Abstract

Background. When processing auditory stimuli, the brain uses pattern recognition to perceive and predict future events. If the presented event deviates from the cognitive prediction, a reaction is elicited. This study aims to find correlations between neural and behavioral reactions to deviant stimuli.

Methods. Two experiments were performed to measure reactions to prosodic feature deviation; one electroencephalography experiment detecting online reactions using auditory rhythm deviation, and one behavioral experiment detecting offline reactions by using 144 Norwegian words where 2/3 of the words had sustained prosodic alterations. The data of the two methods where then compared to find correlations in event related potentials (ERP), response time and accuracy.

Results. We tested the correlation between the ERP data and the behavioral data statistically using a t-test. We found that the mismatch negativity amplitude correlates with response time when it comes to the words that have been manipulated to be unacceptable with ($t(1,9)=-3.3605, \rho=-0.746, p=0.008$). We also found that both word length and tone placement has a main effect on response time, and that accuracy is an effect of tone placement.

Conclusion. There was a correlation in mismatch amplitudes with the response time in the words where prosodic deviants were manipulated to be unacceptable.

Acknowledgements

First of all, I would like to thank my supervisor Giosué Baggio. I would not have had the courage to write this thesis had it not been for his guidance, reassurance, and never-ending patience. A thank you to my co-supervisor Mila Dimitrova-Vulchanova for her competence and support in the initial stages of the project, and to Massimo Lumaca for letting me observe, assist, and utilize his data during and after the EEG-trials. A 'tusen takk' to Randi Nilsen for assisting me through all of her Norwegian expertise, and many thanks to Sobh Chanhoun, Christian Dillner Hagen, and Hendrik Bertil Eshuis for their invaluable technical expertise regarding the programming of the behavioral experiment.

A special thank you to Mona and Tonje, who have been the best companions one could ask for during these two years. Thank you for your companionship during late night frights, temporary meltdowns, and literal ROFLMAOs.

Writing this thesis would not have been the same without the laughter, tears, candy, daily rants, and blackboard drawings behind the walls of study hall D113. Thank you to Ingvild, Sunniva, and Turi Marte who also submitted their thesis this spring, with an extra thank to Turi Marte, for the amazing efforts through the initiative of 'Masterverksted'.

To Jon: I know you always have my back, but during this process you have proved to be a bigger Lene-supporter than you are a Manchester United-supporter. A special thank you for retrieving my sock during *that* moment of morning despair, and for bringing me food when items from the SiT-kiosk did not suffice.

I would also like to thank Cantus for being the best distraction one could ask for, and thank you to the Cantus-girls in 'The Dragvoll Experience' for your constant encouragement, patience and hugs.

Lastly, I would like to thank my friends on Snapchat for not unfollowing me.

List of abbreviations

EEG	Electroencephalography
ERP	Event related potentials
MMN	Mismatch Negativity
MS	Milliseconds
SD	Standard Deviation
SEN	Standard East Norwegian

Table of Contents

List of figures and tables

List of figures

List of tables

1. Introduction

Our perception of the world lays the ground rules for how we predict and process the events that we are exposed to throughout the course of our lives. Whether it is instantaneous sensory stimuli or situational events, how we process the events that are presented to us are based on stored cognitive patterns which are either innate or acquired. When it comes to auditory language perception, these patterns are based on a multitude of factors: Vocabulary, grammar, and pronunciation are all vital elements for language processing, language comprehension, and language production, but some languages put more emphasis on specific linguistic elements than others. This thesis is centered around prosody, which includes language features that have a greater significant influence in tonal languages than in non-tonal languages. As prosody includes features such as intonation, stress, tone, and rhythm, it can be used to convey or emphasize emotion in speech just by making small adjustments to a linguistic segment, word, or phrase. In tonal languages, prosody can also change both the lexical form as well as the semantic meaning of a word or a phrase, and thus be a factor in completely transforming the original segment. Earlier studies propose that speakers of tonal languages are able to be more competent in distinguishing prosodic differences than those who do not speak a tonal language: "Because of the fundamental linguistic importance of tone in these language systems, one might predict a different pattern of prosodic control than is found in speakers of non-tone languages" (Baum & Pell, 1999, p. 585). With perception, prediction and prosody as a base, we wanted to do an auditory experiment using native speakers of Norwegian to try to find correlations between the three.

Our research objectives are the following: we aim to assess whether low-level auditory processes, that are not specifically speech perception processes, are actually recruited during perception of tone (mis)placement in Norwegian words. We measured participants´ sensitivity to irregularities in low-level auditory sequences (pure tone sequences) using the MMN in ERPs, and the perception of tone placement in Norwegian words in an independent behavioral task.

1.1 Hypothesis

Null hypothesis: *There are no correlation between irregularities in low-level auditory sequences (pure tone sequences) using the MMN in ERPs and the perception of tone placement in Norwegian words in an independently conducted behavioral task.*

Evidence against the null hypothesis would come in the form of a correlation between MMN features (e.g., latency or amplitude) and behavioral measures (i.e., response times and accuracy).

2. Theory

2.1 Perception and prediction

When we are exposed to various sensations, our brain starts the perception process to interpret and analyze the incoming sensory stimuli. This process enables the brain to navigate through a variety of responses in order to find the response that is contextually appropriate, but also to prepare us for forthcoming events by relating the perceptive process to sensory patterns that are already stored in our cognition:

> When we perceive a stimulus, our brain generates a complex pattern of neural activity, reflecting the summation of a large number of information-processing stages, some of which correspond to the conscious processing of perceived representations, whereas others reflect nonconscious processing. (Bekinschtein et al., 2009, p. 1672)

Winkler, Denham and Nelken (2009) propose that the theories for how humans interact constantly with the future are typically devised through a method of statistical inference known as *Bayesian inference*, where the probability for a hypothesis is updated when additional evidence or information emerges. They include that "the ´purpose´ of perception is to generate testable hypotheses about the causal structure of the external world, based both on prior knowledge and the current sensory input" (I. Winkler, Denham, & Nelken, 2009, p. 532), supporting our notion regarding prediction being a vital element in directing behavior due to its regularity based information processing.

Whether it is a conscious choice or a subconscious decision, prediction influences our lives, even in the little things. This can be trivial things such as how we expect that a cup of coffee will help us wake up after a bad night's sleep, how we time our travel based on how long we estimate that it will take to arrive to our desired destination, and knowing that there will be a foul smell when opening a trash can lid. These predictions are based on a pattern of regularities that we have acquired through various stimuli; some we have learned through personal experience, while others by acquiring knowledge through external sources such as other people, books, media, cultural influences and so on.

In the instances that our predictions are challenged, we depend on being able to quickly change the pre-determined reaction that our brain already has prepared for itself for. This can happen for example when we expect that the approaching car will stop at the red light as we assume that the driver of the car possesses the same long-term stored knowledge that we do (that the red light means 'stop'), and we thus predict that it will be safe to cross the road based on these stored patterns that we believe we share with the driver. However, if our assumption that the car will stop proves to be wrong, our pre-determined decision to cross the road will be challenged and the brain will immediately withdraw the notion of safe passing by replacing our initial predictions with new predictions that are adapted to the sudden change in events. The brain has now changed its prediction from the previous knowledge that car stopping means safe crossing, to the car not stopping to meaning one might be hit by said car. Although we might not have first-hand experience when it comes to being hit by a car, we do possess implicit knowledge that being hit by a car does not correspond with a desired situational outcome. As predictions vary with time as well as of importance, we need to rely on that our brain processes the specific stimuli not only correctly, but also within the appropriate time frame.

To be able to apply the correct response to its respective stimuli, a distinction between *online* and *offline* responses is needed. According to Waller and Greenauer (2013), the cognitive processing system can be divided into two sub-categories, online processing and offline processing. Where online processing relies on our working memory and deals with how we process our immediate surroundings through perceptual and sensory information in addition to the covariant motor processes of this type of information, offline processing is concerned with long term memory and stored patterns. They use spatial navigation as an example of these processes, where online processing is explained through the continuous information processing when navigating through a well-known space without being consciously aware, while offline processing shows contrast to the online spatial awareness in the instance where we are asked for directions regarding the same space (Waller & Greenauer, 2013).

As we have addressed in this section, prediction can be both universal and subjective. Most of us will instinctively move away when we encounter fire, and we can easily identify the smell of freshly baked bread emerging from a bakery in an environment full of competing olfactory stimuli. We do not have to ever have been burnt in order to know that we should avoid an open flame, and we do not need to have experienced food poisoning to react with disgust to the smell of rotten meat.

2.2 Neuroanatomical basis of language

According to Freberg (2010), a simple classification of the brain is dividing it into four lobes, or sections; the frontal lobe, the parietal lobe, the temporal lobe, and the occipital lobe. In order to process information coming from the sensory systems, the cerebral cortex consists of three functional areas; the sensory cortex, association cortex, and the motor cortex. Various areas of the sensory cortex are found through the occipital, temporal and parietal lobes, with the primary auditory cortex being located in the temporal lobe. Crossman and Neary (2005) further explains the layout some of the neurological processes in the brain relevant to our study: “Nearby regions of the temporal lobe and parietal lobe, most notably the angular gyrus and supramarginal gyrus of the inferior parietal lobule, provide a functional interface between auditory and visual association areas important in naming, reading, writing and calculation” (Crossman & Neary, 2005, p. 140)

To help distinguish different processing tasks in various areas of the brain, we need a tool for classifying different brain areas. As Freberg (2010) claims, the most widely used system to illustrate and label cortical structures is what is known as *Brodmann’s system*, where different sections of the brain are categorized into 52 different areas which each constitute a separate localization of the cortex. The two most known centers for language processing are Brodmann’s area 44 and 45, better known as Broca’s area, which is argued to be the center for speech production, and Brodmann’s area 22, or Wernicke’s area, which is said to be responsible for speech comprehension. Although these two centers have been attributed the processing of language, it appears that larger sections such as the frontal, temporal, and parietal lobes are involved. In addition to these, it occurs that the cingulate cortex, insular cortex, and the basal ganglia are all associated with language processing (Freberg, 2010).

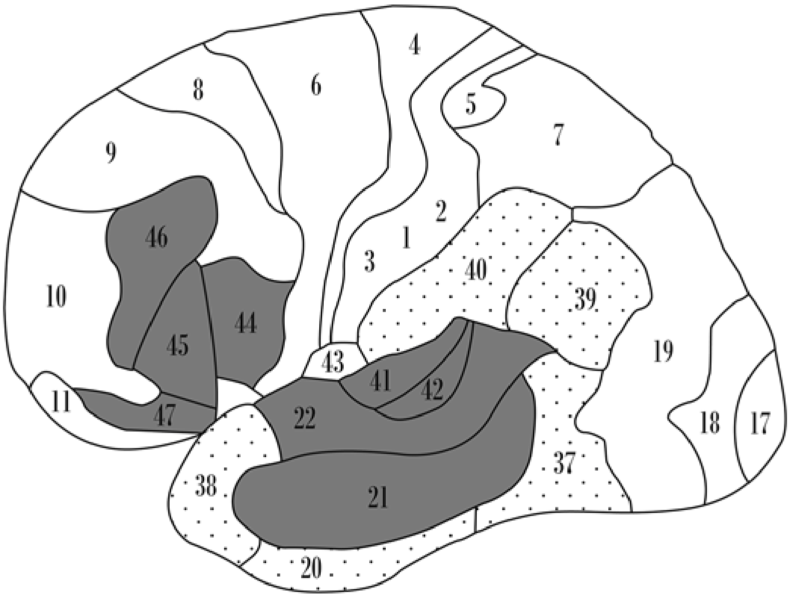

Figure 2.1 Neuroanatomical map of cortical areas of the brain involved in language processing. Newer research proposes a new and expanded way of area classifications regarding Brodmann's classifications, where Wernicke's area is suggested expanded from the core areas BA 21, BA 22, BA 41, and BA42 to an extended Wernicke's area including the areas BA20, BA37, BA38, BA39, and BA40 as pictured (Ardila, Bernal, & Rosselli, 2016) Reprinted with permission.

Crossman & Neary (2005) explain further that Wernicke's area is also known as the 'auditory association cortex', as it is the area that interprets information according to contextual significance. As the frontal, temporal, and parietal lobes of the left hemisphere are responsible for language comprehension and expression, we say that language, as well as mathematical ability, is lateralized to the left hemisphere, while the right hemisphere is better when it comes to musical proficiency and spatial perception. This is also where you find Broca's area, which is said to be an area which contains connections between the occipital, ipsilateral temporal, and parietal lobes that are known to be engaged in language function. Brodmann's area 41 and 42 is where the auditory cortex is located, and it is here we find the location for responsibility of conscious, or offline, sound perception. Surrounding this is the auditory association cortex, the previously mentioned Wernicke's area, an area which is essential in understanding spoken words.

2.2.1 Lateralization of speech

The notion of that speech sounds is both produced and comprehended in the left side of the brain is widely attributed as an accepted one, while non-speech sounds are perceived in the right hemisphere. However, there have been studies pointing to how certain aspects of speech is processed either in the right hemisphere, or as a more complex process including several parts of the brain as a combination of structures. Hagoort and Poeppel (2013) say that as there is now well established that speech perception does not adhere to an isolated brain area but is rather a process that spans across several brain regions throughout the cortex of the brain, there are a number of steps that are taken in order to complete the process of language perception. They start with how the parietal areas, temporal lobe areas, and various frontal regions form the speech recognition system, and then moves to how the processing of speech sounds are due to the mapping of sound input taking place in a ventral temporal lobe pathway, and how Sensorimotor transformations underlying mapping to output representations are enabled by a dorsal patch that include frontal and parietal lobes. To simplify the speech perception structure, they refer to a rough division by Ben Shalom and Poeppel (2008), where memory is credited to the temporal lobe, analysis to the parietal lobe, while unification is attributed to the frontal lobe. In addition to this, Hagoort and Poeppel (2013) argue that areas in the temporal lobe have a role in retrieving and storing speech sounds and words (Peter Hagoort & Poeppel, 2013). A more detailed overview of how the functional neuroanatomy of auditory language comprehension is situated is posed by Friederici (2002) as she explains it as a bilateral temporo-frontal network through the following classifications:

- The left temporal regions support processes that identify phonetic, lexical and structural elements
- The left frontal cortex is involved with sequencing and the formation of structural, semantic and thematic relations
- The right temporal region is thought to support the identification of prosodic parameters
- the right frontal cortex is involved in the processing of sentence melody (Friederici, 2002, p. 79).

We have earlier addressed how sentence or text processing, lexical-semantic processing, and phonological processing is mostly in the left hemisphere, but that there also are activations in the right hemisphere. Hagoort and Poeppel (2013) refer to a series of neuroimaging papers on language processing by Vigneau et al. (2011), where left and right hemisphere activations were compared. For the majority of cases, these activations in the right hemisphere were in what are called homotopic areas, which in this context means that there are two continuous functions leading from one brain area to another, indicating that the interhemispheric influence is strong (Vigneau et al., 2011). One could thus dare to make the assumption that for the majority of the population, with the exception of some of those who are left-handed, the readiness of language largely resides in the left perisylvian cortex, where perisylvian cortex refers to the area nearby the Sylvian fissure (also known as the lateral sulcus). Without effort, speakers produce around two and five words per second, and the listener has to decode this information close to instantaneous in order to process and comprehend the utterance. In addition to this, if the listener is expected to respond, they need to be able to create an appropriate response within seconds of the other speaker completing their utterance. Furthermore, they say that several studies support the notion of language being strongly lateralized to the left in the population majority, but that one cannot say that language function is completely lateralized as there is an increasing amount of evidence that the right hemisphere is involved in essential aspects of language processing. There is thus no way one can claim that specific elements of language processing adhere to particular brain areas as multiple processes are engaged simultaneously when processing language (Peter Hagoort & Poeppel, 2013).

2.2.2 What is sound

We experience the sensation of sound when we are exposed to acoustic stimuli. Brodal (2010) explains how sound waves are comprised of pressure variations, which are amplified by the structures in the middle ear (the eardrum and ossicles), and eventually transduced to an electrical signal by the structures in the inner ear. The electrical signal is further transmitted to the brain by the eight cranial nerve (the vestibulocochlear nerve) (Brodal, 2010). Freberg (2010) argues that although the human can hear sound through both liquids and solids, the medium that is usually used to carry sound is air, where sound velocity is about 330 m per second. The sounds that we hear begin with an object moving in one of these means, creating a disruption which we can measure in form of amplitude and frequency. Where amplitude, or loudness, measures the height of the soundwave, the frequency measure the number of wave cycles per unit of time. We use decibel (dB) to measure the amplitude of a sound wave, where the extremities range from 0 dB, meaning the threshold of hearing, to instant perforation of the eardrum at about 160 dB. Both music and speech are composed of complex mix of frequencies, where F0, or the fundamental frequency, measure a tone’s lowest point of frequency, thus determining its pitch. Human beings usually have their lowest hearing threshold at the frequency around 30 Hz, while the highest frequency is usually around 20kHz. Frequencies that are lower than this are called infrasound, while higher frequencies are called ultrasound (Freberg, 2010). To avoid confusion, it is important to note that in this study, F0 is also known as *intonation*, a prosodic feature that will be addressed further in chapter 2.4.

2.2.3 Auditory perception and processing

According to Warren (2008), the term *pitch* is used to explain how an auditory sensation can be measured on a low to high scale, and is thus limited to a frequency which is audible. When it comes to the presence of pitch in the context of music, we say that pitch has a span of around seven octaves, where the lowest limit is at 41 Hz, and the upper limit is about 4500 Hz. Although the limit of hearing is around 16000 Hz, this upper limit is set due to factors such as quality of tone as well as harmonic considerations as it is proposed as a possibility that higher pitches than this can cause the music to seem amelodic. (R. M. Warren, 2008). Freberg (2010) adds that although a simplistic view on pitch is that it is associated with frequency, other factors, such as the context or intensity of a stimulus, can cause the pitch to vary (Freberg, 2010). As we have addressed pattern recognition in the chapter of perception and prediction, it is important to mention that pattern recognition is also a factor in the neural responses associated with pitch perception. The ability to process repetition of patterns and detect pattern deviations are important factors in both perception of rhythm as well as identifying melodic themes, not only in obvious deviations from the given pattern, but also in more subtle fluctuations in frequency and amplitude. To illustrate this, Winkler et al. (2009) argue how processing of low-level auditory stimuli can be put in context with pattern recognition and prediction:

> We review evidence showing that some processing of regularities occurs at quite low levels in the auditory system and suggest that auditory perceptual objects are mental constructs based on representations of temporal regularities which are inherently predictive, continuously generating expectations of the future behavior of sound sources. (I. Winkler et al., 2009, p. 532)

To process phonological units in speech, the brain quickly identifies regularities and patterns and merges these units into larger segments of speech. According to Hagoort and Levelt (2009), our brain uses less than 200 ms to identify a word, about 320 ms to categorize verbs and nouns morphologically, and 450 ms in total to complete the phonological encoding. This process is also entitled ‘unification’, as it instigates individual preverbal steps to unify into “a linear sequence of speech sounds” (P. Hagoort & Levelt, 2009). Friederici (2002) categorizes these steps of speech processing into three phases:

> Phase 1 (100-300 ms) represents the time window in which the initial syntactic structure is formed on the basis of information about the word category. During phase 2 (300-500 ms), lexical-semantic and morphosyntactic processes take place with the goal of thematic role assignment. During phase 3 (500-1000 ms), the different types of information are integrated. (Friederici, 2002, p. 79)

These processes cause the brain to merge these steps into more complex constructions. Winkler et al. (2009) support this notion by arguing that “representations of auditory regularities serve as perceptual objects” and elaborates this argument by explaining how “auditory objects are described in the brain by predictive rules linking together coherent sequences of sounds” (I. Winkler et al., 2009, p. 538). Friederici (2002) further says that these steps also includes smaller events which can influence language comprehension as “segmental phonemes and suprasegmental phonological information (prosody and pitch) as well as syntactic and semantic information must be accessed and coordinated within milliseconds” (Friederici, 2002, p. 78). This means that during approximately 450-500 ms of processing, the brain needs to accept all the individual linguistic elements presented in this neurological process, or respond quickly to conspicuous irregularities when its prediction fails on the base of pre-existing regularities and patterns. In accordance with these processes, Winkler et al. (2009) make the conclusion “that the auditory objects appearing in perception are based on detecting regular features within the acoustic signal” (I. Winkler et al., 2009).

2.2.4 Auditory deception

Following Tiippana (2014), due to the processes of our brain consequently trying to interpret and comprehend presented acoustic stimuli, the notion of deception in perception is an important factor when it comes to how we perceive speech as the combination auditory and visual signals might confuse the receiver and cause them to perceive an acoustic signal differently that if the stimuli had been presented without interference from visual stimuli. This effect shows how visual stimuli interfere with perception of acoustic stimuli is called *the McGurk effect*, and the most famous example is how a person will think that they hear the letter [d] if they see a film which presents a person who is articulating [g], but where the sound is dubbed so the acoustic stimuli is actually a [b]. The acoustic signal is perceived correctly when there is not interference from the incongruent visual speech stimuli, meaning that the subject simply closes their eyes and all of a sudden will have no problem recognizing the correct auditory stimuli of [d]. This perception deviation is called the fusion effect, as it merges visual and auditory information into the perception of a letter which deviates from both original stimuli (Tiippana, 2014). Warren (2008) pose another example of perceptive deception is what have been called illusory changes through *verbal transformations*. In short, when a subject is presented with the same monosyllabic word two times per second for a total of three minutes, the subject has reported back generally around six different verbal forms instead of the identical syllables presented in the sequence, and also the same illusory change several times, e.g. a sequence of the word ‘right’ is being heard has “ripe, right, white, white-light, right, right-light, ripe, right, ripe, bright-light, right, ripe, bright-light” (R. M. Warren, 2008, p. 205).

2.3 Tonal languages

According to the Merriam-Webster dictionary, a tonal language is defined as a language "in which variations in tone distinguish words or phrases of different meaning that otherwise would sound alike" (Tonal language, n.d.). Some argue that the definition of tone language and languages which uses pitch accents should be separated, but we will use the phrase as an umbrella term based on the dictionary definition. Norwegian is thus considered to be a tonal language, amongst other tonal languages mentioned in Ashby and Maidment (2005) as for example Chinese, Vietnamese, Thai, Zulu, and Navajo (Ashby & Maidment, 2005). When it comes to explaining the features of a tonal language, Baum and Pell (1999) write that tonal languages are languages where prosody serves as a phonemic function, i.e. that one can differentiate a pair of lexical items by their rising and falling tones. In these languages, prosody serves an important function as both the semantical meaning of the word as well as the emotion being conveyed can completely change depending on the tonal contrast, and that the prosodic components provide a basic linguistic function: "Because of the fundamental linguistic importance of tone in these language systems, one might predict a different pattern of prosodic control than is found in speakers of non-tone languages" (Baum & Pell, 1999, p. 585). As this study uses Norwegian words as the base for its behavioral experiment on prosody prediction, being familiar with the Norwegian language structure is important in order to effectively address qualitative and quantitative factors later in the thesis.

2.3.1 Basics principles of the Norwegian language

Store Norske Leksikon, estimate that the majority of Norwegians identify as non-SEN (Standard East Norwegian) users, and only about 15% identify as users of Nynorsk. One can assume that the vast variations of dialects could amount to just as many individual dialectal divisions as there are variations in dialectal identities amongst Norwegian speakers, if not more (Språk i Norge, October 23 2017). As the four main groups can be divided into sub-groups in the form of variations of said dialect, for example main-land trondsk, northern trondsk, coastal trondsk and high-land trondsk (Språk i Norge, October 23. 2017), these sub-groups may again have their own variations within, which can be influenced by factors such as culture, geography, tradition, and economy. An example of such regional variety is addressed by Wetterlin and Lahiri (2012) as they explain how small variations can affect accent distribution:

> The accent distribution of Standard East Norwegian is affected by both the morphology and phonology. Adding new morphemes can affect both accent and stress as well as cause resyllabification and assimilations. Thus, we believe that this interaction can best be accounted for when the morphology and phonology are seen as taking place one level after the other. (Wetterlin & Lahiri, 2012, p.296)

The term 'Nynorsk' is directly translated as 'New Norwegian', and is one of the two official written forms of Norwegian language in Norway, the other being 'Bokmål'. The (arguably) equivalent spoken variations to these two forms are spoken 'Nynorsk' and 'Standard East Norwegian' for 'Bokmål'. In short, where Bokmål was a result of "Norwegializing" the Danish language after Norway had deemed their independence, Nynorsk were founded on the base of reflecting spoken Norwegian dialects in an attempt to form a completely unbiased Norwegian language. The number of dialects spoken in Norway is not defined, but they say that one can roughly estimate the vast varieties in prosodic changes by looking at the four main groups of Norwegian language; East Norwegian, West Norwegian, Trondersk, and Northern Norwegian, which all have dialectal sub-categories (Lundskær-Nielsen, Barnes, & Lindskog, 2005), supporting the categorization found in Store Norske Leksikon.

Prosodically, the Norwegian language possess not only stress as a basic linguistic function, but also tonal accent. Wetterlin and Lahiri (2012) point out that although the North Germanic languages share a common ancestry, they differ when it comes to the tonal prosody of present day. They use Central Swedish in contrast to Standard East Norwegian as an example of how tonal prosody has changed; where in Central Swedish, the pattern of tonal prosody has been generalized, "Standard East Norwegian compounds still reflect the word internal properties of lexical tone accents" (Wetterlin & Lahiri, 2012, p.279). The Norwegian dialects are also categorized by whether they are classified as *high tone dialect* or *low tone dialect*. The pronunciation difference (within the same word) when it comes to the two classes are as follows:

> High tone: Tone is high in the beginning of the word, then falls towards the end.
> Low tone: Tone is low in the beginning of the word, then rises towards the end.

Store Norske Leksikon use the two syllable words /boka/ and /sola/ as examples of words that have a different tone depending on whether their associated dialect is low tone or high tone. It is important to notice that the aspect of pronunciation difference between low tone and high tone do not change the lexical form nor the semantic meaning of the word. The words ‘boka’ and ‘sola’, when pronounced correctly in relation to their associated dialect, will have the same meaning in both variations (Dialektar i Østfold, February 23 2016). In comparison, when changing the starting tone when pronouncing the word /bøner/, as a change in prosody causes the word to have completely different semantic meanings, such as /farmers/, /beans/, and /prayers/ just by small adjustment to the word pronunciation’s prosodic features.

According to Lundskær- Nielsen et al. (2005), South East Norwegian, in this study called Standard East Norwegian, contains nine vowel phonemes, which all can be either long or short depending on their linguistic environment, and five long diphthongs. In addition to this, Standard East Norwegian contain around twenty consonant phonemes, a number that, like diphthongs, can differ due to regional variations.

- Vowels: /i//y/ /e/ /ø/ /æ/ /ʉ/ /a/ /o/ /u/
- Diphthongs: /ei/ /øy/ /æʉ/ /ai/ /oi/
- Consonants: /p/ /b/ /m/ /f/ /v/ /t/ /d/ /s/ /l/ /r/ /n/ /ʈ/ /ɖ/ /ʂ/ /ɭ/ /ɳ/ /k/ /g/ /ç/ /j/ /ŋ/ /h/

They further state that as there are many regional varieties in Norwegian, the realization of individual diphthongs and vowels may vary, and one can encounter dialects where diphthongs seize being diphthongs depending on regional pronunciation. Regarding tone, all monosyllabic words have tone 1 regardless of regional variation. In addition to this, they state that a syllable needs to have stress in order to have tone, and that a polysyllabic word do not contain any lexically specified morphemes, they result in a default tone 2 (Wetterlin & Lahiri, 2012).

2.3.2 Perception and prosody in tonal languages

As prosody plays an important, and often crucial, part when it comes to the basic linguistic functions of a tonal language, users of tonal languages are dependent on their brain quickly being able to quickly detect, and process, these (often subtle) tonal differences when we are communicating. If the brain fails to process the intended word correctly, the speaker's utterance might be unintelligible for the recipient. An example of this in Norwegian can be when the intended utterance being "I found a box of expired beans in my cabinet yesterday" is interpreted as "I found a box of expired farmers in my cabinet yesterday". Most times, the recipient will understand the speaker's intended meaning by the context of the given utterance, but in some cases the tonal difference can be crucial; the words 'deig' (dough) and 'deg' (you) are both pronounced /dei/ in Standard East Norwegian, which can cause confusion if you are e.g. working at a bakery, or being on a first date:

:

1. Jeg hater å jobbe med (I hate working with) /dei/
    a. I hate working with dough
    b. I hate working with you

2. Jeg elsker (I love) /dei/
    a. I love dough
    b. I love you

To make this even more complicated, /dei/ can also mean 'them' in some Norwegian dialects. This means that if you ever work with a baker from the west of Norway who has spent a good amount of years in Oslo, their mixed dialect might cause some confusion as their prosodic as well as lexical patterns might change in the midst of a sentence.

Prosody can also change the lexical form in Norwegian words, but not where it is not applicable according to the prosodic patterns of the Norwegian language. One example can be how the words 'brygge' and 'brygge' have different lexical forms; one being a noun meaning a 'pier', and the other being a verb meaning to 'brew'. If the tone of these words was to be changed from low tone to high tone in an infelicitous context, the processes of native Norwegian's brain would most likely detect such a prosodic error, but as the deviation would not influence the meaning of the word, the listener would not be confused about the speaker's intended meaning.

2.3.3 Norwegian as tonal language

As addressed earlier in the chapter, Ashby and Maidment (2005) explains how lexical tone languages use pitch patterns to distinguish between words that are otherwise identical in pronunciation. This means that one need to pay attention to witch pitch pattern is applied to a word, as the wrong pitch pattern could change the meaning of the word completely. Where Norwegian has two tone patterns, Cantonese Chinese have six. When a language contains some tones that claim moving pitch patterns, meaning that the difference in the movement of the pitch patterns distinguish one word from another, it can also be known as *contour tone language.* A contrast to contour tone languages are level tone languages, also known as register tone languages, in which a one distinguish a word's tones by how their pitch level relates to each other, with examples being African languages such as Bafang and Yoruba (Ashby & Maidment, 2005).

As previously mentioned, spoken words are not just a string of verbal segments, but rather a complex construct of linguistic elements. Rhythmic stress, lexical tone, and intonation are features of speech that are called prosodic features or suprasegmentals. Baum and Pell (1999) refer to a study by Lehiste (Lehiste, 1970) to address the influence of prosody in language:

> As is well known, prosody serves a variety of functions in language processing, from the conveyance of the speaker's emotions to the phonemic use of tone to differentiate lexical items in certain languages. Regardless of function, the same three acoustic parameters serve as primary prosodic attributes: fundamental frequency (F0), duration and amplitude. (Baum & Pell, 1999, pp. 581, 582)

Among the other prosodic functions in language, Lundskær-Nielsen et al. (2005) adds that stress is composed of several features, among them pitch variation, loudness, length, loudness, and intensity. Stress affects syllables, causing the stressed syllable to be more prominent than the unstressed counterparts (Lundskær-Nielsen et al., 2005). Ashby and Maidment (2005) support this notion by stating that stress is a feature used to influence entire syllables rather than shorter phonological segments, and that prosody is used to put emphasis on the desired syllable in order to make it more audible. It can thus can also singlehandedly change a word's lexical class in English, just as it can in Norwegian (Ashby & Maidment, 2005).

According to Ashby and Maidment (2005), three factors that influence a stressed syllable are pitch, loudness, and length. They use the words 'written' [ˈɹɪtən] and 'return' [ɹɪˈtɜːn] to show

how change in stress transforms a word completely just by adding length and emphasis to a phoneme, in this case the last vowel of the word(s). If the stress were interchanged, the words would be hard to distinguish from each other, especially if the speaker also changed the length of the last vowel. Lastly, the pitch of the words is also different, as it in ’written’ the first syllable is high and falling, while in ‘return’ the same pattern occurs in the second syllable. To not be confused with the case of homonyms in Norwegian, see the example of the verb ‘brygge’ and the noun ‘brygge’ mentioned earlier in this chapter, where the pronunciation is the same for both words regardless of their lexical class, while the verb ‘håpe’ and the noun ‘håpet’ are both pronounced /håpe/, but with a difference in stress.

2.3.4 Pitch in non-tonal languages

Koelsch (2013) states that the element of pitch is a fundamental element when it comes to both music and speech, especially when it comes to decoding both grammatical and lexical meaning in tonal languages. However, pitch is also essential in non-tonal languages as the use of intonation in suprasegmental variations are used when conveying meaning (Koelsch, 2013). Ashby and Maidment (2005) report that even though languages like e.g. English and French are not lexical tone languages, all languages use the variation of pitch in order to communicate the desired meaning (Ashby & Maidment, 2005). An example of pitch variation in a non-tonal language is how the pitch changes in the same lexical utterance depending on the utterance being a question or a statement:

1. The blue one
2. The blue one?

Here, the pitch variation does not constitute for a change in the lexical meaning of any of the words involved, as each word still keeps it’s lexical and semantic meaning, but it does influence how the utterance is interpreted in the form of a question or a statement. Such specific pitch variation where the meaning of the utterance is defined by how it is said without changing the meaning of the individual words is known as *intonational phrases*(Ashby & Maidment, 2005). Pitch variations in intonational phrases are especially important when it comes to ambiguous phrases, for example the following phrase:

1. Those who ran quickly got reprimanded
    a. Those who left | quickly got reprimanded
    b. Those who left quickly | got reprimanded

According to Ashby and Maidment (2005), one might argue that pitch variations in intonational phrases have the function of being ‘prosodic commas’ in phrases, as one would use a comma to distinguish the separate entities in the phrase ‘the small cats and dogs’, which could be interpreted either as ‘the small cats as well as the small dogs’, or as only a specific selection of the cats but all the dogs, no matter their size. Another aspect of intonation addressed by Ashby and Maidment (2005) is the notion of *key,* as it affects intonation phrases as a whole. *Key* is used to signal whether we are finishing up our utterance, or have more to add. An example of this can be either in isolated utterance, for example if you are reciting a shopping list where the pitch is consistent throughout the majority of the list until the last item, when the pitch becomes lower in order to signal the end of the recitation. Another example of key are how newsreaders signal the end of one story as they are getting ready to present the next, cabin crew informing about events regarding your flight, or if you are adding a “verbal parenthesis” to a intonational phrase.

I saw her and her sister |the one who works in marketing | at the airport.

A word’s prosody in an intonational phrase can be contextually influenced through what is called *intonational tone*. The use of intonational tone in correlation with prediction is an important prosodic element, as the pattern that we use for predicting the completeness of an utterance is based on how familiar we are with a language or an accent (Ashby & Maidment, 2005). This means that we use our intuitive knowledge of pitch patterns to predict when an utterance is complete, and we will react if our assumption is inaccurate. Intonational tone is also used to signal that new information is added to a conversation, for example when parts of the conversation are already stated and accepted by all parts, but additional information is included:

1. When you get to my house *look under the mat.*

When a speaker wants to convey an attitude, they might apply what Ashby and Maidment (2005) defines as a *nuclear tone.* A nuclear tone starts at the nucleus of an intonational phrase and continues to the end, and is used to convey attitude in a speech phrase. As we saw an example of earlier through sarcasm in ‘You look nice today’, nuclear tone can also define whether an utterance is deemed sympathetic or unsympathetic:

1. *Stop crying*
    - *a.* If uttered with a high tone at the first word, the speaker will sound sympathetic
    - *b.* If uttered with a low tone first, and then falling on the second word, the utterance would be deemed harsh, and rather a command than a soothing statement.

Ashby and Maidment (2005) say that although some of these signals are not directly related to our speech features, some are connected to how we are speaking; loudness, tempo, pause frequency, type of phonation, and pitch range are all signals known as *paralinguistic features,* and all affect how our we convey our intended meaning, as well how what we say are perceived by others. An example here can be how the words uttered clashes with paralinguistic features, such as someone saying that they are ‘really exited’ in a monotone slow pitch signaling quite the contrary, or a person saying they are extremely happy while yelling and having a wide pitch range which would signal anger (Ashby & Maidment, 2005).

To illustrate the signals that can be found in English (non-tonal language) and Norwegian (tonal language) it could be interesting to compare the amount of possible prosodic variations within a phrase. An example for five different meanings in an English phrase could be the following:

1. You look nice today.

When written, variations of this sentence look like they mean the exact same thing; that in fact the person addressed look nice today. However, when this sentence is uttered orally, difference in prosody can alter the meaning completely:

1. <u>You</u> look nice today.
    - a. If standing in a group, you are the only one who looks nice today.
2. You <u>look</u> nice today.
    - a. You look nice, but you smell like rotten trout.

3. You look nice today.
    a. Either an emphasis of the niceness (perhaps you looked mean yesterday), or when put additional stress on, you can be assured that nice is an understatement of how you are looking.
4. You look nice today.
    a. Because yesterday, you looked like a trash can.
5. **You** look nice today.
    a. A different variety of 1), as this one can be perceived as sarcasm when the stress is uttered in a certain way.

Then we look at the (inflected) Norwegian noun ‘bønner’ (beans) as an example. When written, the sentence translates to the simple fact that the sender does not want beans. Extracting five different meanings for the phrase without regarding the prosodic variations of tone (tone 1 and tone 2) in /bøner/ could look like something like this:

1. Jeg vil ikke ha /bøner/
    a. I don’t want beans
        i. I don’t want beans, I want peas.
        ii. I don’t want beans, but I guess I’ll have to eat them.
        iii. It was not me who wanted beans, it was someone else.
        iv. Are you insane, I never asked for this.
        v. Is the right response to you oogling me that I should say no to the beans?

If the prosodic variations of tone were to be an element in these utterances, regardless of the word’s context, we would be presented with more variations than in the English example, for example that we do not want any farmers attending our birthday party. By adding paralinguistic signals, the range of expressive possibilities for either language could seem endless.

2.4 Music and speech

The acoustic cues of music and speech have more in common than just soundwaves. In *Language, Music, and the* Brain (2013), Klaus R. Scherer refers to a study by Patel, Scherer, Bjorkner, and Sundberg ((Patel, Scherer, Bjorkner, & Sundberg, 2011), where ten different actors produced the vowel /a/ based on five different emotions. The researchers then extracted three components of acoustic variations due to emotion; voicing frequency, tension, and perturbation. Scherer addresses this study further regarding how they found that the emotions consisted of a specific combination acoustic parameters which reflected a distinct blend of physiological voice control parameters, an analysis Scherer argues similarly can be applied to music as variety of prosodic cues are used similarly when it comes to conveying emotion in music and speech. In what he refers to as *cross-modal patterns of acoustic cues for discrete emotions,* he lists different emotions with their acoustic counterparts in both vocal expressions and music performance*:*

- Anger
    - Fast speech or rate tempo
    - High voice intensity or sound level
    - Much variability in voice intensity or sound level
    - Much high-frequency energy
    - High F0/pitch level
    - Much F0/pitch variability
    - Rising F0/pitch contour
    - Fast voice onsets or tone attacks
    - Microstructural irregularity
- Fear
    - Fast speech or rate tempo
    - Low voice intensity or sound level (except in panic fear)
    - Much variability in voice intensity or sound level
    - Little high-frequency energy
    - High F0/pitch level
    - Little F0/pitch variability

    - Rising F0/pitch contour
    - A lot of microstructural irregularity
- Happiness
    - Fast speech rate or tempo
    - Medium high voice intensity or sound level
    - Medium high-frequency energy
    - High F0/pitch level
    - Much F0/pitch variability
    - Rising F0/pitch contour
    - Fast voice onsets or tone attacks
    - Very little microstructural regularity
- Sadness
    - Slow speech rate or tempo
    - Low voice intensity or sound level
    - Little variability in voice intensity or sound level
    - Little high-frequency energy
    - Low F0/pitch level
    - Little F0/pitch variability
    - Falling F0/pitch contours
    - Slow voice onsets or tone attacks
    - Microstructural regularity (Scherer, 2013, p. 125).

We use these acoustic variants of auditory stimuli to predict upcoming events based on our pre-existing cognitive patterns. An example can be in the form of a horror movie, where acoustic elements combined to create so-called ‘eerie’ music is used to create suspenseful auditory stimuli for the viewer (listener). Paired with corresponding visual stimuli, the brain retrieves information about what one can expect at the end based on pre-existing cognitive patterns. In this case, one would usually expect something frightening to happen and the acoustic cues for fear as described in Scherer’s categorization would apply to any speech sounds produced. However, if the prediction is wrong and the expected frightening event is absent at the end of the auditory and visual climax, one will return (presumably) to where one was before the eerie stimuli was introduced. A lot of films now utilize this retracted state after a failed prediction, by presenting the frightening event when the viewer least expects it, often during stimuli comprised of acoustic cues that are associated with happiness and safe surroundings.

As we have addressed how screaming functions as an instinctive reaction, Scherer (2013) says that humans have kept some form of primal, non-linguistic vocalizations that are similar to many mammal species; spontaneous vocal reactions to submission, fear, anger and aggression in animals can be compared to what he calls nonlinguistic human affect vocalizations, or interjections, where utterances similar to ‘oh’, ‘ai’, and ‘ii’ can remind us of animal vocalizations. When these vocalization reactions extend to become sociocultural norms, Scherer refers to Ekman and Friesen’s postulated requirements for these reactions termed vocal *affect emblems* and how they evolve from spontaneous expression of emotions into referential meanings:

1. Existence of a verbal “translation”
2. Social agreement on its meaning
3. Intentional use in interaction
4. Mutual understanding of the meaning
5. Sender assumes responsibility in emblem production (Scherer, 2013, p. 133).

Scherer says that as these bursts of affects usually consist of repeated or single sounds, they have later evolved into more complex sound structures both when it comes to syntactic aspects and intonation patterns due to melodic resemblance. He suggests that due to the pragmatics of this emotional signaling, there is a possibility that singing predates speech (Scherer, 2013), which supports the notion of shared neurological features for speech and music. D. Robert Ladd (2013) also draws parallels for phonetics in both language and music, and says that phonetics in language can be argued to have its musical counterparts in elements such as musical pitch and melodic structure by saying that “both music an language are evolutionary built on the ability to assemble elements of sound into complex patterns, and that what is unique about human language is that this elaborate combinatorics system incorporates compositional referential semantics” (Ladd, 2013, p. 287). According to Warren (2008), the ability to differentiate between a variety of combination of units is a significant factor in both speech comprehension and music appreciation. As both speech and music is considered to have similar rates when it comes to syllable and melodic note processing (about 150 ms for both), he says that it has been long assumed that the listener needs to be familiar with the tonal pattern as well as the tones’ frequency and duration when it comes to recognizing and identifying a specific melody:

> It has been assumed for some time that the ability to distinguish between different arrangements of the same sounds requires that listeners be able to identify the order of components. However, recent evidence indicates that permuted orders of speech sounds or tones, and permuted orders of unrelated sounds (such as hisses, tones, and buzzes) can be distinguished without the ability to identify the orders within the sequences (or even the component sounds themselves). (R. M. Warren, 2008, p. 149)

2.4.1 Comparing structures of music and language

Stefan Koelsch (2013) refers to how musical elements are sequenced into regularity-based arrangements as *musical syntax,* but at the same time advises one to not see this system at unitary as syntactic organization comes in several different categories. In his discussion about how the cognitive processes involved in musical syntax processing, he lists seven sub-processes:

- *Element extraction*, where small elements are extracted from the continuous flow of auditory information. In music, this can be chords and tones, while the equivalents in language are suggested to be words and phonemes.
- *Knowledge-free structuring* is when you do not need to possess long-term knowledge of a structure in order to be able to detect and react to elements that seem out of place. An example here is how the brain can detect tones that are off-key after a single key music passage has been established.
- *Musical expectancy formation.* In contrast to the previous sub-process of *knowledge-free structuring,* this process is based on regular patterns that are stored in a format of long-term memory. An example of this can be how one expects a certain tone to be represented next in a tonal interval due to long-term memory patterns.
- *Structure building* is a term which suggests that tonal music should be viewed as a hierarchical structure based on auditory working memory, similarly to how tree diagrams are used to depict syntactic structures.
- *Structural reanalysis and revision* happens when a hierarchical structure needs to be revised. To illustrate, Koelsch used 'garden path sentences' to display an example of how issues of ambiguity can cause grounds for revision. These sentences are

grammatically correct, but a reader will most likely deem them incorrect. The most used example of a garden path sentence is Thomas Bever's "The horse raced past the barn fell".

- *Syntactic integration.* As a sentence consists of various syntactic features, tonal music uses elements such as meter, harmony, and melody as its constituents in order to create a coherent structural representation.
- *Large scale structuring* is the final sub-process of musical syntax processing, and while the previously mentioned sub-processes have been relevant for phrase structure, large scale structuring concerns with how music forms are structured outside of phrasing. An example here can be how a song starts with two verses in a row, then goes into a chorus, then a verse, then a bridge, before ending in a chorus (Koelsch, 2013, pp. 142-145).

2.5 Neural correlates of prosodic processing

To understand how prosody affects language perception, we need to understand how prosody is processed in the brain. As we have mentioned in chapter 2.2, the lateralization of cognitive function in the brain means that a given function is located in one hemisphere or the other. Freberg (2010) states that when it comes to speech perception, the majority of complex verbal language processing is argued to be located in the left hemisphere. When we are exposed to sound, both hemispheres process the sound, but the contralateral hemisphere will complete the listening task more quickly than the other. When it comes to the matter of prosody, Freberg refers to a study by Charbonneau, Scherzer, Aspirot and Cohen (Charbonneau, Scherzer, Aspirot, & Cohen, 2003) where fMRI (functional magnetic resonance imaging) showed that the right hemisphere participates in evaluation emotional tone in spoken language, but is far less adequate in processing prosody compared to the left hemisphere. When it comes to the location of pitch perception, Freberg refers to a study by Schlaug, Jancke, Huang, and Steinmetz (Schlaug, Jancke, Huang, & Steinmetz, 1995) where results suggested that perfect pitch perception for musicians is mediated in the left hemisphere (Freberg, 2010). Hickok and Poeppel (2000) offer a more elaborate description on how the processes of speech perception are mediated:

> “From this point, however, we will argue that there are at least two distinct pathways that participate in speech perception in a task-dependent manner, and that they are more strongly lateralized to the left hemisphere. The first is a ventral pathway, which probably involves cortex in the vicinity of the temporal-parietal-occipital junction. This pathway appears to be important for interfacing sound-based representations of speech with widely distributed conceptual representations, and therefore is involved in tasks that require explicit access to certain sub-lexical speech segments.” (Hickok & Poeppel, 2000, p. 131)

As we are constantly exposed to various types of acoustic stimuli, our brain needs to be able to perceive and process every single input in an effective manner. If the brain was to detect and react to every single prosodic irregularity it encountered over time, the energy output and capacity would suffer as it would constantly be alerting us to trivial errors that would not have a significant influence on language processing and comprehension. Winkler et al. (2009) say that a lot of the sounds around us are *ambient*, meaning that they show up as continual fluctuating energy on a waveform, and that speech sounds have a regular pattern of soundwaves in a waveform. If we encounter the same irregularities over time, the brain will start to accept the irregularities as correct input: “The auditory system continuously searches for regularities within the acoustic signal. Primitive regularities may be encoded by neurons adapting their response to specific sounds” (I. Winkler et al., 2009, p. 532). This means that when we are exposed to a language over time, we learn the language’s acoustic signals and our brain start to accept them as regularities. This also goes for encountering new acoustic signals; our brain will eventually adapt and stop regarding these as irregularities and instead accept them as regularities within our regular pattern. However, Lundskær-Nielsen et al. (2005) argue that as we get older, we find ourselves being so accustomed to the specific sounds of the language(s) we are proficient in that it can be hard, and sometimes impossible, to distinguish between speech sounds that we are not familiar with. An example is how a native English speaker can find it challenging to distinguish between the vowel [ɪ] from the vowel [e], as they are used to the difference between [ɪ] and [ɛ] (Lundskær-Nielsen et al., 2005).

2.5.1 PET and fMRI studies on prosodic processing

Earlier, the knowledge of prosodic processing in the brain where scarce: “despite its importance in communication, the neural systems responsible for the production and comprehension of prosody remain largely unspecified” (Baum & Pell, 1999, p. 582). In the later years, however, PET and fMRI studies have further contributed to our understanding of prosodic processing. Friederici (2002) addresses some of these studies on prosodic processes in her paper *Towards a neural basis of auditory sentence processing:*

> The functional neuroanatomy of prosodic processes has been specified in recent studies using PET and fMRI. At the segmental level, pitch discrimination in speech syllables correlates with an increased activation in the right prefrontal cortex [Zatorre, Mondor, and Evans (1999)]. Violations of pitch for lexical elements in a tonal language, such as Thai, results in modulation of activation in the left frontal operculum adjacent to Broca’s area [Gandour et al. (2000)] (Friederici, 2002, p. 82)

> A recent fMRI experiment that systematically varied the presence of pitch information (normal intonation versus synthesized, flattened intonation) and of syntactic information (normal speech versus synthesized, delexicalized speech) at the sentential level identified modulations in activity of the right peri-sylvian cortex. In particular, the right superior temporal region and the fronto-opercular cortex were identified as regions that support the processing of suprasegmental information [Meyer et al. (in press)](Friederici, 2002, p. 83)

Although Friederici (2002) says that the neuroanatomical data that is available are suggestive, she concludes the section on prosodic processes by stating that “Overall, although limited, the data available indicate that a temporo-frontal network that is predominantly within the right hemisphere supports prosodic processes and that prosodic information can influence syntactic processes” (Friederici, 2002, p. 83).

2.6 The current study

As tonal languages are sensitive to prosodic variations, users of these languages depend on the brain quickly being able to process, detect, and differentiate between, small acoustic changes that could change the meaning of the word. Our assumption was that native tonal language users would show sensitivity to prosodic changes in both rhythm and spoken words. Our aim was to assess whether low-level auditory processes, that are not specifically speech perception processes, are actually recruited during perception of tone (mis)placement in Norwegian words.

2.7 Electroencephalography and event related potentials

Electroencephalography is a technique for monitoring the surface electrical activity of the brain (Crossman & Neary, 2005) and is a method widely used in both the clinical setting, e.g. for diagnosing epilepsy and other brain dysfunctions, as well as in research. Warren (2008) explains how electroencephalography is used to record electrical effects that follow neural activity caused by auditory stimuli by placing several sensors on a subject’s scalp in order to read the event related potentials (ERPs). Hence, ERP represent the EEG signals cohering to a specific event. In the case of this study, the sensory event was in the form of auditory stimuli.

The first elicited responses in humans have been observed within 10 seconds of the presented stimuli, the ‘middle latency’ responses are found between 10 and 40 ms, and the ‘late latency’ between 50-300 ms, each latency indicating different levels of processing (R. M. Warren, 2008). Hence, as we will further elaborate later, the parts of the ERP-signal we are interested in takes place in the ‘late latency’ period.

3. Method

In order to be able to test our hypothesis, two experiments were performed where both experiments were approved by the Norwegian Centre for Research Data[1] (NSD). The first was an experiment based on electroencephalography (EEG, while the second experiment was a behavioral experiment using offline responses to measure response time and accuracy. Participants were informed of the purpose of the trial and had to give their signed informed consent before being enrolled.

The experiments will be addressed individually in the upcoming sections.

3.1 Experiment 1: EEG

3.1.1 Participants

The EEG experiment was originally tested on a group of 17 participants (n=17, M= 6 and F=11, mean age 25.9), where none of the participants identified themselves as having musical training. The EEG experiment took place in the Language Acquisition and Language Processing Lab at NTNU Dragvoll, and was conducted by a PhD Candidate. The subjects, including the participants of the main group (n=11, M=4 and F=7, age range: 21-40, age mean: 28.45), were asked to watch an excerpt from the subtitled silent movie “The Artist” (Hazanavicius, Dujardin, & Bejo, 2011) while ignoring sounds coming from the loudspeaker placed in the room. The sounds emitted were 80 dB. They were instructed to try to minimize 16 movements, among them blinking and eye movements to try to reduce noise possibly affecting the EEG measurement and hence the ERP-data.

---

[1] http://www.nsd.uib.no/nsd/english/index.html

3.1.2 Auditory stimuli

The auditory stimuli consisted of three recording sessions of sine-waves equitone[2] rhythms of five beats, where the order of sessions or conditions were counterbalanced across the participants. The rhythms were transposed randomly at three fundamental frequencies: 315 Hz, 397 Hz and 500 Hz. The sessions consisted of three variants of entropy[3]; high-entropy, E=2, low-entropy, E=1, and control, E=0. Each of the sessions were about 15 minutes long, and consisted of four blocks with 100 stimuli in each block, totaling 1200 stimuli per session. There were three types of stimuli presented in each session; one standard stimuli type, and two deviants. Every block consisted of 80% standard stimuli which were 1300 ms in duration with 700 ms interstimulus interval, or 1600 ms with 900 ms of interstimulus interval. In 10% of the trials, early deviants which altered the rhythmic contour (in this study referred to as *contour deviant*) were presented when the fourth beat changed 300 ms earlier than in the standard stimuli. In the remaining 10%, a timing deviant, meaning a deviant that preserved the original rhythmic contour, was produced as the fourth beat was anticipated at 100 ms. All of the deviants as well as the standards were presented in a pseudorandom order, where the only constraint was that two deviants could not occur in a row.

---

[2] Meaning that they are pure tones. In depth, Povel and Okkerman define equitone sequences as “sequences of tones that are identical in all respects: frequency, spectral composition, intensity, and duration. The only parameter varied in these sequences is the time-interval between tones” (Povel & Okkerman, 1981, p. 565).

[3] The amount of differences in probabilities

3.1.3 EEG-recording and processing

The four scalp quadrants were posterior, central left, central right, and anterior, with nine channels in each. 0-320 ms with four windows of 80 ms were selected according to the characteristics of timing for the MMN, and again after examining the ERP data. The EEG data was recorded by using an EEG actiCap with 64 Ag-AgC1 electrodes, where the data was recorded at a sampling rate of 1000 Hz. The electrodes were referenced to FCz (fronto, central, and midline) and arranged in the international 10-20 system (see fig 3.1.3.1). Finally, the EEG data was offline downsampled from 1000 Hz to 500 Hz.

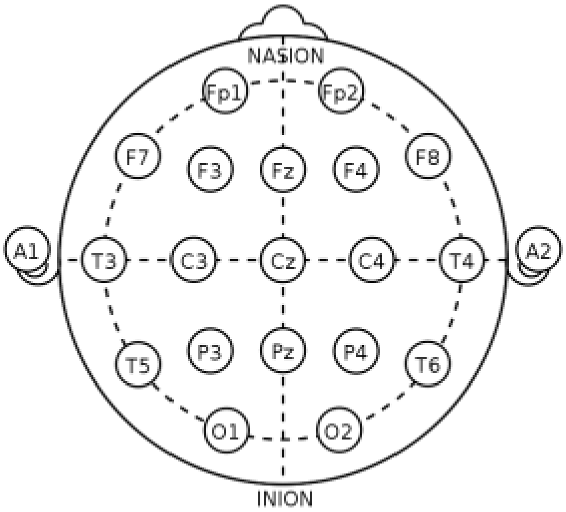

Figure 3.1. Showing the electrode placing according to the 10-20 system. https://upload.wikimedia.org/wikipedia/commons/7/70/21_electrodes_of_International_10-20_system_for_EEG.svg. Public domain.

A bandpass-filter using Matlab toolboxes ERPlab and EEGlab was regulated between 0.1 and 30 Hz with a roll-off of 12db per octave, before re-referencing the EEG offline to the mastoid channels average. The EEG was then segmented into epochs of 700 ms relative to the fourth beat onset, from -100 to 600 ms. Any epoch with EEG responses surpassing ±70 µV in a moving window of 200 ms were excluded from the average due to being regarded as artifacts, meaning that they are signals that are mainly of ocular, muscular, and mechanical origin instead of cerebral. The schematic layout of the EEG sampling in relation to the rhythm stimuli is presented in figure 3.2.

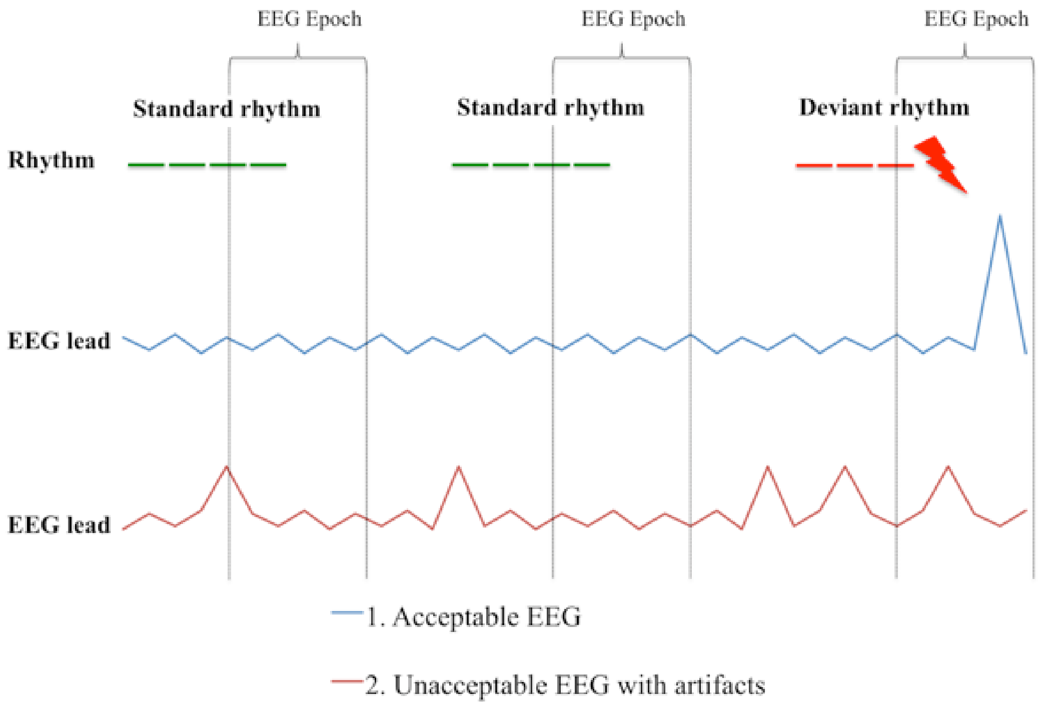

Figure 3.2 Schematic illustration of the layout of the EEG experiment with a thought acceptable and unacceptable EEG lead in relation to standard and deviant rhythms. Sampling of EEG epochs of 700milliseconds duration being visualized in relation to standard and deviant rhythms.

3.1.4 ERP data analysis

The epochs that were not excluded were averaged separately according for the standard and deviant condition, then subtracting standard for deviant ERPs were used to compute deviant-standard difference waves for each participant as such a process isolates the component of, among other effects, mismatch negativity which is a fronto-centrally distributed negative component which peaks between 100 and 200 ms when an auditory mismatch is detected. The presence of mismatch negativity (MMN) was first tested using a four-way ANOVA where the mean amplitude for MMN was used as a dependent variable for each of the three conditions. Four factors, including stimulus type in the form of standard or deviant, four levels of temporal window, four levels of quadrant, and 9x4 levels of Electrode were used in this testing, and The

Greenhouse-Geisser correction[4] was applied when applicable, while in post-hoc analyzes Wilcoxon signed-rank tests[5] were applied.

3.2 Experiment 2: Behavioral

3.2.1 Participants

The behavioral experiment was conducted on two different groups: The main group, which consisted of 11 of the individuals that had previously been tested in the EEG-experiment (n=11, M=4 and F=7, age range: 21-40, age mean: 28.45), and a control group (n=20, M=10 and F=10, age range 18-40). The remaining six individuals from the EEG-experiment were lost to follow-up. The groups consisted of both right-, and left-handed participants, no one had auditory or vision problems other than normal use of glasses/contact lenses, and all the participants involved in the analysis were Norwegian native speakers and used Norwegian as their main language of daily communication. A self-report questionnaire was used to assess language background and proficiency.

3.2.2 Auditory stimuli

The auditory stimuli were collected randomly from two-, three-, and four-syllable Norwegian words used in daily events and media outlets. Of the original 180 collected words, 144 were chosen due to syllable category and tone category. In sum, there were 48 disyllabic words, 48 words with 3 syllables, and 48 with 4 syllables. Each of the syllabic categories were then again divided into tone 1 and tone 2, with 24 words of each tone for each of the three syllable categories. Words that were considered Norwegian homonyms were excluded from the test, while arbitrary compound words were included. To ensure that the required specifications for the experiment were implemented, the words were controlled by an Associate Professor at the Norwegian section at Department of Language and Literature The auditory stimuli were then recorded by a female Norwegian native speaker who had Standard East Norwegian as her

---

[4] A correction used to adjust the degrees of freedom in the ANOVA test in order to produce a more accurate significance (*p*) value

[5] A non-parametric, paired difference test that is used to compare two samples

dialect. The recordings were made in the Phonetics Laboratory at the Department of Language and Literature at NTNU Dragvoll.

The words were then segmented into individual files by using the program Audacity. For manipulation of the words, we used the program PRAAT to change the prosodic features of the chosen words into different classifications (see table 4.1), counting 18 variants of Norwegian words. Each of the syllable sets had their own sub-set of variants in the form of Tone 1 and Tone 2, and each of those sub-sets were then again divided into Acceptable (the word's original/non-manipulated form), Not Acceptable, and Borderline Acceptable.

Table 3.1 The different variations of the words. 2S = two syllables. 3S = three syllables. 4S = four syllables. ACC = acceptable. UNACC = unacceptable. BA = borderline acceptable. T1 = Tone 1. T2 = Tone 2.

2 syllable	**3 syllable**	**4 syllable**
2S_ACC_T1	3S_ACC_T1	4S_ACC_T1
2S_UNACC_T1	3S_UNACC_T1	4S_UNACC_T1
2S_BA_T1	3S_BA_T1	4S_BA_T1
2S_ACC_T2	3S_ACC_T2	4S_ACC_T2
2S_UNACC_T2	3S_UNACC_T2	4S_UNACC_T2
2S_BA_T2	3S_BA_T2	4S_BA_T2

A template for each of the individual categories were manually drawn by extracting the pitch tier from the completed manipulation by manipulating one or more of the five main pitch points extracted in the program: The original word was manipulated by using the "manipulate (to manipulation)" feature, before going to "view and edit", then "stylize pitch" through to "sound &pulses (pitch and duration)", before finally using the "extract pitch tier" function. The original pitch tier for each word was then manually replaced by their appropriate constituent. The exact same manipulation template was used for all the words within the same syllable count and manipulation, e.g. the same template for "unacceptable" was used for all the 2 syllable words and the same template for "borderline" was used for all the two-syllable words. This procedure was then repeated with the assigned templates for the three and four syllable words.

See the following figures 3.3 to 3.5 for the manipulated variations for the three-syllable word 'oppmuntre'.

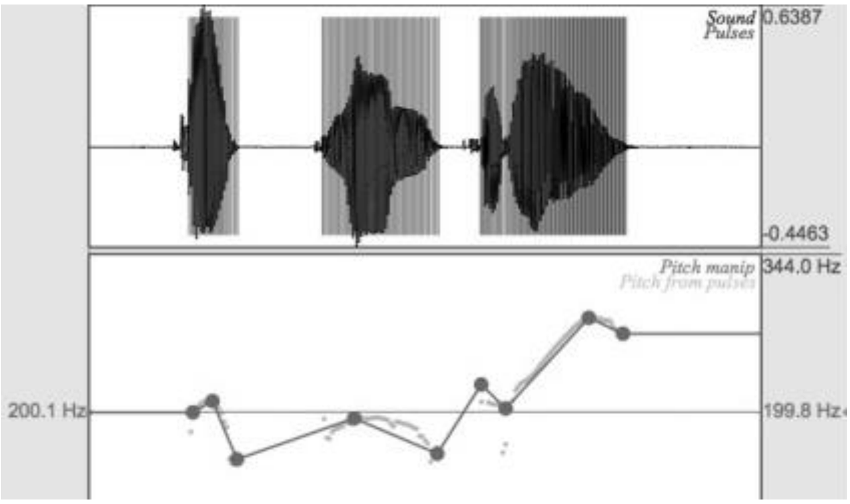

Figure 3.3. Original, 'acceptable' pitch for the three-syllable word 'oppmuntre' visualized in the program PRAAT.

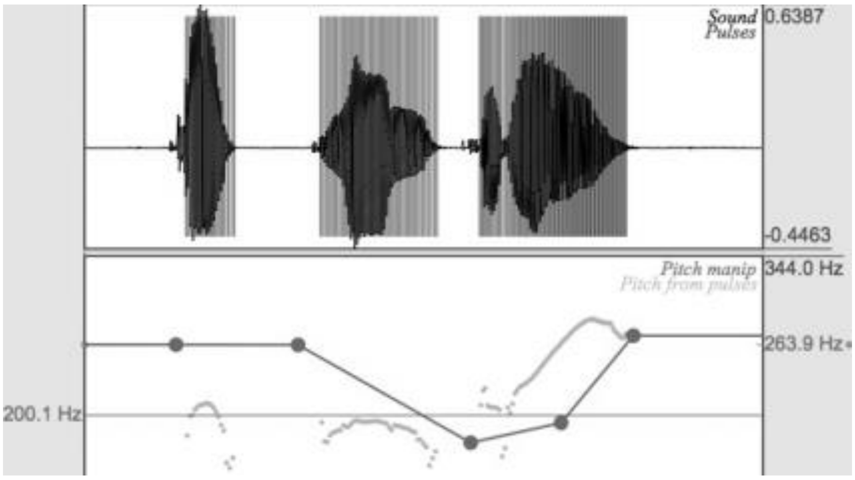

Figure 3.4 Manipulated pitch to 'borderline acceptable' for the three-syllable word 'oppmuntre' visualized in the program PRAAT.

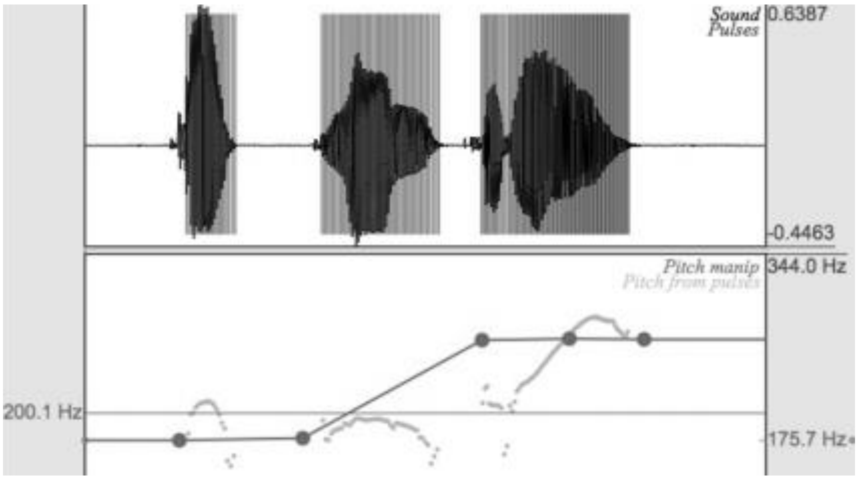

Figure 3.5 Manipulated pitch to 'unacceptable' for the three-syllable word 'oppmuntre' visualized in the program PRAAT

3.2.3 Behavioral testing

We used E-prime version 2 on a Dell M6800 to code the behavioral experiment. The words were randomly assigned into two (2) different blocks consisting of seventy-seven (77) words with equal numbers of each of the given parameters; number of syllables, tone classification, manipulation (original/acceptable, borderline acceptable, and unacceptable). The borderline words were manipulated to be perceived as equivocal, meaning that both 'incorrect' and 'correct' responses would be accurate responses. These words were used as filler words, with arbitrary occurrence in the behavioral test. The keyboards keys 'z' and 'm' constituted the reaction for 'yes' and 'no', with a shift in which letter constituted which respective response after a mid-experiment break. The same Bose 35 noise canceling headset were used for all the participants to ensure minimal sound obstruction from the environment. The participants were instructed to react by intuition to whether they considered the word spoken as Norwegian pronunciation by pressing the keys representing either 'yes' or 'no'. They were also told to only focus on the tone of the word, and disregard the 'computerized' sounds that could occur in some of the manipulations after. The participants were asked to react as quickly as possible after the auditory stimuli had been given, and reaction time limit was 3000 ms before the program continued with the next stimuli regardless of whether the participant had responded or not.

The data was then structured syllable-wise and then alphabetically before calculating the combined average response times and accuracies of each of the differently manipulated syllables. Microsoft Excel was used to calculate the averages and standard deviation of the response time (RT) and accuracy data acquired through the behavioral testing. The data averages of the 'borderline acceptable' words were calculated, but omitted from the results chapter. The calculations for the 'borderline acceptable' words are included in appendix C.

4. Results

4.1 Results from the EEG / ERP experiment, and the correlation with the behavioral experiment

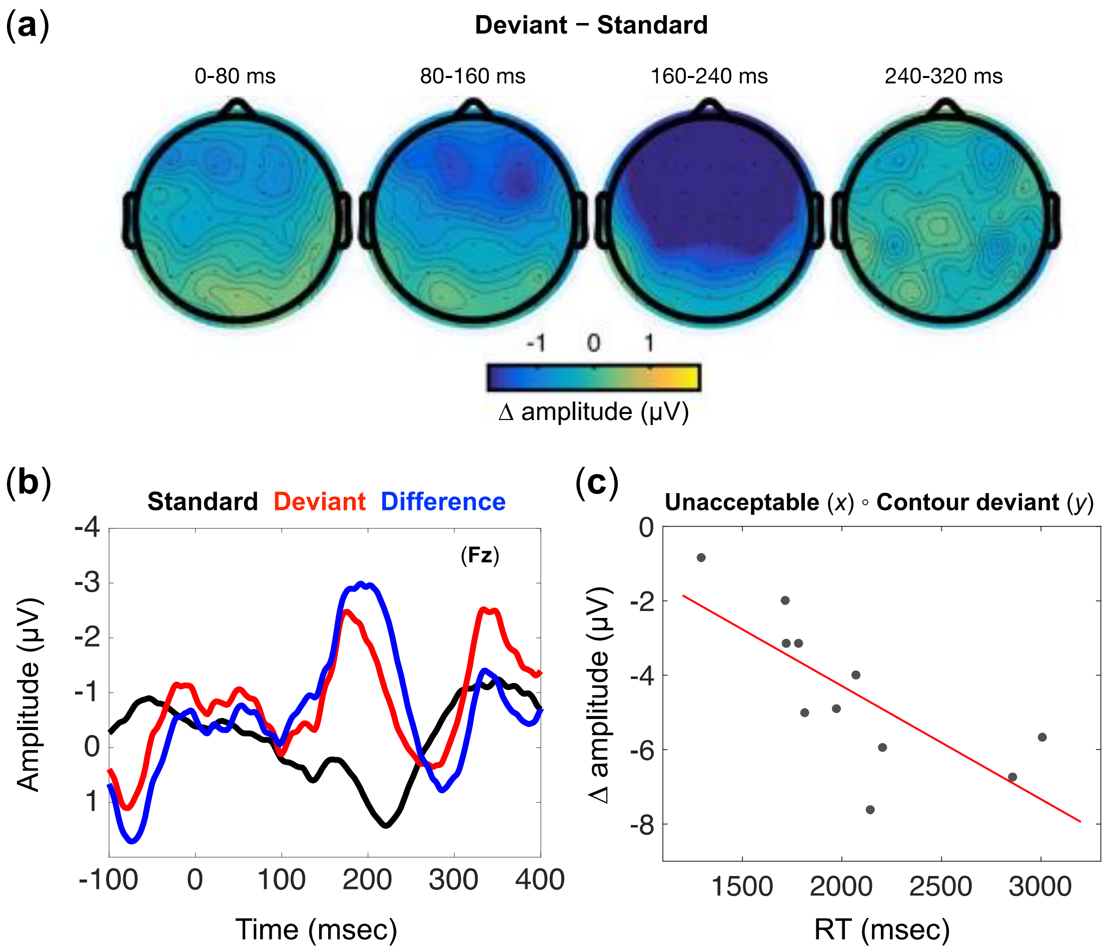

Figure 4.1 a) Visualization of MMN results in relation to described time intervals. b) Graph describing the timing where the mismatch negativity reaches the highest amplitude. c) Correlation between the ERP results and the behavioral results showing how the amplitude of the MMN in the ERP experiment correlates to the response time in the behavioral task.

As we can see from figure 4.1, the ERP-results show an increasing amplitude starting at 80 ms and dissolving after 240 ms, with its highest impact during 160-240 ms. The analyzes for the behavioral results and ERP-results show how the amplitude for the mismatch negativity correlates with response time in the behavioral trial.

4.2 Evaluation of the null hypothesis

Statistical evaluation of the null hypothesis was done using a t-test. The null hypothesis was rejected, as the mismatch amplitude correlates with response time when it comes to the words that have been manipulated to be unacceptable with ($t(1,9)=-3.3605, \rho=-0.746$, $p=0.008$). See figure 4.1.

In sum, we see that the results of the ERP seem to predict the behavioral outcome in unacceptable words.

4.3 Results from behavioral experiment

4.3.1 Result from behavioral testing of the main group

Table 4.1 Female response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the main group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1481.2 (± 499.3)	0.92 (± 0.16)
	UNACCEPTABLE	1938.9 (± 722.3)	0.22 (± 0.11)
Three	ACCEPTABLE	1535.9 (± 205.2)	0.99 (± 0.02)
	UNACCEPTABLE	1857.2 (± 411.0)	0.14 (± 0.12)
Four	ACCEPTABLE	1846.1 (± 353.6)	0.96 (± 0.09)
	UNACCEPTABLE	2059.4 (± 496.8)	0.16 (± 0.18)

Table 4.2 Male response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the main group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1742.6 (± 568.8)	0.98 (± 0.03)
	UNACCEPTABLE	2142.5 (± 631.4)	0.33 (± 0.12)
Three	ACCEPTABLE	1702.1 (± 202.7)	0.98 (± 0.03)
	UNACCEPTABLE	2282.6 (± 408.5)	0.31 (± 0.09)
Four	ACCEPTABLE	1876.0 (± 420.9)	1.00 (± 0.00)
	UNACCEPTABLE	2267.9 (± 391.6)	0.45 (± 0.22)

Table 4.3 Combined total response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the main group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1576.2 (± 513.8)	0.94 (± 0.13)
	UNACCEPTABLE	2012.9 (± 665.7)	0.26 (± 0.12)
Three	ACCEPTABLE	1596.3 (± 211.2)	0.99 (± 0.03)
	UNACCEPTABLE	2011.9 (± 444.4)	0.20 (± 0.14)
Four	ACCEPTABLE	1857.0 (± 358.3)	0.98 (± 0.08)
	UNACCEPTABLE	2135.2 (± 452.9)	0.27 (± 0.24)

4.3.2 Results from behavioral testing of the control group

Table 4.4 Female response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the control group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1427.2 (± 161.7)	0.97 (± 0.03)
	UNACCEPTABLE	1639.0 (± 245.3)	0.44 (± 0.29)
Three	ACCEPTABLE	1553.5 (± 160.1)	1.00 (± 0.00)
	UNACCEPTABLE	1913.7 (± 329.2)	0.36 (± 0.27)
Four	ACCEPTABLE	1720.2 (± 119.4)	0.99 (± 0.03)
	UNACCEPTABLE	1988.7 (± 277.1)	0.53 (± 0.34)

Table 4.5 Male response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the control group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1506.2 (± 367.7)	0.98 (± 0.03)
	UNACCEPTABLE	1909.6 (± 559.2)	0.56 (± 0.21)
Three	ACCEPTABLE	1600.6 (± 452.1)	0.99 (± 0.03)
	UNACCEPTABLE	2121.5 (± 646.4)	0.41 (± 0.24)
Four	ACCEPTABLE	1851.8 (± 446.1)	1.00 (± 0.00)
	UNACCEPTABLE	2247.6 (± 700.6)	0.58 (± 0.23)

Table 4.6 Combined total response-time (in milliseconds, ms) and accuracy averages for acceptable and unacceptable two-syllable words, three-syllable words and four-syllable words in the control group.

Syllable	Manipulation	RTs (ms) (± SD)	Accuracy (± SD)
Two	ACCEPTABLE	1466.7 (± 207.9)	0.98 (± 0.03)
	UNACCEPTABLE	1774.3 (± 392.3)	0.50 (± 0.25)
Three	ACCEPTABLE	1577.0 (± 168.2)	1.00 (± 0.02)
	UNACCEPTABLE	2017.6 (± 440.2)	0.39 (± 0.25)
Four	ACCEPTABLE	1786.0 (± 161.1)	1.00 (± 0.02)
	UNACCEPTABLE	2118.1 (± 455.1)	0.56 (± 0.28)

4.3.3 Statistical analyses of behavioral results

We used the program R[6] to run the statistical tests for the experiments, and found effects in response times (RT) in ‘tone placement’ and ‘word length’ in both the main group and the control group. The effects were found by using ANOVA, which is an analysis of variance, to see how much the independent variables have affected the dependent variables, in this case the independent variable being the word manipulations and the dependent variables being accuracy (ACC) and response time (RT).

In the main group, the main ANOVA effects on response time (RT) of ‘tone placement’ were F(1,10)=8.005, p=0.0179, see Figure 4.2a), ‘word length’ were (F(1,10)=10.18, p=0.009, see Figure 4.2b), while on accuracy (ACC) of ‘tone placement’ the main ANOVA effect was (F(1,10)=196.5, p<0.001; Figure 4.2c).

The main ANOVA effects found in the control group were similar: The main ANOVA effects on response times (RT) of ‘tone placement’ were (F(1,19)=19.41, p=0.0003, see Figure 4.2d), and for ‘word length’ they were F(1,19)=98.73, p<0.001, see Figure 4.2e). Lastly, the main effect on accuracy (ACC) of ‘tone placement’ was F(1,19)=86.05, p<0.001, see Figure 4.2f).

---

[6] https://www.R-project.org/.

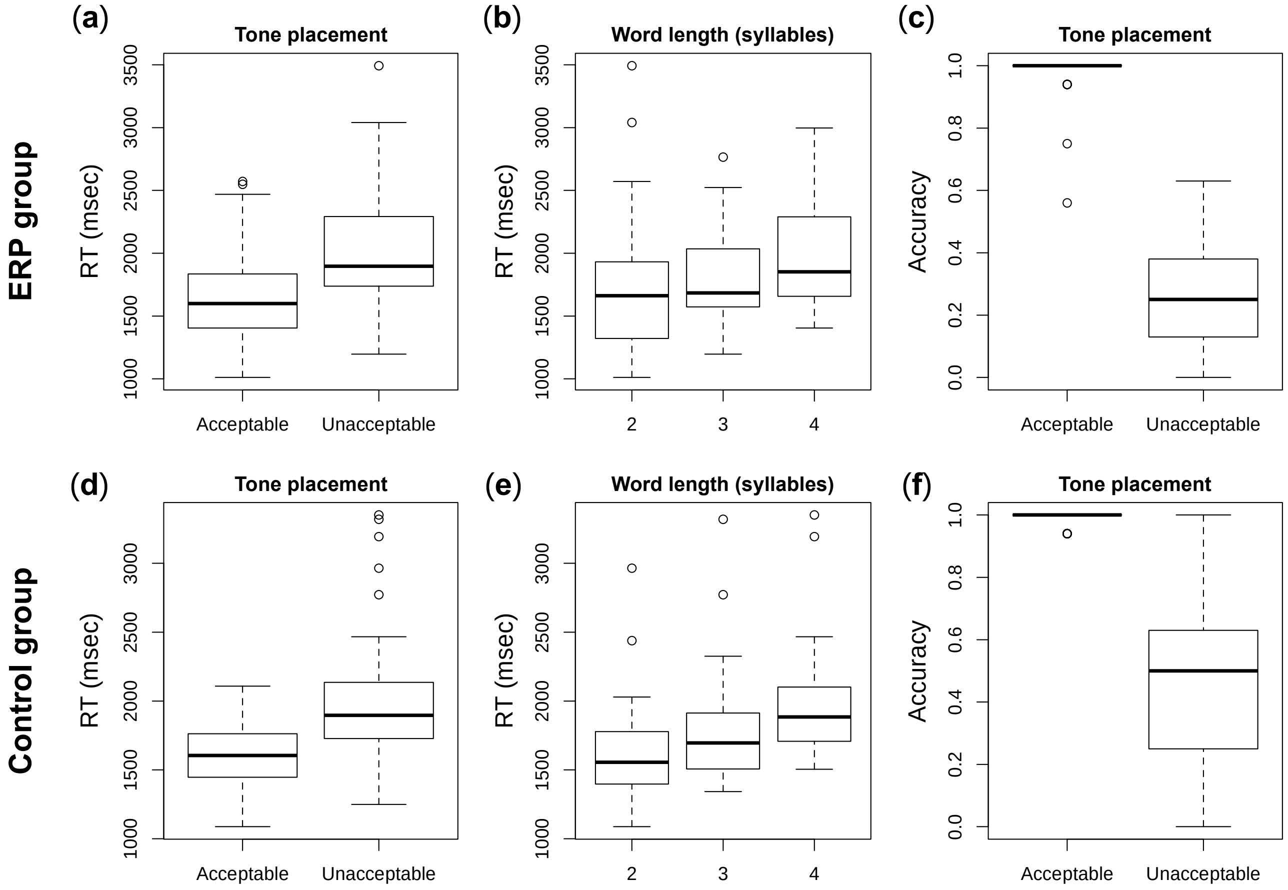

Figure 4.2. Showing results for response times and accuracy in relation to number of syllables and tone placement for the main (ERP) group and control group

These data indicate that both tone placement and word length has a main effect on response time independent of each other, and also that accuracy was affected by tone placement.

5. Discussion

5.1 Summary of the main findings

We assessed whether low-level auditory processes, that are not specifically speech perception processes, are actually recruited during perception of tone (mis)placement in Norwegian words. We measured participants´ sensitivity to irregularities in low-level auditory sequences (pure tone sequences) using the MMN in ERPs, and the perception of tone placement in Norwegian words in an independent behavioral task. Our results show that the results of the ERP do seem to predict the behavioral outcome in unacceptable words through both response time and accuracy.

5.2 Assessing the ERPs with the behavioral data

As the deviant response to the change in the rhythmic pattern is elicited starting at 80 ms and ending at around 240 milliseconds, our data supports the notion that early, low-level auditory processes are involved in detecting higher-level prosodic and grammatical features. The noun 'håpet' and the verb 'håpe' are both pronounced /håpe/, but with a difference in tone; tone 1 for the noun, and tone 2 for the verb. If we were to apply our results in regard to perception and the brains' ability to detect deviation in prosodic pattern for these two words, we could propose that a Norwegian native speaker would be able to detect the word's initial prosodic feature, in this case its toneme, and based on the stored prosodic language patterns already be able to identify the word category during what Friederici calls phase 0. As one of phase 0 functional processes is phoneme identification, and as a toneme is a phoneme, one can perhaps argue that a Norwegian speaker's word form and categorization processes happen earlier than posed in Friederici's original phase 1. An example to illustrate how this supposed early detection of tone in Norwegian predicts the word form is by putting each of the words 'håpet' (noun, tone 1) and 'håpe' (verb, tone 2) in a syntactic context appropriate for each of their respective word category (the bracketed words can be omitted in Norwegian natural speech).

1. Håpet brast (for henne)
2. Det er lov å håpe

If our assumptions prove to be correct, a Norwegian speaker should be able to correctly place the word /håpe/ in its correct syntactic environment solely based on the intonation in the initial phoneme, in this case depending on which of the two tones that is present in /hå/. The initial phoneme would then determine the word category in an early phase and would thus predict the syntactic structure, in this case based on the perceived pattern depending on whether the word /håpe/ is processed as a noun or a verb.

When it comes to correlations between our ERP-data and behavioral data, figure 4.1 c) show how the amplitude of the contour deviant in our ERP-experiment correlate with the response times of the unacceptable words in our behavioral test. The results indicate that the amplitude impact follow the response time, more specifically that a short response time indicates a lower amplitude while larger amplitude reaction is evoked incrementally as the response time increases.

5.3 Assessing the mismatch negativity in prosody prediction

Our results indicate that ERPs can be a tool to predict the behavioral outcome in unacceptable words through both response time and accuracy. As shown in fig 4.2.a), the ERP show neurological responses starting at around 80 ms before disappearing at approximately 240 ms.

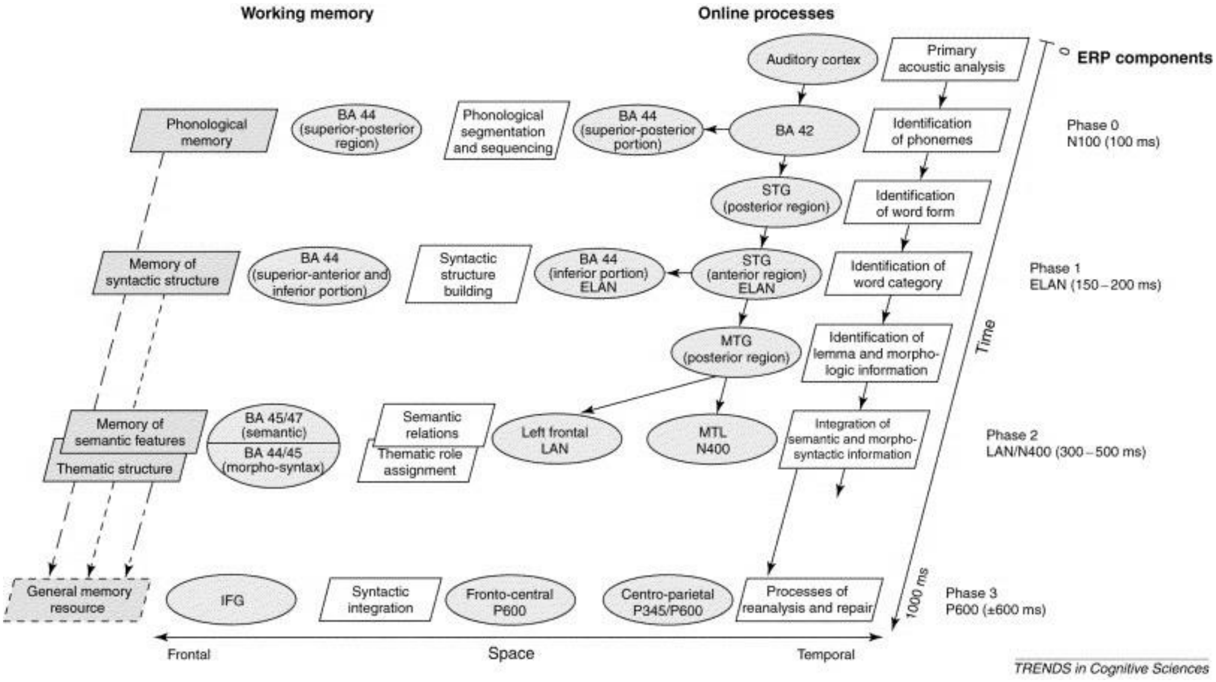

Figure 5.1. Neurocognitive model of auditory sentence processing where the boxes represent the functional processes while the ellipses represent the underlying neural correlate identified either by fMRI, PET or ERPs. (A.D. Friederici, 2002, Towards a neural basis of auditory sentence processing, p.79. Trends in Cognitive Sciences, 6(2). Reprinted with permission.)

When we look at Friederici's map of auditory processing phases, we see that phase 0 (first 150 ms) includes the following functional processes; primary acoustic analysis, identification of phonemes, and identification of word form (all in the temporal area), phonological memory (frontal area), and phonological segmentation and sequencing in between. The specific neural correlates for this phase are generated by regions including BA42, BA44, and the superior temporal gyrus (posterior region). If N100[7] was to be elicited, we would see this at around 100 ms.

---

[7] Winkler et. al (2009) explains N100 (also called N1) as being elicited "by sudden changes in sound energy, such as sound onset or an abrupt change in the spectral make-up of a continuous sound (I. Winkler et al., 2009)

The time interval representing phase 1 is when we see an elicited reaction across the frontal and temporal areas. This coincides with Friederici's mapping of phase 1 as it starts at 150-200 ms and expands up to approximately 300 ms. The functional processes in this phase are identification of word category and identification of lemma[8] and morphologic information in the temporal lobe, memory of syntactic structure in the frontal lobe, while we find syntactic structure building in-between. The specific neural correlates to this phase are BA44 and the anterior region of the superior temporal gyrus, with the middle temporal gyrus in the transition to phase 2. The ERP elicited in this area at around 150-200 ms is referred to as ELAN, or early left anterior negativity. As our ERP results disintegrate at around 240 ms, we do not go any further into phase 2, which has more to do with semantic and syntactic relations, and phase 3 which is concerned with further syntactic integration.

The ERPs elicited in our experiment show highest impact between 160 and 240 ms, which according to Friederici's model is the phase of identification of word category and syntactic structure building. Before this, we also see an emerging reaction in the initial phases of auditory processing as the brain starts to register a deviation in the perceived rhythmic pattern. It would be interesting to see if one can use the implications for such a reaction in future research to try to see whether one can suspect to find evidence of sensitivity in early prosodic detection in native Norwegian speakers. This hypothesis deserves further testing using ERPs, in conditions where the MMN is elicited not by sequences of pure tones, as was the case in the present study, but by actual Norwegian words. Preliminary evidence in favor of this hypothesis is discussed below.

To confer with later studies, studies by Winkler et. al (2009) say that there are "some regularities that can only be detected by persons with previous specialized training" (I. Winkler et al., 2009, p. 534), and further say that an example of this type of specialized training can be learning to speak a different language, bringing us back to the subject of whether bidialectism can be argued to be a type of bilingualism. This idea corresponds to our idea of how early detection of prosodic irregularities have an impact on word perception due to the brain detecting prosodic deviants during the phase of identification of phonemes, phonological segmentation and sequencing, and phonological memory. With this in mind, our results also coincide with a study

---

[8] A lexeme's logical form

by Winkler et al. (1999), where a group of Hungarians were divided into two sub-groups, one in which all of them spoke fluent Finnish, and the other one where none had any knowledge of Finnish. Through a vowel distinguishing trial, ERP results showed that those who spoke fluent Finnish reacted to the contrast between /e/ and /æ/ when /æ/ was presented during a continuous repetition of the /e/ vowel. The Hungarians who did not speak Finnish did not elicit a reaction, as /æ/ is used in Finnish, but not Hungarian. The results showed that in the Hungarians who spoke Finnish, MMN was elicited in the between 100 and 200 ms, with its top impact at about 150 ms (István Winkler et al., 1999). According to Friederici's model this is phase 1, where the ELAN is elicited in the functional processes of identification of word category and syntactic structure building. Winkler et al. (1999) conclude their abstract with the result indicating that "the fluent Hungarians developed cortical memory representations for the Finnish phoneme system that enabled them to preattentively categorize phonemes specific to this language" (István Winkler et al., 1999, p. 638). However, it is important to note that the MMN in our experiment was elicited by pure tones, and not complex speech tones. Yet, this kind of result linking the MMN to phonological processing directly as found in Winkler et al. (1999), provides some support for the hypotheses put forward in this section.

5.4 Assessing the behavioral data

The behavioral results for the acceptable words indicate a high level of correct detection of acceptable Norwegian words as the lowest combined accuracy being 0.94 (SD ± 0.13), suggesting that native Norwegians are competent in recognizing the prosodic patterns of Standard East Norwegian. For the unacceptable words, the lowest combined accuracy was 0.20 (SD ± 0.14).

It might be worth mentioning that the highest accuracy for unacceptable words for both main group and control group were the four syllable words, where the combined main group had an accuracy of 0.27 (SD ±0.24), while the control group's combined result for the four syllable words were as high as 0.56 for the accuracy (SD ± 0.28).

The statistical analyses of the findings in the main group and control group were similar, showing that both 'word length', and 'tone placement' had significant effects on response times.

5.4.1 Qualitative indications of the behavioral results

Despite working with a small sample, some factors emerged when analyzing the linguistic backgrounds of the subjects, and thus allowing the extraction of interesting factors to consider. As the quantitative analyses results of the small sample group showed an effect on response time prediction as well as accuracy, we need to consider the qualitative elements that could tell us more about the variables that could influence the quantitative results in our experiment. The following analyses are based on data found in Appendix B.

The subjects who identified their main dialect as Standard East Norwegian (SEN), had the lowest accuracies for the 'unacceptable' manipulations, with 0.21 as the highest score whereas the total average for the sample was 0.24. The average for those who did not identify as SEN speakers, was at 0.32. The SEN speakers also mostly had the shortest response times when it came to the words that were manipulated to an 'unacceptable' state, 1759.0 ms on average, compared to the rest of the group whose average response time to the same manipulations were 2221.5, totaling a difference of 462.5 ms on average for the same independent variables. These findings could indicate that the SEN speakers were a little to not at all hesitant to accept the correctness of their intuitive reactions, which again could help argue that there seems to be a greater acceptance for prosody modification in Norwegian dialectal speakers when it comes to intuitive offline reactions. The SEN-speaking person who scored the highest in this group had spent more years in Trondelag or further north than any of the other three in the SEN-group, thus further strengthening this assumption. This participant scored an accuracy of 0.21 for the unacceptable manipulations, whereas the others were all 0.10 or below.

The two that had lived in a country where English (a non-tonal language) was the main language had two of the lowest accuracy scores for the 'unacceptable' manipulations, while the subject who had the most extensive influence of other languages in their linguistic background did not only have English as second native language, but had also lived in China for some time, and additionally had basic language skills in several other tonal languages than Norwegian. This subject had one of the best scores in accuracy for the 'unacceptable' manipulations, and a lower response time than the average for both 'acceptable' and 'unacceptable' word manipulations. The difference between these two variances of 0.07 and 0.42 means a gap of 0.35. However, the two others placing above 0.40 in the ‘unacceptable’ accuracy gave no info of being influenced by other languages, other than being non-SEN speakers.

Lastly, the three subjects that identified themselves as users of Nynorsk did better in the accuracy for the ‘unacceptable’ manipulation than the average of the group. This finding might also help the notion that those who identify as non-SEN speakers, but rather belong to the demography of Nynorsk, have a better tolerance for prosody changes, but also are better at separating incorrect prosodic variations from correct variations.

It is important to note that the subject group assessed themselves to be either advanced or fluent in one or more other languages than Norwegian in addition to also claiming to be influenced by other Norwegian dialectal varieties. Our linguistic pattern says a lot about who we are, but at the same time it is common to change our speech features depending on our environment. The communication accommodation theory by Howard Giles (2014) addresses how people adjust how they communicate in order to accommodate others, a common trait for Norwegians who have relocated to another dialectal area and start to be influenced by the regional patterns of conduction (Giles, 2014). We are thus not in a position to claim that the results indicate that mixing of dialects or being proficient in another language cause a sensitivity to prosodic deviants, however, it might be an interesting aspect to pursue in the future.

5.5 A sociolinguistic approach to prosodic perception in Norwegian language

It is difficult to address the results in this study without mentioning what can be called ‘the bilingual-predicament’ that is often brought up when talking about Norwegian dialects. To surely assess our study in comparison to other studies, we need to address the notion of bilingualism in order to support our study on the correct terms. This is where it gets complicated, as the standard definition of being bilingual means having the ability to speak two languages fluently[9], but as we are discussing neural correlations regarding prosodic features in language, some studies, amongst them studies that argue against the standard definition and say that neural processes in bidialectism are similar to the neural processing occurring in bilingual brains. Two examples of studies which challenge the standard definition of bilingualism are “*The effect of childhood bilectalism and multilingualism on executive control*” by (Antoniou,

---

[9] There are several definitions of bilingualism, but this is one of the standard definitions

Grohmann, Kambanaros, & Katsos, 2016), and "*The effect of bidialectal literacy on school achievement*" by (Vangsnes, Söderlund, & Blekesaune, 2017), where both of them argue that being bidialectal can be compared to being bilingual. Throughout this discussion, both arguments will be taken into consideration when addressing relevant issues.

Our results indicate that native speakers of Norwegian seem to easily be able to detect a prosodically incorrect Norwegian word. However, as Norwegian dialects consist of sub-categories upon sub-categories of variants, one cannot state with any certainty whether an incorrect response is in fact incorrect, or if it is just another discrepancy of a sub-categorical dialect variation? As we only have two tones in Norwegian, low-tone and high-tone, the answer to this when faced with prosodic variations in dialects should in theory be "yes", especially when faced with prosodic variations in dialects. But as Norwegians are constantly exposed to prosodic diversity in form of both Norwegian variations as well as influence by other languages, how can one be sure that our brain responses are distinctive correct or incorrect when processing and predicting incoming speech? As a speaker of a tonal language needs to be able to quickly distinguish between prosodic variations within a word, one might assume that they outperform a non-tonal language speaker when presented with tasks that require quick prosodic detection.

It is common for languages to merge and evolve following contact with other languages and cultures, especially as factors such as immigration, tourism, and general cultural exchange make it increasingly easy to communicate and connect across borders. As the world becomes smaller and smaller, we are presented with larger prosodic variations in dialects than the ones that comes from "internal" differences. Non-native Norwegian speakers often bring their own prosodic features, or lack thereof, into the language learning process, creating what is considered "gebrokken" (broken) Norwegian. From a listener's point of view, one of the interesting things about spoken broken Norwegian is that many native Norwegians are prone to detecting the small change(s) to Norwegian prosodic features just as they would detect syntactical and lexical features, such as omitting neuter and instead use feminine or masculine gender for nouns. Complete fluency can be almost impossible to acquire for a non-native Norwegian speaker who has learned Norwegian as a second language, despite mastering both vocabulary and grammar in a close to perfect manner. They might even have acquired phonological elements from their new dialect (such as the West-Norwegian "skarre-/r/"), but there will still probably be little linguistic tells that will be detected by a native Norwegian speaker and thus reveal that the speaker does not have Norwegian as their native language. This notion leads us to the question of how we can locate the extremities of prosodic constraint, or whether it is possible to measure the scope of prosodic features in any way.

5.6 Comparison with other studies

As this thesis has argued for how native Norwegian speakers can identify the word category in some words just by its initial phoneme, there are studies that argue against implications of conscious phonemic sensitivity. Hickok and Poeppel (2000) argue that one does not consciously acknowledge phonemic segments, but rather interprets the word as a whole:

> Note that such tasks are fundamentally different from tasks that involve auditory comprehension: when one listens to an utterance in normal conversation, there is no conscious knowledge of the occurrence of specific phonemic segments, only the content of the message is consciously retained”. (Hickok & Poeppel, 2000, p. 134)

Warren (2008) also seem to support this notion as he says that there have been reports of experiments where a phoneme has been replaced with a noise burst or a cough, in which the subjects were not able to the identify the phoneme that was deleted from the utterance, and says that this strengthens the notion that comprehension comes from predictions based on acoustic pattern recognition as both the ‘real’ and the ‘restored’ phoneme seems to be inferred entities (R. M. Warren, 2008). He also says that as processing follows reception, there is evidence that we do not process individual words immediately after reception, but rather that there is a delay in the perceptual organization until more words are received in order to minimize errors (R. M. Warren, 2008). When it comes studies specifically centering the processing of syntactic and prosodic information in regard to auditory language comprehension, Friederici (2002) mentions that as there are few behavioral studies that have looked at these possible interactions, and in the few ones that have been performed, there have been studies that support the indication of such interaction, two of them being (P. Warren, Grabe, & Nolan, 1995) and (Kjelgaard & Speer, 1999). However, in regards to psycholinguistic experiments on language comprehension, she mentions that the two main classes of models used to account for the behavioral data are based on reading based data, and not principally on the role that prosodic elements have in spoken sentences (Friederici, 2002).

We thus need to be careful in regard to favoring indications of online or offline prosodic sensitivity, as much more research is needed in order to be able to argue such a hypothesis.

5.7 Limitations of the study

Amongst the limitations of this study are the small group of participants, which was partly caused by merely half of the EEG-study participants not being able to perform the behavioral task due to logistics/ traveling costs out of the span of this project. Even with a complete ratio of EEG/behavioral participants, the study would benefit from a larger number of subjects.

‘Weaknesses’ that are hard to omit include the computerized pitch manipulation of the words, as some of the higher pitch frequencies in the manipulated words caused an amount of words to sound synthetic, or “metallic”. Although the participants were told to disregard this, it could be a factor in how the prosodic alterations were perceived by the participants. A possible way to omit this problem could be by using word recordings of actual Norwegian dialects and counter them with either non-native Norwegian users who speak broken Norwegian, or having Norwegian words be pronounced by a speaker with no knowledge of the Norwegian language. However, the possibility of a perfect execution of a dialectal variety experiment as the one that we have performed in this study is unlikely due to the massive individualism in the large variety of Norwegian dialects. Another ‘weakness’ that is difficult to omit is how the experiments are performed in controlled environment. As the experiment is not performed in a natural speech setting but rather in a controlled environment with isolated words that have been manipulated digitally, it can thus be argued that we can only say that the data *indicate* results that are relevant for natural speech processing, not that they are decisive facts. Hickok and Poeppel (2000) say that speech perception experiments executed in a controlled environment do not extend to all of the brain areas that are normally a part of speech processing in natural language: “Specifically, the set of cognitive and neural systems involved in performing traditional laboratory speech perception tasks, such as syllable discrimination or identification, only partially overlap those involved in speech perception as it occurs during natural language comprehension” (Hickok & Poeppel, 2000, p. 131).

Lastly, when learning a new language, detaching from one’s native language’s prosodic features is often one of the more challenging features of language acquisition. Thus, proficiency is a subjective matter, where self-assessment is probably the least reliable source for measuring language competence. In our study, the majority of our subjects assessed themselves to be either fluent or advanced, which can cause imbalance when trying to assess someone’s language competence as defining a person’s proficiency in another language is a subjective matter.

6. Conclusion

We assessed whether low-level auditory processes, that are not specifically speech perception processes, are actually recruited during perception of tone (mis)placement in Norwegian words. We measured participants´ sensitivity to irregularities in low-level auditory sequences (pure tone sequences) using the MMN in ERPs, and the perception of tone placement in Norwegian words in an independent behavioral task. Our results show that the results of the ERP do seem to predict the behavioral outcome in unacceptable words through both response time and accuracy.

In chapter five, we assess the elicitation of mismatch negativity in pure tone sequences, and ask whether this result could be grounds for a similar test regarding implications of possible early prosodic detection in Norwegian speakers. As mentioned, we believe that such a hypothesis deserves further testing using ERPs in conditions where the MMN is elicited by actual Norwegian words, not by sequences of pure tones.

6.1 Further research/ thoughts

1. If we can stretch our findings to say that speakers of tonal languages are more sensitive to distinguishing all kinds of prosodic features, although we do not compare a tonal language to a non-tonal language in our study, we open up to a whole new sub-set of hypotheses for language acquisition and musicality, and also the correlations between the two. One could perhaps argue that tonal language speakers are more capable of not only learning new languages, but also that they are more likely to acquire a native-like pronunciation compared to a non-tonal speaker, or that tonal language speakers have a better chance of being musically talented as pitch perception is an important factor when creating and producing music.

2. As we have argued that we see effects of sensitivity in prediction when it comes to manipulation of typical Norwegian words, how would the experiment have changed if we tested for the same prosodic acceptance when manipulating loan words?

3. Dyslexia: In Freberg (2010), it says that “most cases of dyslexia involve poor phonological awareness, or the ability to discriminate verbal information at the level of speech sounds, or phonemes, as evidenced by their difficulties with words that rhyme (p.395), and also that they process speech sounds slower than non-dyslexics. What differences in response times and accuracies would we have found if we tested the same experiment on dyslexic subjects?

4. If there one could argue for higher sensitivity of prosodic detection in native tonal language speakers, what would happen if we tested the prosodic sensitivity of 100 non-musically trained non-tonal language native speakers, 100 non-musically trained tonal language native speakers, 100 musically trained non-tonal language native speakers, and 100 musically trained tonal language native speakers, and compared their results?

The prosodic sensitivity should also be a factor to recognize in different work fields. If one group of people are more prone to quickly detect prosodic inconsistencies on the base of both learned and instinctive predicted patterns, the ability to detect important discrepancies due to deception, coercion, intent, or even overt meanings could prove to be an immensely important skill. In for example forensic linguistics, where profiling based on patterns is a crucial element when mapping someone’s intelligence, academic background, age, dialectal background and so on.

7. References

An Introduction to Sphericity. (2017). Homepages.gold.ac.uk. Retrieved 3 May 2017, from http://homepages.gold.ac.uk/aphome/spheric.html

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition, 149*, 18-30. doi:https://doi.org/10.1016/j.cognition.2015.12.002

Ardila, A., Bernal, B., & Rosselli, M. (2016). How Localized are Language Brain Areas? A Review of Brodmann Areas Involvement in Oral Language. *Archives of Clinical Neuropsychology, 31*(1), 112-122. doi:10.1093/arclin/acv081

Ashby, M., & Maidment, J. A. (2005). *Introducing phonetic science*. Cambridge ; New York: Cambridge University Press.

Baum, S. R., & Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13*(8), 581-608.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences, 106*(5), 1672-1677. doi:10.1073/pnas.0809667106

Brodal, P. (2010). *The Central Nervous System* (4th ed. ed.). Oxford: Oxford University Press, USA.

Charbonneau, S., Scherzer, B. P., Aspirot, D., & Cohen, H. (2003). Perception and production of facial and prosodic emotions by chronic CVA patients. *Neuropsychologia, 41*(5), 605-613.

Crossman, A. R., & Neary, D. (2005). *Neuroanatomy : an illustrated colour text* (3rd ed.). Edinburgh ; New York: Elsevier/Churchill Livingstone.

Freberg, L. A. (2010). *Discovering biological psychology* (2nd ed. ed.). Belmont, Calif: Wadsworth.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *J Cogn Neurosci, 12*(1), 207-222.

Giles, H. (2014). Communication Accommodation Theory (pp. 208-210).

Hagoort, P., & Levelt, W. J. (2009). Neuroscience. The speaking brain. *Science, 326*(5951), 372-373. doi:10.1126/science.1181675

Hagoort, P., & Poeppel, D. (2013). The Infrastructure of the Language-Ready brain. In M. A. Arbib (Ed.), *Language, music, and the brain : a mysterious relationship*. Cambridge, Mass: MIT Press.

Hazanavicius, M., Dujardin, J., & Bejo, B. (2011). The Artist. Norge: Scanbox Entertainment Norway.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci, 4*(4), 131-138.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity. *Journal of Memory and Language, 40*(2), 153-194. doi:10.1006/jmla.1998.2620

Koelsch, S. (2013). Emotion in Action, Interaction, Music, and Speech. In M. A. Arbib (Ed.), *Language, music, and the brain : a mysterious relationship*. Cambridge, Mass: MIT Press.

Ladd, D. R. (2013). An Integrated view of Phonetics, Phonology, and Prosody. In M. A. Arbib (Ed.), *Language, music, and the brain : a mysterious relationship*. Cambridge, Mass: MIT Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass: M.I.T. Press.

Lundskær-Nielsen, T., Barnes, M. P., & Lindskog, A. (2005). *Introduction to Scandinavian phonetics : Danish, Norwegian and Swedish*. København: Alfabeta.

Patel, S., Scherer, K. R., Bjorkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: the role of voice production. *Biol Psychol, 87*(1), 93-98. doi:10.1016/j.biopsycho.2011.02.010

Povel, D.-J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 30*(6), 565-572. doi:10.3758/bf03202011

Scherer, K. R. (2013). Neural Correlates of Music Perception. In M. A. Arbib (Ed.), *Language, music, and the brain : a mysterious relationship*. Cambridge, Mass: MIT Press.

Schlaug, G., Jancke, L., Huang, Y. X., & Steinmetz, H. (1995). In-Vivo Evidence of Structural Brain Asymmetry in Musicians. *Science, 267*(5198), 699-701. doi:DOI 10.1126/science.7839149

Tiippana, K. (2014). What is the McGurk effect? *Front Psychol, 5*, 725. doi:10.3389/fpsyg.2014.00725

Vangsnes, Ø. A., Söderlund, G. B. W., & Blekesaune, M. (2017). The effect of bidialectal literacy on school achievement. *International Journal of Bilingual Education and Bilingualism, 20*(3), 346-361. doi:10.1080/13670050.2015.1051507

Vigneau, M., Beaucousin, V., Herve, P. Y., Jobard, G., Petit, L., Crivello, F., . . . Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*(1), 577-593. doi:10.1016/j.neuroimage.2010.07.036

Waller, D., & Greenauer, N. (2013). The Function, Structure, Form, and Content of Environmental Knowledge. In B. H. Ross (Ed.), *Psychology of Learning and Motivation*. Waltham, UNITED STATES: Elsevier Science.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes, 10*(5), 457-486. doi:10.1080/01690969508407112

Warren, R. M. (2008). *Auditory perception : an analysis and synthesis* (3rd ed. ed.). Cambridge: Cambridge University Press.

Wetterlin, A., & Lahiri, A. (2012). Tonal alternations in Norwegian compounds (Vol. 29, pp. 279).

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci, 13*(12), 532-540. doi:10.1016/j.tics.2009.09.003

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., . . . Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*(5), 638-642. doi:10.1111/1469-8986.3650638

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage, 10*(5), 544-554. doi:10.1006/nimg.1999.0491

8. Appendices

8.1 Appendix A

List of words used in the behavioral experiment.

Word	Word length	Tone	Type of manipulation
Bursdag	2	1	Original
Firma	2	1	Original
Flere	2	1	Original
Kabel	2	1	Original
Mindre	2	1	Original
Påstå	2	1	Original
Såpass	2	1	Original
Verden	2	1	Original
Bursdag	2	1	Borderline
Firma	2	1	Borderline
Flere	2	1	Borderline
Kabel	2	1	Borderline
Mindre	2	1	Borderline
Påstå	2	1	Borderline
Såpass	2	1	Borderline
Verden	2	1	Borderline
Bursdag	2	1	Not acceptable
Firma	2	1	Not acceptable
Flere	2	1	Not acceptable
Kabel	2	1	Not acceptable
Mindre	2	1	Not acceptable
Påstå	2	1	Not acceptable
Såpass	2	1	Not acceptable
Verden	2	1	Not acceptable
Elske	2	2	Original
Ikke	2	2	Original
Innhold	2	2	Original
Lomme	2	2	Original
Mening	2	2	Original
Noen	2	2	Original
Panne	2	2	Original
Plage	2	2	Original
Elske	2	2	Borderline
Ikke	2	2	Borderline

Innhold	2	2	Borderline
Lomme	2	2	Borderline
Mening	2	2	Borderline
Noen	2	2	Borderline
Panne	2	2	Borderline
Plage	2	2	Borderline
Elske	2	2	Not acceptable
Ikke	2	2	Not acceptable
Innhold	2	2	Not acceptable
Lomme	2	2	Not acceptable
Mening	2	2	Not acceptable
Noen	2	2	Not acceptable
Panne	2	2	Not acceptable
Plage	2	2	Not acceptable
Egentlig	3	1	Original
Genseren	3	1	Original
Innrømme	3	1	Original
Offentlig	3	1	Original
Oppmuntre	3	1	Original
Passkontroll	3	1	Original
Påvirke	3	1	Original
Studie	3	1	Original
Egentlig	3	1	Borderline
Genseren	3	1	Borderline
Innrømme	3	1	Borderline
Offentlig	3	1	Borderline
Oppmuntre	3	1	Borderline
Passkontroll	3	1	Borderline
Påvirke	3	1	Borderline
Studie	3	1	Borderline
Egentlig	3	1	Not acceptable
Genseren	3	1	Not acceptable
Innrømme	3	1	Not acceptable
Offentlig	3	1	Not acceptable
Oppmuntre	3	1	Not acceptable
Passkontroll	3	1	Not acceptable
Påvirke	3	1	Not acceptable
Studie	3	1	Not acceptable
Byggeplass	3	2	Original
Husleie	3	2	Original
Høyere	3	2	Original
Lykkelig	3	2	Original
Meningen	3	2	Original

Saksøke	3	2	Original
Tankefull	3	2	Original
Utlandet	3	2	Original
Byggeplass	3	2	Borderline
Husleie	3	2	Borderline
Høyere	3	2	Borderline
Lykkelig	3	2	Borderline
Meningen	3	2	Borderline
Saksøke	3	2	Borderline
Tankefull	3	2	Borderline
Utlandet	3	2	Borderline
Byggeplass	3	2	Not acceptable
Husleie	3	2	Not acceptable
Høyere	3	2	Not acceptable
Lykkelig	3	2	Not acceptable
Meningen	3	2	Not acceptable
Saksøke	3	2	Not acceptable
Tankefull	3	2	Not acceptable
Utlandet	3	2	Not acceptable
Anbefale	4	1	Original
Avslappende	4	1	Original
Kaffetrakter	4	1	Original
Livsnødvendig	4	1	Original
Ombestemme	4	1	Original
Oppdatering	4	1	Original
Pågripelse	4	1	Original
Studiene	4	1	Original
Anbefale	4	1	Borderline
Avslappende	4	1	Borderline
Kaffetrakter	4	1	Borderline
Livsnødvendig	4	1	Borderline
Ombestemme	4	1	Borderline
Oppdatering	4	1	Borderline
Pågripelse	4	1	Borderline
Studiene	4	1	Borderline
Anbefale	4	1	Not acceptable
Avslappende	4	1	Not acceptable
Kaffetrakter	4	1	Not acceptable
Livsnødvendig	4	1	Not acceptable
Ombestemme	4	1	Not acceptable
Oppdatering	4	1	Not acceptable
Pågripelse	4	1	Not acceptable
Studiene	4	1	Not acceptable

Barnehage	4	2	Original
Eiendommen	4	2	Original
Gjennomføre	4	2	Original
Grensekontroll	4	2	Original
Overbevist	4	2	Original
Overlate	4	2	Original
Redningsmannskap	4	2	Original
Tidligere	4	2	Original
Barnehage	4	2	Borderline
Eiendommen	4	2	Borderline
Gjennomføre	4	2	Borderline
Grensekontroll	4	2	Borderline
Overbevist	4	2	Borderline
Overlate	4	2	Borderline
Redningsmannskap	4	2	Borderline
Tidligere	4	2	Borderline
Barnehage	4	2	Not acceptable
Eiendommen	4	2	Not acceptable
Gjennomføre	4	2	Not acceptable
Grensekontroll	4	2	Not acceptable
Overbevist	4	2	Not acceptable
Overlate	4	2	Not acceptable
Redningsmannskap	4	2	Not acceptable
Tidligere	4	2	Not acceptable

8.2 Appendix B

Questionnaire (in Norwegian) used for the qualitative assessment.

Bakgrunnsinformasjon for forskningsprosjekt om lesing og ordprosessering

Tusen takk for at du har sagt ja til å delta i vårt forskningsprosjekt om lesing og ordprosessering. I dette skjemaet ber vi om bakgrunnsinformasjon som er nødvendig for at resultatene fra undersøkelsen skal kunne brukes.

Alle opplysningene du gir her, vil senere bli behandlet uten direkte gjenkjennende opplysninger. En kode knytter deg til dine opplysninger gjennom en deltakerliste. Det er kun autorisert personell knyttet til prosjektet som har adgang til deltakerlisten og som kan finne tilbake til infoen. Del B og C av dette skjemaet vil bare oppbevares med koden. All informasjon vil bli anonymisert ved prosjektslutt. Det vil ikke være mulig å identifisere deg i resultatene av studien når disse publiseres.

Legg merke til at skjemaet har 4 sider.

Skjemaet leveres direkte til meg eller sendes på e-post til

Med takknemlig hilsen,

Del A: Personlig informasjon

Fag/Yrke: ____

Fødselsår: ________________

Kjønn ☐ Kvinne ☐ Mann

Bostedskommune: __________________________

Deltakerkode:

(Fylles inn av prosjektleder)

Del B: Språklig bakgrunn

Morsmål

Er norsk morsmålet ditt?

☐ Ja ☐ Nei

Hvis ja, har du andre morsmål i tillegg?

☐ Ja ☐ Nei

Hvis ja, hvilke(t) språk? ______________________________

Hvilket språk bruker dere hjemme? ______________________________

På norsk, hvilken dialekt snakker du? ______________________________

Hvor i Norge har du bodd, og hvor lenge?

Kommune	**Antall år totalt**

Hvor ofte leser du tekst skrevet på bokmål?

hver dag flere ganger per uke et par ganger i uken av og til aldri

Hvor ofte skriver du tekst på bokmål?

hver dag flere ganger per uke et par ganger i uken av og til aldri

Hvor ofte leser du tekst skrevet på nynorsk?

hver dag flere ganger per uke et par ganger i uken av og til aldri

Hvor ofte skriver du tekst på nynorsk?

hver dag flere ganger per uke et par ganger i uken av og til aldri

Vil du definere deg selv som en som bruker bokmål?

ikke i det hele tatt nesten mer eller mindre stort sett fullstendig

Vil du definere deg selv som en som bruker nynorsk?

ikke i det hele tatt nesten mer eller mindre stort sett fullstendig

Engelsk og andre fremmedspråk

I **engelsk**, hvordan vurderer du ferdighetene dine på hvert av disse områdene?

	Grunnleggende	**Middels**	**Avansert**	**Flytende**
Lesing				
Skriving				
Snakke				
Lytte				
Totalt				

Har du bodd i, eller hatt lengre opphold i, et land hvor engelsk er hovedspråk?

☐ Ja ☐ Nei

Hvis ja, hvor lenge varte oppholdet/oppholdene?______________________________

____

Har du bodd i, eller hatt lengre opphold i, et land hvor annet enn engelsk er hovedspråk?

☐ Ja ☐ Nei

Hvis ja, hvor var det, og hvor lenge varte oppholdet/oppholdene?

____

Hvilke språk kan du utover morsmålet ditt og engelsk? *(Hvis du ikke snakker andre språk, gå til **del C**)*

Språk	**Nivå**			
	Grunnleggende	**Middels**	**Avansert**	**Flytende**
Tysk				
Fransk				
Spansk				
- angi språk				
- angi språk				
- angi språk				

Del C: Andre faktorer i språklæring

Har du, eller har du hatt, problemer med synet utover normal brillebruk?

☐ Ja ☐ Nei

Har du, eller har du hatt, problemer med hørselen?

☐ Ja ☐ Nei

Har du, eller har du hatt, språkvansker av noe slag (spesifikke språkvansker, lese-/lærevansker eller lignende)?

☐ Ja ☐ Nei

Har du, eller har du hatt, andre diagnoser som kan tenkes å påvirke språklæring (ADHD, autisme eller lignende)?

☐ Ja ☐ Nei

Er du venstrehendt?

☐ Ja ☐ Nei

8.3 Appendix C

Raw data from the behavioral experiment for main and control group accordingly.

Main group

Observation	Subject	Gender	Syllables	Manipulation	Response time (ms)	Accuracy
s1	101	F	2s	ACC	1289,5	1
s2	101	F	2s	UNACC	1698,06	0,25
s3	101	F	2s	BORD	1884,63	1
s4	101	F	3s	ACC	1518,81	1
s5	101	F	3s	UNACC	2035,81	0,38
s6	101	F	3s	BORD	1988,06	1
s7	101	F	4s	ACC	1657,31	1
s8	101	F	4s	UNACC	2189,25	0,13
s9	101	F	4s	BORD	2229,56	1
s10	102	M	2s	ACC	1644,56	1
s11	102	M	2s	UNACC	1931,56	0,25
s12	102	M	2s	BORD	2020,06	1
s13	102	M	3s	ACC	1618,63	1
s14	102	M	3s	UNACC	2292,69	0,44
s15	102	M	3s	BORD	2256,75	1
s16	102	M	4s	ACC	1868,94	1
s17	102	M	4s	UNACC	2204,5	0,63
s18	102	M	4s	BORD	2283,88	1
s19	103	F	2s	ACC	1366,69	1
s20	103	F	2s	UNACC	3493,06	0,25
s21	103	F	2s	BORD	2209,75	1
s22	103	F	3s	ACC	1572,94	1
s23	103	F	3s	UNACC	2523,63	0,06
s24	103	F	3s	BORD	2486,5	1
s25	103	F	4s	ACC	1714,94	1
s26	103	F	4s	UNACC	2997,19	0,13
s27	103	F	4s	BORD	3020,5	1
s28	104	F	2s	ACC	1320,88	1
s29	104	F	2s	UNACC	1678,94	0,19
s30	104	F	2s	BORD	1405,25	1
s31	104	F	3s	ACC	1595,94	1
s32	104	F	3s	UNACC	1737,75	0,06

s33	104	F	3s	BORD	1451,31	1
s34	104	F	4s	ACC	2364,5	1
s35	104	F	4s	UNACC	1743,63	0,06
s36	104	F	4s	BORD	1720,5	1
s37	105	F	2s	ACC	1574,81	0,94
s38	105	F	2s	UNACC	1723,44	0,06
s39	105	F	2s	BORD	1624,5	1
s40	105	F	3s	ACC	1599,81	1
s41	105	F	3s	UNACC	1848,94	0,13
s42	105	F	3s	BORD	1748,69	1
s43	105	F	4s	ACC	1835,38	1
s44	105	F	4s	UNACC	1873,38	0
s45	105	F	4s	BORD	1832,44	1
s46	106	F	2s	ACC	1256,69	0,94
s47	106	F	2s	UNACC	1896,00	0,38
s48	106	F	2s	BORD	1606,06	1
s49	106	F	3s	ACC	1390,50	1
s50	106	F	3s	UNACC	1623,38	0,19
s51	106	F	3s	BORD	1751,69	1
s52	106	F	4s	ACC	1656,88	1
s53	106	F	4s	UNACC	1824,00	0,44
s54	106	F	4s	BORD	1908,38	1
s55	107	M	2s	ACC	2571,5	1
s56	107	M	2s	UNACC	3041	0,31
s57	107	M	2s	BORD	2789,19	1
s58	107	M	3s	ACC	2002,25	1
s59	107	M	3s	UNACC	2765,63	0,31
s60	107	M	3s	BORD	2532,19	1
s61	107	M	4s	ACC	2469,63	1
s62	107	M	4s	UNACC	2769,19	0,56
s63	107	M	4s	BORD	3168,44	1
s64	108	M	2s	ACC	1434,81	0,94
s65	108	M	2s	UNACC	1566,56	0,5
s66	108	M	2s	BORD	1621,56	1
s67	108	M	3s	ACC	1557,00	0,94
s68	108	M	3s	UNACC	1766,25	0,25
s69	108	M	3s	BORD	1587,06	1
s70	108	M	4s	ACC	1641,44	1
s71	108	M	4s	UNACC	1815,63	0,5
s72	108	M	4s	BORD	2053,13	1

s73	109	F	2s	ACC	1010,81	1
s74	109	F	2s	UNACC	1205,81	0,13
s75	109	F	2s	BORD	1248,31	1
s76	109	F	3s	ACC	1203,31	1
s77	109	F	3s	UNACC	1196,19	0
s78	109	F	3s	BORD	1289,63	1
s79	109	F	4s	ACC	1403,81	1
s80	109	F	4s	UNACC	1481,94	0
s81	109	F	4s	BORD	1387,31	1
s82	110	F	2s	ACC	2548,81	0,56
s83	110	F	2s	UNACC	1876,75	0,31
s84	110	F	2s	BORD	1943,94	1
s85	110	F	3s	ACC	1869,69	0,94
s86	110	F	3s	UNACC	2034,44	0,19
s87	110	F	3s	BORD	2285,69	1
s88	110	F	4s	ACC	2289,94	0,75
s89	110	F	4s	UNACC	2306,69	0,38
s90	110	F	4s	BORD	2328,69	1
s91	112	M	2s	ACC	1319,31	1
s92	112	M	2s	UNACC	2030,75	0,25
s93	112	M	2s	BORD	1800,44	1
s94	112	M	3s	ACC	1630,38	1
s95	112	M	3s	UNACC	2305,63	0,25
s96	112	M	3s	BORD	2182,25	1
s97	112	M	4s	ACC	1523,94	1
s98	112	M	4s	UNACC	2282,13	0,125
s99	112	M	4s	BORD	2456,38	1

Control group

Observation	Subject	Gender	Syllables	Manipulation	Response times (ms)	Accuracy
s1	44FR	F	2s	ACC	1466	1
s2	44FR	F	2s	UNACC	1747,56	0,75
s3	44FR	F	2s	BORD	1802,75	1
s4	44FR	F	3s	ACC	1442,63	1
s5	44FR	F	3s	UNACC	1918,38	0,75
s6	44FR	F	3s	BORD	1941,75	1
s7	44FR	F	4s	ACC	1699,75	0,9375
s8	44FR	F	4s	UNACC	1884,125	1,00
s9	44FR	F	4s	BORD	1908,19	1
s10	58FR	F	2s	ACC	1279,81	1
s11	58FR	F	2s	UNACC	1400,75	0,125
s12	58FR	F	2s	BORD	1295,56	1
s13	58FR	F	3s	ACC	1397,56	1
s14	58FR	F	3s	UNACC	1460,38	0,00
s15	58FR	F	3s	BORD	1444,0625	1
s16	58FR	F	4s	ACC	1612,88	1
s17	58FR	F	4s	UNACC	1636,3125	0,00
s18	58FR	F	4s	BORD	1619,19	1
s19	59FR	F	2s	ACC	1392,56	0,9375
s20	59FR	F	2s	UNACC	1365,50	0,9375
s21	59FR	F	2s	BORD	1384,3125	1
s22	59FR	F	3s	ACC	1483,38	1
s23	59FR	F	3s	UNACC	1894,19	0,63
s24	59FR	F	3s	BORD	1688,5	1
s25	59FR	F	4s	ACC	1709,94	1
s26	59FR	F	4s	UNACC	1719,50	1,00
s27	59FR	F	4s	BORD	1701,0625	1
s28	60F	F	2s	ACC	1388,13	0,9375
s29	60F	F	2s	UNACC	1966,00	0,44
s30	60F	F	2s	BORD	1810,4375	1
s31	60F	F	3s	ACC	1891,25	1
s32	60F	F	3s	UNACC	2251,375	0,25
s33	60F	F	3s	BORD	2267,63	1
s34	60F	F	4s	ACC	1888,3125	0,9375
s35	60F	F	4s	UNACC	2164,75	0,63
s36	60F	F	4s	BORD	2181	1

s37	61F	F	2s	ACC	1358,69	1,00
s38	61F	F	2s	UNACC	1485,63	0,56
s39	61F	F	2s	BORD	1449,9375	1
s40	61F	F	3s	ACC	1454,50	1
s41	61F	F	3s	UNACC	1705,06	0,38
s42	61F	F	3s	BORD	1681,88	1
s43	61F	F	4s	ACC	1596,75	1
s44	61F	F	4s	UNACC	1797,31	0,625
s45	61F	F	4s	BORD	1819,50	1
s46	63F	F	2s	ACC	1673,25	0,94
s47	63F	F	2s	UNACC	1913,25	0,13
s48	63F	F	2s	BORD	1971,13	1
s49	63F	F	3s	ACC	1666,75	1
s50	63F	F	3s	UNACC	2325,31	0,13
s51	63F	F	3s	BORD	2219,13	1
s52	63F	F	4s	ACC	1814,94	1
s53	63F	F	4s	UNACC	2252,75	0,31
s54	63F	F	4s	BORD	2338,19	1
s55	65F	F	2s	ACC	1488,6875	0,9375
s56	65F	F	2s	UNACC	1734,8125	0,56
s57	65F	F	2s	BORD	1866,69	1
s58	65F	F	3s	ACC	1596,375	1
s59	65F	F	3s	UNACC	2210,25	0,69
s60	65F	F	3s	BORD	2157,38	1
s61	65F	F	4s	ACC	1706,31	1
s62	65F	F	4s	UNACC	2467,00	0,56
s63	65F	F	4s	BORD	2437,44	1
s64	66F	F	2s	ACC	1649,38	1,00
s65	66F	F	2s	UNACC	1763,69	0,3125
s66	66F	F	2s	BORD	2157,19	1
s67	66F	F	3s	ACC	1633,56	1,00
s68	66F	F	3s	UNACC	1897,69	0,25
s69	66F	F	3s	BORD	1882,25	1
s70	66F	F	4s	ACC	1828,00	1
s71	66F	F	4s	UNACC	2098,88	0,4375
s72	66F	F	4s	BORD	2120,06	1
s73	71F	F	2s	ACC	1449,69	0,9375
s74	71F	F	2s	UNACC	1764,19	0,50
s75	71F	F	2s	BORD	2030,94	1
s76	71F	F	3s	ACC	1616,00	1

s77	71F	F	3s	UNACC	2119,56	0,4375
s78	71F	F	3s	BORD	2157,69	1
s79	71F	F	4s	ACC	1826,31	1
s80	71F	F	4s	UNACC	2155,88	0,625
s81	71F	F	4s	BORD	2191,06	1
s82	80FR	F	2s	ACC	1126,25	1,00
s83	80FR	F	2s	UNACC	1248,81	0,06
s84	80FR	F	2s	BORD	1545,25	1
s85	80FR	F	3s	ACC	1352,81	1,00
s86	80FR	F	3s	UNACC	1355,00	0,06
s87	80FR	F	3s	BORD	1405,19	1
s88	80FR	F	4s	ACC	1518,44	1
s89	80FR	F	4s	UNACC	1710,44	0,06
s90	80FR	F	4s	BORD	1618,44	1
s91	43MR	M	2s	ACC	1360,9375	1
s92	43M	M	2s	UNACC	1540,25	0,25
s93	43M	M	2s	BORD	1749,13	1
s94	43M	M	3s	ACC	1394,75	1
s95	43M	M	3s	UNACC	1910,4375	0,0625
s96	43M	M	3s	BORD	1727,25	1
s97	43M	M	4s	ACC	1865,75	1
s98	43M	M	4s	UNACC	1893,88	0,1875
s99	43M	M	4s	BORD	1984,75	1
s100	45ML	M	2s	ACC	1087,0625	0,9375
s101	45M	M	2s	UNACC	1266	0,5625
s102	45M	M	2s	BORD	1512,4375	1
s103	45M	M	3s	ACC	1341,6875	0,9375
s104	45M	M	3s	UNACC	1531,0625	0,25
s105	45M	M	3s	BORD	1514,5625	1
s106	45M	M	4s	ACC	1504,125	1
s107	45M	M	4s	UNACC	1601,1875	0,5
s108	45M	M	4s	BORD	1604,6875	1
s109	50MR	M	2s	ACC	1401,6875	1
s110	50MR	M	2s	UNACC	1791,5625	0,625
s111	50MR	M	2s	BORD	1837,75	1
s112	50MR	M	3s	ACC	1559,8125	1
s113	50MR	M	3s	UNACC	1911,4375	0,25
s114	50MR	M	3s	BORD	1875,5	1
s115	50MR	M	4s	ACC	1703,25	1
s116	50MR	M	4s	UNACC	2070,375	0,375

s117	50MR	M	4s	BORD	2054,4375	1
s118	53MR	M	2s	ACC	1569,5	1
s119	53MR	M	2s	UNACC	2964,875	0,75
s120	53MR	M	2s	BORD	2578,5	1
s121	53MR	M	3s	ACC	1591,25	1
s122	53MR	M	3s	UNACC	3320,3125	0,5625
s123	53MR	M	3s	BORD	3322,375	1
s124	53MR	M	4s	ACC	1868	1
s125	53MR	M	4s	UNACC	3194,25	0,6875
s126	53MR	M	4s	BORD	3595,5625	1
s127	57Mr	M	2s	ACC	1414,8125	0,9375
s128	57Mr	M	2s	UNACC	2438,875	0,3125
s129	57Mr	M	2s	BORD	2922,3125	1
s130	57Mr	M	3s	ACC	1621,875	1
s131	57Mr	M	3s	UNACC	2772,5625	0,625
s132	57Mr	M	3s	BORD	2288,875	1
s133	57Mr	M	4s	ACC	2108,625	1
s134	57Mr	M	4s	UNACC	3351,25	0,625
s135	57Mr	M	4s	BORD	3129,625	1
s136	62ML	M	2s	ACC	1453,1875	0,9375
s137	62ML	M	2s	UNACC	1860,375	0,5625
s138	62ML	M	2s	BORD	2020,625	1
s139	62ML	M	3s	ACC	1465,8125	1
s140	62ML	M	3s	UNACC	2051,5625	0,4375
s141	62ML	M	3s	BORD	1870,0625	1
s142	62ML	M	4s	ACC	1692,8125	1
s143	62ML	M	4s	UNACC	1932,5625	0,9375
s144	62ML	M	4s	BORD	2965	1
s145	67ML	M	2s	ACC	1623,4375	1
s146	67ML	M	2s	UNACC	1845,5	0,6875
s147	67ML	M	2s	BORD	1770,625	1
s148	67ML	M	3s	ACC	1865,4375	0,9375
s149	67ML	M	3s	UNACC	1915,125	0,5625
s150	67ML	M	3s	BORD	1954,875	1
s151	67ML	M	4s	ACC	1961,75	1
s152	67ML	M	4s	UNACC	2104,75	0,375
s153	67ML	M	4s	BORD	2098,3125	1
s154	68Mr	M	2s	ACC	1421,875	1
s155	68Mr	M	2s	UNACC	1868,0625	0,5
s156	68Mr	M	2s	BORD	1835,5625	1

s157	68Mr	M	3s	ACC	1581,4375	1
s158	68Mr	M	3s	UNACC	2091,3125	0,5
s159	68Mr	M	3s	BORD	1972,0625	1
s160	68Mr	M	4s	ACC	1883,875	1
s161	68Mr	M	4s	UNACC	2152,75	0,625
s162	68Mr	M	4s	BORD	2067,875	1
s163	69Mr	M	2s	ACC	2028,8125	0,9375
s164	69M	M	2s	UNACC	1813,3125	0,9375
s165	69M	M	2s	BORD	1859,4375	1
s166	69M	M	3s	ACC	1897,8125	1
s167	69M	M	3s	UNACC	1828,875	0,75
s168	69M	M	3s	BORD	2185,5625	1
s169	69M	M	4s	ACC	2014,5	1
s170	69M	M	4s	UNACC	1945,3125	0,875
s171	69M	M	4s	BORD	2195,5	1
s172	70ML	M	2s	ACC	1701	1
s173	70M	M	2s	UNACC	1706,875	0,375
s174	70M	M	2s	BORD	1747	1
s175	70M	M	3s	ACC	1686	1
s176	70M	M	3s	UNACC	1881,875	0,0625
s177	70M	M	3s	BORD	1940,25	1
s178	70M	M	4s	ACC	1914,875	1
s179	70M	M	4s	UNACC	2229,25	0,5625
s180	70M	M	4s	BORD	2179,625	1