $\eta^2 = .06$, and the interaction of anterior-to-posterior and left-to-right, $F(_{2,38}) = 13.77$, $p < .001$, $\eta^2 = .01$, reflecting that also the sounds from the standard direction elicited a brain response that was (although on average having a positive amplitude) lowest over central scalp areas compared to both the left and the right hemisphere (with this pattern more pronounced over frontal and central than parietal sites).

Statistical MMN

A sMMN was elicited as difference between the irregular (i.e., less often – 10%) and the regular (i.e., more often – 90%) word endings. Given that the first part of the ending (-tel) was identical for both endings (-telya or –telyem), the onset of the brain response had to be adjusted for the length of -tel (164 ms): the used time window between 300 to 500 ms is therefore in fact between 136 to 336 ms (relative to the end of -tel). Within that time window, the brain response to the irregular stimuli has a more negative amplitude than the brain response to the regular stimuli, resulting in a sMMN. The sMMN (see Figure 4) is most prominent over fronto-central scalp sites, and generally largest over midline electrodes (but with a slight lateralisation to the left) and has an amplitude maximum of approximately 1 µV peaking at around 240 ms (404 ms relative to the onset of –tel). That is, compared to the phMMN, it had a similar scalp distribution, but a smaller amplitude size (2 vs. 1 µV) and a higher latency (150 vs. 240 ms).

[Insert Figure 4 here]

The sMMN was statistically analysed with an ANOVA for a time window of 136 to 336 ms relative to when the two endings became distinguishable. The ANOVA had the factors condition (regular = high transition probability vs. irregular = low transition probability), anterior-to-posterior (anterior, central and posterior) and left-to-right (left, midline and right). It revealed a significant effect of condition, $F(_{1,19}) = 14.40$, $p < .001$, $\eta^2 = .11$, and an interaction of condition and left-to-right, $F(_{2,38}) = 9.16$, $p = <.001$, $\eta^2 = .01$. This reflects that the amplitude difference in the brain responses to the irregular vs. the regular endings could be observed over the whole scalp with the largest differences to be found around the midline (see Figure 5).

[Insert Figure 5 here]

Behavioural results

In addition to assessing the neurophysiological correlates of the acquisition of the transition probabilities between word roots and endings (i.e., the sMMN), an implicit-memory-based familiarity test was used to assess the behavioural correlates of this acquisition process. In the familiarity test, participants had to decide which of two words (one with the regular and one with the irregular ending) sounded more familiar. Given the decision between two sequences, the chance level was 0.5 and correct recognition significantly above that level would indicate successful acquisition at the behavioural level.

[Insert Figure 6 here]

The recognition performance (see Figure 6, left panel), was assessed over the 6 blocks of the experiment. Whereas the recognition performance was not above chance level for the first two blocks ($p > 0.123$), it was above chance level for the blocks three to six: $t(19) > 3.44$, $p \leq 0.001$, $d > 0.77$, with a correct recognition rate of at least 67.5%. The recognition performance over the whole experiment (i.e., the mean of all six blocks) was 66.0%, which was also significantly above chance level: $t(19) = 3.69$, $p = .001$, $d = 0.83$. An ANOVA assessing the recognition performance over the experimental blocks (block 1 to block 6) revealed a main effect of block, $F(_{5,95}) = 3.75$, $p = .004$, $\eta^2 = .07$.

Participant assessed their confidence regarding whether they had chosen the correct sequence on a scale from 1 to 5 (Figure 6, right panel). They rated their confidence at a mean of 3.08 in the first block and between 3.61 and 3.73 for the second to the sixth block. An ANOVA with the factor blocks revealed a main effect of blocks, $F(_{5,95}) = 8.15$, $p < 0.001$, η^2

= .09. It was furthermore assessed whether the confidence ratings (over the whole experiment) differed between trials where the “correct” word (i.e., one containing an ending that had a 90% transition probability) was chosen, compared to those where the “incorrect” word was chosen (i.e., one with a 10% transition probability). The mean confidence ratings were M = 3.11 for “incorrect” and M = 3.60 for “correct” trials, leading to statistically significant difference between those two conditions: $F(_{1,18}) = 14.46$, $p = 0.001$, $\eta^2 = 0.13$.

At the end of the experiment, there was an additional explicit-memory-based familiarity test. Here, the two endings used in the experiment were compared to four endings that were phonologically similar to the presented ending but not used in the experiment. Ratings for each ending could range from 0 to 100, with 0 signifying absolute certainty of the endings’s absence, 50 signifying uncertainty, and 100 signifying absolute certainty of the endings’s presence. The data are shown in Figure 7. The mean scores for the not presented decoy endings were “Kyem” = 76.1, “Selyem” = 78.6, “Teloj” = 30.6, and “Telu” = 36.3; for the presented endings, they were “Telya” = 52.1, and “Telyem” = 91.3.

[Insert Figure 7 here]

Discussion

The main finding of this study is that a change in the stimuli from high frequency to low frequency triggered an MMN response. This finding is in compliance with our hypothesis. As the results from the EEG analysis showed, both sMMN and phMMN were elicited. Both types of MMN showed significant results at p <.01. The sMMN were elicited at 136 – 336 ms after the onset of the syllable that distinguished the two endings *–ja/-jem*, and the phMMN was elicited 100 – 200 ms after the beginning at *–tel-*, (i.e., the onset of the direction change). The results from the EEG analysis show that a sMMN could be elicited by linguistic stimuli.

Both the sMMN and the phMMN had a prominence along the midline with a slight lateralisation towards the left hemisphere, as well as a prominence over frontal and central scalp sites (and somewhat lower amplitudes at parietal sites). This is consistent with language processing, both in general as well as in response to the acquisition of language-related regularities through SL, being more lateralised to the left hemisphere, and with earlier reports of the MMN having a distribution most prominent over fronto-central to central scalp sites (cf. Näätänen et al., 2007; Koelsch et al., 2016; Tsogli et al., 2019). When compared to the sMMN reported by Tsogli et al. (2019; using sound timbre stimuli) our results are overlapping (Näätänen et al., 2007; Plante et al., 2015; Plante et al., 2014; Plante et al., 2017) in that their results also revealed a prominence over midline scalp sites, but different in that our results had a slightly more leftwards lateralisation than theirs. The explanation for this may be that language stimuli have an influence on which parts of the brain become more activated when using linguistic stimuli. The lateralisation of the phMMN to the left hemisphere has been described also by Partanen et al. (2011), and in natural language studies by the use of fMRI (Plante et al., 2014; Plante et al., 2015). The EEG analysis showed that the amplitude of the phMMN was larger than the amplitude of the sMMN. This finding is well in accordance with that of previous experiments (e.g., Tsogli et al. 2019, Tsogli et al. 2022).

Another main finding of this study is that the participants exhibited learning in different modalities, i.e., not only neurophysiological (in the sMMN) but also when using the implicit-memory based familiarity test at the end of each block. Here, participants achieved a performance above chance level from block 3 to 6. In previous experiments using a similar experimental paradigm (e.g., Tsogli et al., 2019), no performance above chance level could be documented. It may be worthwhile to ask what drove that increase in behavioral performance. The two key differences are (1) that the stimuli used in the current experiment were linguistic in nature (whereas those used by Tsogli et al., 2019, were non-linguistic), and (b) that naturalistic language stimuli were used (i.e., words that are “legal” words in Russian, in contrast to the often rather artificial stimuli typically used in SL experiments) which provided additional prosodic cues. Both might have helped the acquisition: (a) by using stimuli that perhaps fit better than the sound timbres what our auditory system (that is highly specialized to process linguistic information) is used to process, and (b) by helping processing stages that occur before the extraction and acquisition of the transition probabilities, particularly the segmentation process. In most SL experiments, participants must solve two interacting tasks in combination: (a) segmentation and (b) rule extraction and acquisition. Here, rule extraction and acquisitions enables the segmentation of the stream of sounds, while successful segmentation helps, vice versa, rule extraction and acquisition. Prosodic cues help the segmentation process, thereby supporting the process of rule extraction and acquisition. It has been found, at least in infants, that prosodic cues not only contribute to segmentation, but are even weighted stronger when prosodic and statistical cues are conflicting (Johnson & Jusczyk, 2001). The fact that the task of acquiring the rules, given the facilitated segmentation process, was easier in the current experiment, may – at least in part – account for the better recognition performance in the familiarity test.

While the implicit-memory-based familiarity test and the sMMN were significant, the explicit-memory-based measurements were more ambiguous: Participants were not able to reliably distinguish endings presented during the experiment from decoy endings that weren't presented. The reason for this discrepancy may be that reliably reproducing such endings requires an explicit memory representation, whereas SL of transition probabilities primarily relies upon implicit memory. Thus, our results suggest that the participants did not acquire unambiguous explicit knowledge of the regularities underlying the arrangement of the language stimuli, while they clearly developed an implicit-memory-reliant, pattern-based understanding of these regularities, as indicated by the ERP results. This confirms earlier evidence and theories that SL, as reflected in the sMMN, is primarily implicit. SL can be denoted as one of several approaches to implicit, incidental learning (see Parruchet and Pacton, 2006), which is in keeping with our results.

One should take into consideration that the two familiarity tests are testing two substantially different aspects of learning, and that the findings should be accordingly. The implicit-memory-based test is affected by the segmentation and rule extraction and acquisition processes. The participants were asked to judge the frequency of items, rather than to retain the items themselves in memory, and so implicit (unconscious) representations were sufficient to make judgements. The explicit-memory-based test on the other hand, required the participants to actively remember the endings, relying on knowledge that goes beyond that what typically is acquired through SL. This will be further discussed below. In this case the participants were subject to mechanisms that affect word learning, and more interestingly, the mechanisms that influence syllable learning (see Hamanda & Goya, 2015; Cardoso, 2011, for effects, or lack thereof, of how syllable structure affects perception and learning).

However, this implicit learning was impacted by the content of the stimuli. The redundancy of input variables when understood in the context of SL, considering the extraction-and-integration framework proposed by Theissen et al. (2013), can be expanded by considerations regarding the facilitating nature of suprasegmental elements by Shukla et al. (2007). As discussed, SL, in its most basic form, dictates that unlikely phoneme successions in continuous speech usually entail a word boundary. Word boundaries would, in addition, be marked by supra-segmental prosodic cues such as stress and intonation. Whilst it is initially difficult for unfamiliar speakers to decode and use prosodic cues, it will, given that they are occuring again and again, be possible to extract and use such cues with repeated exposure to help speech segmentation. It has been shown, at least in infants, that phoneme combinations that are unlikely to occur within one word are interpreted as word on- and offsets (Mattys and Jusczyk, 2001). The SL process is then not only aided by what phonemes the listener is able to discriminate, but also these super-phonemic factors, which gives a variability in the informational input which in turn facilitates learning (cf. Eidsvåg et al. 2015).

The redundancy of the input, and how these different aspects are weighted might have different effects on the listeners. The results from the behavioural analysis might be an indication that prosodic information, found in natural language stimuli may both help and mislead the participants. This is despite the fact that the effects of the prosodic aspects of the stimuli were minimised by choosing word roots based on their phonological content and shape, rather than any particular prosodic profile. The aim was to use word roots that were two syllables in length, and that differed as much as possible from each other in their phonetic content.

We must take into consideration that the mother tongues of the participants can affect the learning of natural language stimuli, since the first language has been found to play an important role in the acquisition of a second language (Hamanda & Goya, 2015; Cardoso, 2011). When the linguistic structures are similar, this facilitates learning, whereas language acquisition would become more difficult when the linguistic structures strongly differ from each other (Hamada & Goya, 2015). It is important to consider the first language of the participants. Participants were quite varied in their language background, and the degree to which they were multilingual: 12 of the participants spoke Norwegian as their mother tongue, while the remaining eight spoke Tigrinya, German, Cantonese, Mandarin, Turkish, Urdu, Romanian or Arabic, with all of them reporting to speak English well, and some reporting fluency in a third language (typically Norwegian for L1-speakers of non-Norwegian languages). For example, consonant-heavy codas are permitted in Russian and other other Slavic languages, exemplified in toponyms like the Russian *Ноябрьск* (/nɐˈjabrʲsk/) or the Polish *Bygdoszcz* (/’ bɨdɡɔʂt͡ʂ/). Such codas are permitted in German, with words like <wirfst>, [‘viʁfst] (“you throw”) or <tropft>, [‘tʁɔpft] (“it drips), whilst in Cantonese syllables can only either be open or have a single consonant coda, and consonant clusters are forbidden (Matthews & Yip,1994). Akin to what Hamanda and Goya (2015) showed for the Japanese L1 speakers, when compared to English L1 speakers, the influence of the phonotactics L1 had an impact on learning.

Interestingly, the different language backgrounds notwithstanding, the participants generally indicated certainty for the presence of stimuli with a coda. The results that showed that the participants favoured closed syllables is contrary to some research on L2 learning, which indicate that open syllables generally are easier to learn (Hamanda & Goya, 2015; Cardoso, 2011), whilst others argue that the syllable structure that an individual is more likely to find easiest is language dependent (Cutler, 1997).

Taken together, whereas participants would show clear signs of the acquisition of implicit-memory-based representations of the transition probabilities between word roots and endings, as indicated by the sMMN and their performance in the implicit-memory-based familiarity test, the results regarding explicit learning are ambivalent at best. Although the knowledge about the transition probabilities can be regarded as quite complex, and quite definitely going beyond capabilities of the sensory memory (cf. Tsogli et al., 2019), participants acquired that knowledge relatively quick, demonstrated by the presence of the sMMN from the first block and the above-chance-level performance in the implicit-memory-based familiarity test from block 3 onwards.

Conclusion

In this study we examined whether an MMN would be elicited in participants exposed to linguistic stimuli, and whether there would be an explicit-memory-based representation of those linguistic stimuli and their occurrence. The analysis showed that both sMMN and phMMN were elicited, as expected. Both the sMMN and the phMMN responses we found to be slightly lateralised to the left hemisphere.

In addition, the participants did also establish implicit-memory-based knowledge though SL, as documented by their (above chance) performance in the familiarity tests ending each block. In contrast, explicit knowledge of the underlying structures of the language stimuli was rather not established. This could indicate that given that SL is a learning mechanism that primarily relies on implicit learning, the established knowledge resulting from such learning is easier picked up by measures relying on implicit memory.

References

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology andl Biofeedback*, *42*(4), 247-255. https://doi.org/10.1007/s10484-017-9371-3

Brain Products GmbH. (2019). *BrainVision Recorder*. BrainVision Recorder. https://www.brainproducts.com/solutions/recorder/

Cardoso, W. (2011). The development of coda Perception in Second, language phonology: A variationist perspective. *Second Language Research, 27(4)*, 433–465. https://doi.org/10.1177/0267658311413540

Cutler, A. (1997). The comparative perspective on spoken-language processing. *Speech Communication, 21*(1), 3–15. https://doi.org/10.1016/S0167-6393(96)00075-1

Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods 134*, 9–21

Eidsvag, S. S., Austad, M., Plante, E., & Asbjornsen, A. E. (2015). Input Variability Facilitates Unguided Subcategory Learning in Adults. *Journal of Speech, Language, and Hearing Research*, *58*(3), 826-839. https://doi.org/10.1044/2015_JSLHR-L-14-0172

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental review*, *37*(Sep), 66-108. https://doi.org/10.1016/j.dr.2015.05.002

Hamada, M., Goya, H. (2015). Influence of Syllable Structure on L2 Auditory Word Learning. *Journal of Psycholinguistic Research 44*, 141–157 . https://doi.org/10.1007/s10936-014-9284-8

Hugdahl, K. (1995). *Psychophysiology : the mind-body perspective*. Harvard University Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language, 44*(4), 548–567. https://doi.org/10.1006/jmla.2000.2755

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131-141. https://doi.org/10.1111/j.1467-7687.2008.00740.x

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179-190. https://doi.org/10.1111/j.1469-8986.2008.00752.x

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016b). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports*, *6*(1), 19741-19741. https://doi.org/10.1038/srep19741

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831-843. https://doi.org/10.1038/nrn1533

Matthews, S. & Yip, V. (1994). *Cantonese: A Comprehensive Grammar*. London: Routledge

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121. https://doi.org/10.1016/S0010-0277(00)00109-8

Näätänen, R. (1995). The Mismatch Negativity: A Powerful Tool for Cognitive Neuroscience. *Ear and Hearing*, *16*(1), 6-18. https://doi.org/10.1097/00003446-199502000-00002

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical neurophysiology*, 118(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *International Journal forof Psychophysiology*, *48*(2), 179-188. https://doi.org/10.1016/S0167-8760(03)00053-9

Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. (2011). Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles: Linguistic multifeature MMN paradigm. *Psychophysiology*, *48*(10), 1372-1380. https://doi.org/10.1111/j.1469-8986.2011.01214.x

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of memory and language*, *39*(2), 246-263. https://doi.org/10.1006/jmla.1998.2576

Plante, E., Almryde, K., Patterson, D. K., Vance, C. J., & Asbjørnsen, A. E. (2015). Language lateralisation shifts with learning by adults. *Laterality*, *20*(3), 306-325. https://doi.org/10.1080/1357650X.2014.963597

Plante, E., Patterson, D., Dailey, N. S., Kyle, R. A., & Fridriksson, J. (2014). Dynamic changes in network activations characterize early learning of a natural language. *Neuropsychologia*, *62*(Sep), 77-86. https://doi.org/10.1016/j.neuropsychologia.2014.07.007

Plante, E., Patterson, D., Gómez, R., Almryde, K. R., White, M. G., & Asbjørnsen, A. E. (2015). The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics*, *36*, 17-34. https://doi.org/10.1016/j.jneuroling.2015.04.005

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjørnsen, A. E. (2017). An fMRI study of implicit language learning in developmental language impairment. *Neuroimage Clinical*, *14*(C), 277-285. https://doi.org/10.1016/j.nicl.2017.01.027

Saarinen, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, *3*, 1149– 1151.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1-32. https://doi.org/10.1016/j.cogpsych.2006.04.002

The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The Extraction and Integration Framework: A Two-Process Account of Statistical Learning. *Psychological Bulletin, 139*(4), 792-814. https://doi.org/10.1037/a0030801

Tsogli, V. Jentschke, S., Daikoku, T., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports*, *9*(1), 5563–5563. 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): e0263373. https://doi.org/10.1371/journal.pone.0263373 Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1-2), 239-252. https://doi.org/10.1016/S0006-8993(96)01008-6

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1-2), 239-252. https://doi.org/10.1016/S0006-8993(96)01008-6

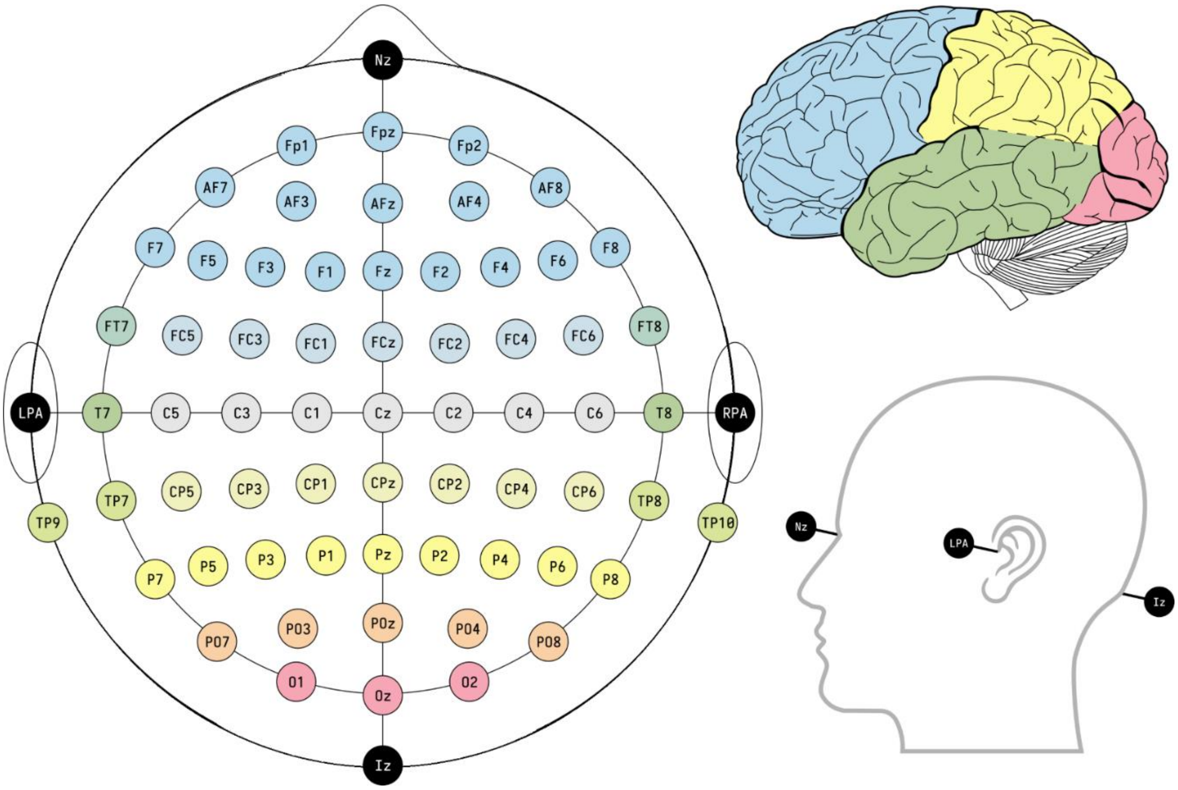
FOOTNOTES

[1] The participants and electrodes in question that were removed were: participant no. 9, electrodes CP1 and O1. Participant no. 12, electrodes FP2 and F4. Participant no. 13, electrode P4. Participant no. 15, electrode F5.

Figures

Figure 1

Electrode layout used during the experiments: The electrodes in colours (i.e., not those in black) were recorded during the experiment. TP9 served as reference electrode during the acquisition, Iz as ground electrode. Not shown are the electrodes used to track vertical (on the right chin) and horizontal eye movements (at the outer canthi of the eyes on either side).



Note: Image adapted from "EEG 10-10 system with additional information" by L. R. Krohl, 2020. (https://commons.wikimedia.org/wiki/File:EEG_10-10_system_with_additional_information.svg) CC 0

Figure 2

Average ERPs (left panel) and scalp distribution (right panel) comparing the irregular and the regular stimuli generating the phMMN. The sound came either from a standard direction (regular stimuli) or the opposite site (irregular stimuli; representing a change in the physical characteristics eliciting the phMMN). The time window (100 to 200 ms after stimulus onset) for which the scalp distribution is shown is indicated by the box with the dotted lines in the ERPs. The scalp distribution was maximal over fronto-central electrodes and was slightly lateralised to the left hemisphere. The same time window was used in the statistical analyses.

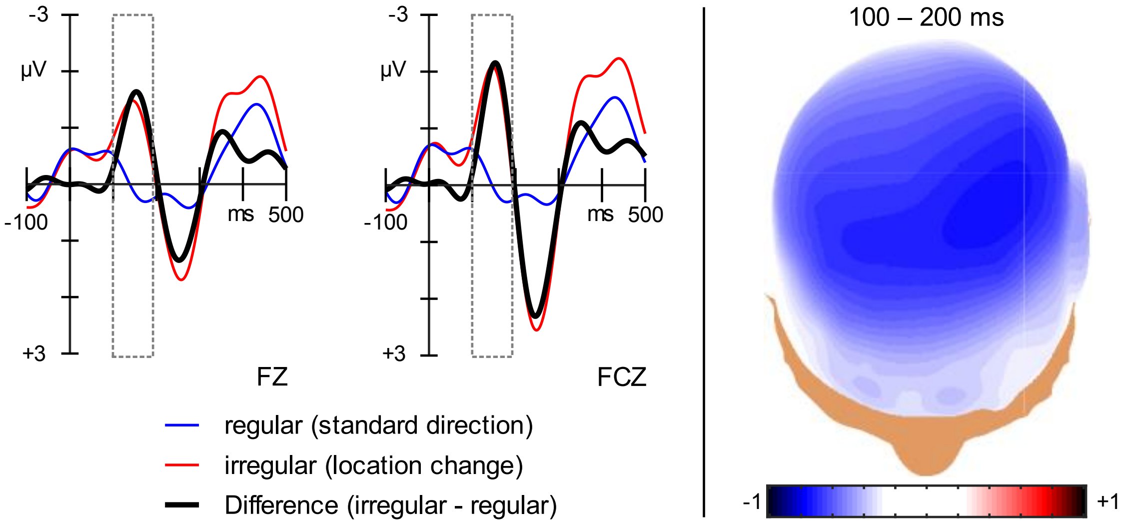


Figure 3

Amplitude differences between physically regular (black lines; the sounds coming from the standard direction) and the irregular stimuli (grey lines; the sounds coming from the opposite direction) throughout the different regions of interest (ROIs) used in the statistical analyses. The three diagrams show the anterior to posterior axis, and within each diagram, the left to right distribution can be found on the x-axis. The greatest amplitude difference was observed at scalp areas around the midline. The amplitude difference was, in addition, most pronounced over central sites, a bit smaller at frontal sites and lowest over parietal sites. In addition, there is a slight lateralisation towards the left hemisphere, best observable at central scalp sites.

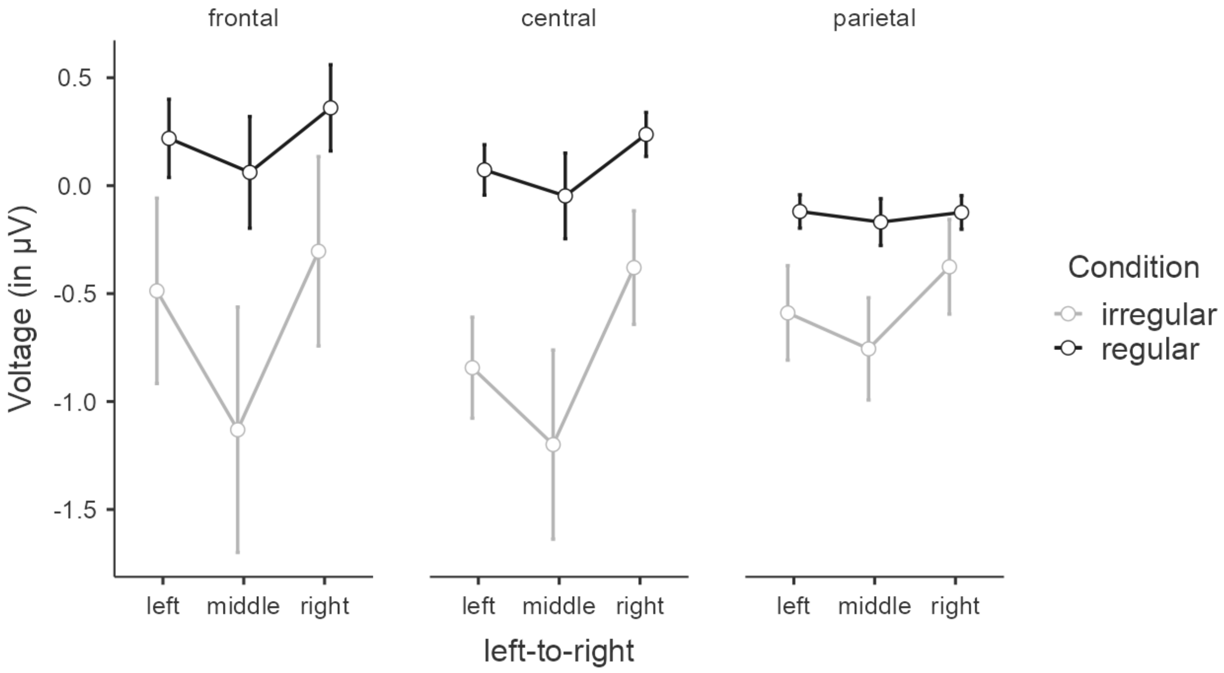


Figure 4

Average ERPs (left panel) and scalp distribution (right panel) comparing the irregular and the regular endings generating the sMMN. Each of the different word roots in the experiment could be followed by an ending with high transition probability (90%, regular ending) or with low transition probability (10%, irregular ending). The difference between the brain response to the irregular vs. the regular endings represents the sMMN. The time window (300 to 500 ms after the onset of the ending) that was used for the scalp distribution is indicated by the box with the dotted lines in the ERPs. The scalp distribution was maximal over fronto-central electrodes and was slightly lateralised to the left hemisphere. The same time window was used in the statistical analyses. The regular and the irregular endings shared the first syllable (-tel), lasting 164 ms, which is indicated by the grey box in the ERPs. Therefore, the chosen time window equates to 136 to 336 ms after the endings began to differ (i.e., the end of –tel).

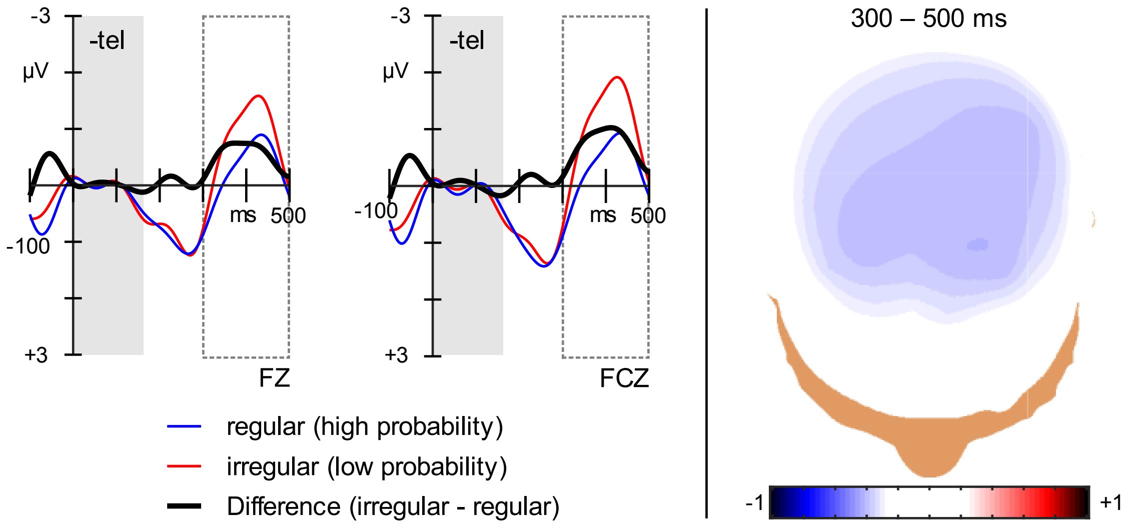


Figure 5

Amplitude differences between the statistically regular (black lines) and the irregular stimuli (grey lines) throughout the different regions of interest (ROIs) used in the statistical analyses. The three diagrams show the anterior to posterior axis, and within each diagram, the left to right distribution can be found on the x-axis. The greatest amplitude difference could be observed at scalp areas around the midline. The amplitude difference was, in addition, more pronounced over frontal and central and frontal sites than over parietal sites along the anterior-posterior axis of the scalp. In addition, there is a slight lateralisation towards the left hemisphere, best observable at central scalp sites. The greatest activation during irregular stimuli occurred in the frontalo-central areas of the scalp. The interaction of the factors condition (irregular vs. regular) and anterior-to-posterior scalp distribution was not significant.

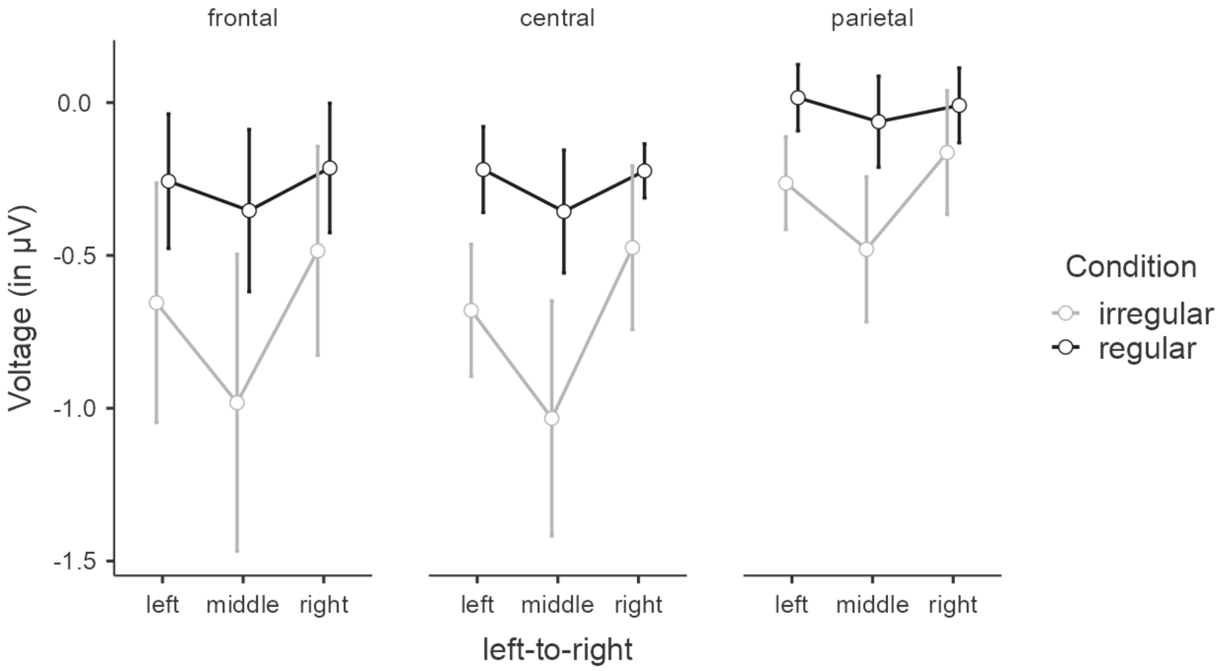


Figure 6

Recognition performance (left panel) and confidence ratings for the implicit learning task implicit-memory-based familiarity test at the end of each block, representing how often the regular ending was chosen (0 – never, 1 – always). Given that the participants had to decide between two sequences, the base rate was at 0.5 (indicated by the dotted line) and a recognition performance above that level can be regarded to reflect the implicit learning of the transition probabilities between the word roots and the endings. After deciding about which sequence sounded more familiar, the participants had to give a rating how confident they were about their decision (right panel).

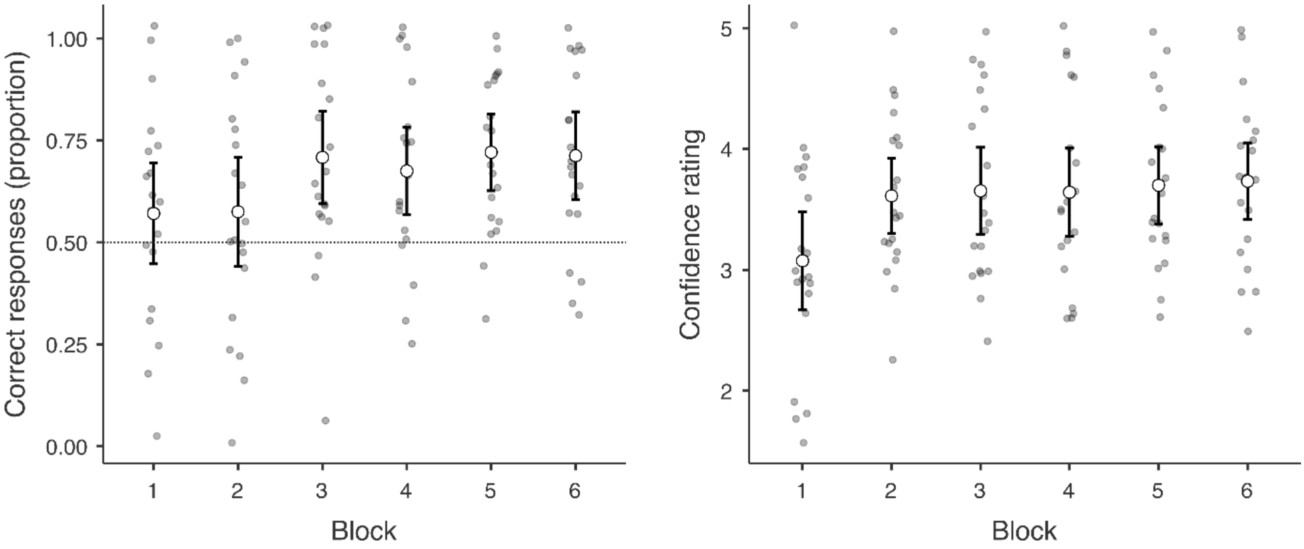
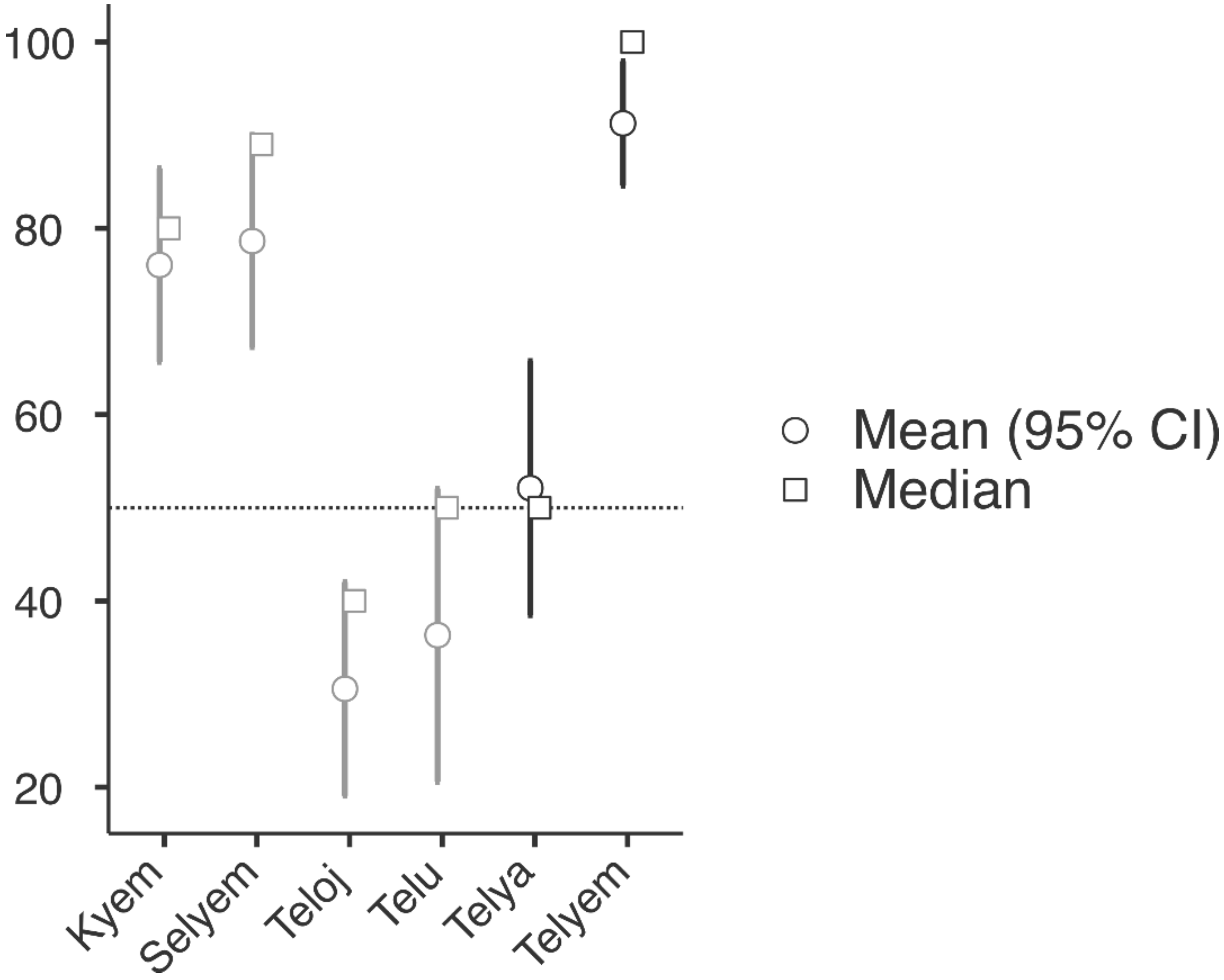


Figure 7

Confidence ratings for the different possible endings, in the explicit-memory-based test after the experiment (0 = Absolutely certainty of absence, 50 = Unsure, and 100 = Absolutely certainty of presence). The black graphs represent the endings used in the experiment, the grey graphs the decoy alternatives.



9.0 Appendixes

Appendix A

Declaration of consent form

Declaration of consent

I have received and understood information regarding the research project Exploring the MMN of statistical learning in a natural linguistic context –a pilot study, and been given the possibility to ask questions. I hereby consent to:

- ☐ Participate in the collection and production of EEG-data
- ☐ Answering the questionnaire after the data production
- ☐ That my anonymous data be stored in a public repository

I consent to the use and handling of my data until the end of the research project.

(Participants signature, place, date)

Appendix B
Informational leaflet provided prior to participation

Information regarding participation in language processing research

Dear participant,

You now have the possibility to take part in a research project that aims to explore the ways the brain responds to an unknown language. The purpose of this document is to inform you of the goals of the project and what participation entails for you.

Purpose

This study aims to explore the ways the brain responds to being exposed to an unknown language. To do this, we will run electroencephalographic (EEG) experiments with audio-based linguistic stimuli. This experiment will take place once per participant, and it will take place in a laboratory. The study is a pilot, which means that the total number of participants will range from 20 to 30. As the epithet “pilot” indicates, this study could lead the way to more extensive research into this subject field.

Who is responsible for the research project?

The responsibility lies with the university of Bergen, and we are part of the Faculty of Psychology. The study is part of a research collaboration between the research groups Forskningsgruppen for kognisjon og læring (The research group for cognition and learning) and Hjerne- og musikkgruppen (The brain and music group). The experiments are part of a masters’ thesis in speech-language pathology.

Who can participate?

We are looking for healthy adults between the ages of 18 and 65. There are five basic criteria for participation. Participants must:

- Be right handed
- Be unfamiliar with Russian and other Slavic languages
- Not be dependent on hearing aids, have cochlear implants or be deaf/hard of hearing
- Have no diagnosed neurological or psychological conditions
- Have a maximum of two years musical training outside of mandatory schooling.

What does participation entail for you?

The study uses data obtained through electroencephalography (EEG), which measures the electrical activity of the brain through electrodes placed on the scalp. In the experiment, these electrodes are placed in a cap and placed over the head, and a helping gel is applied to the scalp to insure good connectivity. The gel used is salt-based ang should be allergy-free. After the data has been collected you will have the possibility to wash it out of your hair. We will supply towels and shampoo, but you might want to bring your own comb/hairbrush. In addition to the electrodes in the cap, seven additional electrodes will be fastened to the face, and these points of the face will need to be disinfected to ensure good connectivity

. When all preparations are done you will hear a series of auditory stimuli through a pair of headphones. After listening you'll be asked to fill in a short questionnaire.

No psychological or physical complications are to be expected from this kind of experiment. If you at any point during the experiment feel negatively affected, this must be communicated to the study administrators at once.

The whole experiment, preparations included, is expected to last about 2 hours (120 minutes). The testing will take place in our laboratory in Christies gate 12, at the psychological faculty of the university of Bergen. You will be compensated 200 NOK for you participation, in the form of universal gift cards.

Participation is voluntary

It is voluntary to participate in this study. If you choose to participate, you can withdraw your consent at any time, without giving any reason. In this case, all of your information will be deleted. If you do wish to withdraw from the experiment, please inform the study administrators. There are no consequences for not wanting to participate, or withdrawing consent at any point.

Your privacy – how we store and utilise your personal information

We will only use your personal information for the goals previously described in this document. We treat your information confidentially and in keeping with current legislation.

No personal information will be stored digitally. Rather, it will be stored on paper, in a locked cabinet in an area with restricted access to the public. Only members of the study will handle the information.

All your information will be rendered anonymous, disconnected from your name or any other information that could link your data back to you.

What happens to your personal information once the study is concluded?

The data that will be collected will be made anonymous, and thus it will not be possible to trace the participants of the study. Your contact information will be stored separately from the data and will be deleted as soon as the data collection phase of the study is finished. This means that your contact information cannot be connected to the data collected in the experiment.

The data collection phase of this study is scheduled to be completed June of 2023. After this, we intend to publish our results. The publication will be based on averages from several participants, and no data can be used to trace you specifically. Until the article has been published you have the possibility to withdraw your consent.

After publication, all of your personal information (e.g., consent forms), will be deleted. In accordance with the principle of open and public research, ***anonymous*** data may be rendered accessible to the public through a third-party storage location (e.g., Open Science Foundation). Public data will not be sored in a way that individual participants and be traced.

The goal of the rendering these data public is for them to be utilised in the best manner possible. For instance, other researchers will be able to use these data and reanalyse them, or use them to answer other research questions.

On what basis do we keep and use your personal information?

We store and utilise your personal information strictly for research purposes, and with your consent.

If you have further questions about the study, please contact:

Aina Cecilie Klinge or Olav Tidemann Garli.

If you have further questions about your rights, or wish to exercise them, please contact:

- Project leader: Arve Egil Asbjørnsen, Faculty of Psychology
- Our ombudsman for privacy: Janecke Helene Veim

Kind regards,

Arve Egil Asbjørnsen	Aina Cecilie Klinge and Olav Tidemann Garli
Project leader	*Students*
(Researcher/advisor)	

Appendix C

Pre and post-test self-reporting forms

Participant number:____

Self-report questionnaire

Pre-test

Gender:____

Age:____

Knowledge of languages:

1) What is your mother tongue?
    - ☐ Norwegian
    - ☐ Other

If your answer is something else than/ in addition to Norwegian, please specify:

____

2) What other languages do you know in addition to your mother tongue, Swedish and Danish? To what degree do you know these?

	Not at all	A little	Medium proficiency	Fluently	As mother tongue
English					
German					
French					
Spanish					
Polish					

If you know other languages not listed above, please list them below together with your level of proficiency:

____

____

____

3) Do you have any knowledge of Russian grammar?

- ☐ Yes
- ☐ No

4) Have you taken music lessons (beyond mandatory schooling)?
   - ☐ Yes
   - ☐ No

5) Do you play an instrument?
   - ☐ Yes
   - ☐ No

If you answered yes, please specify:

____
____
____

6) How many years have you been playing?

____

Self-report questionnaire

Post-test

Questions about the stimulus material

1) How conscious were you of searching for patterns of regularity in what you heard?
   - ☐ Not at all
   - ☐ To a minor degree
   - ☐ To a medium degree
   - ☐ To a large degree
   - ☐ To a very large degree

2) Did you think that the audio contained some pattern you could follow?
   - ☐ Not at all
   - ☐ To a minor degree
   - ☐ To a medium degree
   - ☐ To a large degree

☐ To a very large degree

If so, please specify:

3) Please provide a rating of how sure you are of the appearance of these word endings (0 = Certain it was absent, 50 = unsure, 100 = sure it was present)

-kyem		-telya	
-selyem		-teloj	
-telu		-telyem	

4) Did you recognise any of the words? If so, please specify:

5) If you have any comments to the experiment in general, please write them below:

That was all, thank you for answering our questionnaire!

Appendix D

Lyst til å delta i EEG-eksperiment?

Vi søker deltakere til et forskningsprosjekt hvor formålet er å undersøke hva som skjer i hjernen når vi lytter til et ukjent språk. Eksperimentet vil finne sted på lab i Christies gate 12, og det vil ta omtrent 2 timer å gjennomføre (inkl. forberedelser og pauser). Datainnsamlingen vil skje i løpet av oktober, og man blir kompensert med 200 kroner for å delta.

EEG utføres ved at du får en hette på hodet hvor det festes elektroder. Elektrodene festes til hodebunnen med en gel, denne vaskes av etter at eksperimentet er ferdig (NB: vi har sjampo, hårføner og håndklær!)

For å delta i studien må du:

- Være mellom 18 og 65 år
- Høyrehendt
- Ikke ha kjente språkvansker eller vansker med hørselen
- Ikke ha eller ha hatt kjente nevrologiske eller psykiske lidelser
- Ikke være kjent med russisk eller andre slaviske språk

Ønsker du mer informasjon eller å delta i studien?

Ta gjerne kontakt!

Aina Klinge	Olav Garli
Aina.Moe@student.uib.no	Olav.Garli@student.uib.no

Veiledere: Arve Asbjørnsen, Sebastian Jentschke og Sunniva Sørhus Eidsvåg