Acoustics, phonetics, or phonology?

A neurolinguistic perspective on neural and behavioural asymmetries in vowel processing

Inauguraldissertation

zur Erlangung des Akademischen Grades

eines Dr. phil.,

vorgelegt dem Fachbereich 05 – Philosophie und Philologie

der Johannes Gutenberg-Universität

Mainz

von

Miriam Riedinger

aus Sigmaringen

2023

Datum des Prüfungskolloquiums: 17. Mai 2024

Content

List of tables

List of figures

I Introduction

Language, especially spoken language, is the most important means for human communication. For successful communication, language comprehension is an integral part. Part of this comprehension process, especially in oral communication, is speech perception. Speech perception can be defined as a "set of operations that transform an auditory signal into mental representations of a type that can make contact with internally stored information – that is, words" (Poeppel & Monahan, 2008, p. 80). To be able to identify words from the acoustic signal, the units constructing those words have to be recognised and processed. In most theories, the smallest units comprising words are phonemes, but there are also theories that utilise sub-phonemic units like distinctive features (for an overview, see Kazanina et al., 2018). Nevertheless, an important task for successful speech processing is to recognise and to be able to differentiate between different speech sounds. Thus, for speech sound discrimination, e.g., vowels, mental representations of those speech sounds are necessary.

In recent years, these mental representations have been vastly and extensively researched. Research has been focussing mental representations of vowels and the question which properties of vowels are involved in vowel processing and vowel discrimination. In this context, a widely described phenomenon is the presence of directional asymmetries in neural and behavioural effects in vowel discrimination. Directional asymmetry refers in this context to effects that tend to be asymmetrically pronounced in relation to the presentation order of same stimuli (Johnson, 2015). These directional asymmetries can be used "as a window into the speech perception process" (Johnson, 2015, p. 100) and, therefore, offer a means of investigating mental representations of speech sounds. In the literature, two models are often discussed regarding mental representation of vowels in adult listeners: the Featurally Underspecified Lexicon (Lahiri & Reetz, 2002a, 2010a) model and the Natural Referent Vowel (Polka & Bohn, 2011a) framework. What unites both models is that they both make specific hypotheses about directional asymmetries. Their approaches in explaining these asymmetrical effects and, therefore, their assumptions regarding features involved in mental presentations and vowel processing are different. While the first mentioned model assumes sparse and abstract phonological features, the latter one utilises phonetic features often leading to contradicting hypotheses regarding directional asymmetries. However, evidence for both competing models have been described in the literature. But a systematic and extensive model comparison is missing.

Therefore, this dissertation sets out to compare both models in neural and behavioural investigations regarding German vowel processing. Thus, the following research question shall be answered: **Are either phonological or phonetic features involved in German vowel processing?** Additionally, the following neurolinguistic studies of German vowel processing will also investigate whether neural and behavioural asymmetric effects are influenced by other factors not included in the models. Here, the studies will especially focus on the psychoacoustic factor of perceived loudness. In hearing research, it is well known that physical loudness (as intensity) and perceptual loudness (as perceived loudness) are different concepts and are measured on different scales (Florentine et al., 2011). Although perceived loudness is related to and heavily determined by sound intensity, two sounds of equal perceived loudness may well have different levels of sound intensity (Moore, 2005; Yanushevskaya et al., 2013). However, in psycho- and neurolinguistic research of vowel processing only sound intensity is usually controlled for, ignoring possible influences of perceived loudness on neural and behavioural discrimination effects. Therefore, perceived loudness influences will also be investigated in the conducted research of this dissertation.

The organisation of the dissertation is as follows.

Chapter 2 is the theoretical foundation of the later described empirical investigations. First, directional asymmetry as a concept will be briefly described. Additionally, a review of measures and tasks involved in investigations of directional asymmetries in speech sound processing will be conducted before focusing on the main methodology of this dissertation. Here, the oddball paradigm for stimulus presentation will be presented. Furthermore, the ERP component mismatch negativity, used in this dissertation for neural investigations, will be explained. In the second section of the theoretical chapter, both models, the Featurally Underspecified Lexicon (Lahiri & Reetz, 2002a, 2010a) and Natural Referent Vowel (Polka & Bohn, 2011a), will be described in-depth. For each model, the fundamental basics will be explained before pointing out the key features involved in vowel processing. Afterwards, the specific hypotheses and approaches in explaining directional asymmetries will be introduced. Each model description will conclude with a review of empirical evidence found in the research literature.

Chapter 3 focuses on the conducted empirical investigations of this dissertation regarding German vowel processing. In the first section, a series of studies using minimal pairs to investigate five German long vowel contrasts will be described and discussed. In the first series of experiments, minimal pairs were used to gain a more natural listening situation. Here, the research focused on the model comparison and identifying (additional) factors influencing German vowel discrimination. In the second section, a different series of studies using isolated models will be presented. Here, isolated German long vowels were used to test if previously found effects and influences could be replicated while reducing the complexity of possible (extra-)linguistic influences. Once more, the second series aimed to compare the discussed models. Furthermore, these experiments also will try to replicate found influences on vowel processing of the first experimental series.

In **chapter 4**, a summary of the dissertation, especially of the empirical investigation, will be given. Furthermore, this chapter will draw a conclusion on the found effects and will be answering the research question before concluding by pointing out future research ideas.

II Theoretical Background

1. Asymmetric vowel processing: neural and behavioural measures

A widely observed phenomenon in the research of vowel processing is the notion that effects often tend to be asymmetric. It has been shown that the presentation order of stimuli is related to asymmetric effects, i.e., it matters which sound is presented first. For example, when testing the discrimination of two vowels (e.g., [i] and [e]), one can present the vowels in two possible orders: the high vowel followed by the mid vowel ([i]—[e]) or in the reverse order ([e]—[i]). In a variety of tasks and measures, there have been differently pronounced effects depending on the presentation orders of the same stimuli (e.g., Eulitz & Lahiri, 2004; Polka & Bohn, 2003a; Scharinger, Monahan, et al., 2012). In case of the example introduced above, there are different effects expected for [i] – [e] and [e] – [i].
Thus, directional asymmetries in vowel processing can be defined as effects of varying intensity in the processing of the same speech sounds depending on the order of presentation.

This chapter will review different tasks and measures involved in the study of directional asymmetries in speech sound processing. Although there is some evidence of asymmetric sinusoidal tone processing (Peter et al., 2010), this dissertation will focus on directional asymmetries in speech sound processing only. Furthermore, as methodology for this dissertation, an in-depth description of the commonly used oddball-paradigm and the neural measure of Mismatch Negativity (MMN) will be given.

1.1 Measures and tasks in directional asymmetries

Directional asymmetries in speech sound processing have been found in a variety of measures and tasks. Asymmetric effects arise in behavioural as well as neural measures with relation to presentation orders.
On a behavioural level, studies typically employ accuracy measures as well as measuring reaction times. Regarding accuracy, directional asymmetries have been described using different operationalisations. Most commonly the mean percentage of correct responses showed different pronounced effects (Hwang et al., 2010; Kriengwatana & Escudero, 2017; Masapollo et al.,

2015; Tyler et al., 2014). Evidence of asymmetric effects in accuracy data has also been found in the analysis of error rates (Scharinger & Idsardi, 2014) with one stimulus order more error prone than the reversed one. Another widely used measure for the accuracy of discrimination are A' scores, which represent discrimination performances[1]. Here as well, multiple studies showed directional asymmetries depending on stimulus order (Y. Liu et al., 2021; Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka & Bohn, 1996). Another behavioural measure in investigating directional asymmetries is reaction time (RT), that are associated with underlying cognitive processes in task performance and therefore cognitive load and processing difficulties (De Boeck & Jeon, 2019). In the study of speech sound processing and directional asymmetries, RT is a widely used measure of task performance which allows consideration about processing difficulties. Multiple studies found asymmetric RTs depending on the stimulus order (Friedrich et al., 2008; Højlund et al., 2019; Hwang et al., 2010; Johnson, 2015; Lahiri & Reetz, 2010a; Riedinger et al., 2021; Scharinger, Domahs, et al., 2016; Scharinger & Idsardi, 2014; Scharinger & Lahiri, 2010; Schild et al., 2012). Savela and colleagues (2003), for example, revealed directional asymmetries in vowel processing in dependence phonemic status of the investigated vowels: When an unfamiliar vowel was presented first, processing was more difficult which resulted in longer RTs.

However, there is large evidence that directional asymmetries dependent on stimulus order are also present on a neural level. Different methods have been utilised for the investigations. For example, in an fMRI study Scharinger et al. (2016) provided evidence that activation levels in the superior temporal sulcus vary between reversed presentation orders of the same vowels. Some studies employed MEG measurements for the investigation of asymmetric effects but concentrated on amplitude and latency asymmetries of the electrophysiological effects (Scharinger, Idsardi, et al., 2011b; Scharinger, Merickel, et al., 2011; Scharinger, Monahan, et al., 2012, 2016; Tavabi et al., 2009).

But mostly, EEG or ERP measurements have been used in the investigation of directional asymmetries on a neural level. Several studies described asymmetric amplitude effects in different ERP components (mostly N100 and MMN) (e.g., de Jonge & Boersma, 2015; de Rue et al., 2021; Hestvik et al., 2020; Højlund et al., 2019; Kotzor et al., 2020; Nudga et al., 2021; Scharinger, Monahan, et al., 2012) but also asymmetric latencies between the presentation orders of the same speech sounds (e.g., de Rue et al., 2021; Eulitz & Lahiri, 2004; Ikeda et al., 2002). Furthermore, some studies were able to identify asymmetric effects in time-frequency analyses of EEG

[1] A' scores are described as non-parametric, dependent variable with values varying between 0 and 1. An A' score of 1 indicating perfect discrimination while a value of 0.5 can be interpreted as chance level (Masapollo, Polka, & Ménard, 2017; Polka & Bohn, 1996)

signals depending on the presentation order of speech sounds (Polka et al., 2021; Scharinger, Monahan, et al., 2016).

For the investigation of directional asymmetries in speech sound processing a variety of tasks and designs have been utilised in the past decades. Most of the studies used discrimination tasks, but some studies also used identification (Hwang et al., 2010) or priming designs (Friedrich et al., 2008; Lahiri & Reetz, 2010a; Llompart, 2021; Scharinger, 2006; Scharinger & Idsardi, 2014; Scharinger & Lahiri, 2010; Schild et al., 2012). The largest number of studies used discrimination designs in researching speech sound processing. In the investigation of directional asymmetries in infants' speech perception, paradigms like head-turn or preferential looking tasks (Bohn & Polka, 2001; Nam & Polka, 2016; Polka & Bohn, 1996) and visual habituation (Pons et al., 2012; Sundara et al., 2008) are often used. To study perceptual asymmetric effects in adults, different discrimination tasks have been used: Multiple studies employed same/different tasks where participants have to indicate if the pairwise presented speech sounds are perceived as the same or as different sounds (Bohn & Polka, 2001; Karypidis, 2007; Y. Liu et al., 2021; Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Masapollo et al., 2018; Polka, 1991; Scharinger, Merickel, et al., 2011; Scharinger, Domahs, et al., 2016). Other studies used AXB tasks including variations like XAB (Kriengwatana & Escudero, 2017; Polka & Bohn, 1996; Sundara & Polka, 2008; Tyler et al., 2014). Here participants have to decide whether sound X is the same as A or B (Gerrits & Schouten, 2004). Most of the research in directional asymmetries, however, implemented the so-called oddball paradigm (e.g., Cornell et al., 2011, 2013; Eulitz & Lahiri, 2004; Meng et al., 2021b, 2021a; Riedinger et al., 2021; Scharinger, Monahan, et al., 2012, 2016; Sundara & Polka, 2008; Tavabi et al., 2009; Yu & Shafer, 2021), or variations such as the roving standard design (Chládková et al., 2021; Nudga et al., 2021). The oddball paradigm, as part of the main methodology of this dissertation, will be further explained in section 1.2.

Directional asymmetries in speech processing do not only occur in different measures, tasks and designs but are also evident in different populations as well as languages. While impaired populations are understudied (Cummings et al., 2020), several studies found evidence for directional asymmetries in infants and children (e.g., Chládková et al., 2021; Cummings et al., 2020; Nam & Polka, 2016; Polka & Bohn, 1996, 2003a, 2011a; Pons et al., 2012; Tsuji & Cristia, 2017). However, the majority of evidence of asymmetric speech sound processing stems from adult participants (e.g., Eulitz & Lahiri, 2004; Fu & Monahan, 2021; Hestvik & Durvasula, 2016; Hwang et al., 2010; Kriengwatana & Escudero, 2017; Law et al., 2013; L. Liu et al., 2018; Masapollo, Polka, Molnar, et al., 2017; Nudga et al., 2021; Politzer-Ahles et al., 2016; Schluter et al., 2017; Yu & Shafer, 2021).

Similarly, directional asymmetries have been studied in a variety of languages. Typically, asymmetric effects in speech sound processing have been investigated in German (e.g., Cornell et al., 2011; Eulitz & Lahiri, 2004; Polka & Bohn, 1996; Riedinger et al., 2021; Scharinger, Bendixen, et al., 2012; Scharinger, Domahs, et al., 2016; Scharinger et al., 2010), English (e.g., Cummings et al., 2017; Hestvik & Durvasula, 2016; Y. Liu et al., 2021; Masapollo et al., 2015; Nam & Polka, 2016; Polka et al., 2021; Scharinger, 2006; Yu & Shafer, 2021; Zhao et al., 2019), and French (Karypidis, 2007; Y. Liu et al., 2021; Masapollo et al., 2018; Masapollo, Polka, & Ménard, 2017; Nam & Polka, 2016; Polka et al., 2021; Polka & Bohn, 1996; Tyler et al., 2014; Zhao et al., 2019). But there is also evidence from Turkish (Lipski et al., 2007; Scharinger, Idsardi, et al., 2011b), Czech (Chládková et al., 2021; Nudga et al., 2021), Chinese (Fu & Monahan, 2021; L. Liu et al., 2018; Meng et al., 2021b, 2021a; Politzer-Ahles et al., 2016), Japanese (Hestvik et al., 2020; Ikeda et al., 2002), Danish (Højlund et al., 2019), Arabic (Schluter et al., 2017), Russian (Schluter et al., 2017), Catalan/ Spanish (Pons et al., 2012) and Finnish (Savela et al., 2003).

In the last paragraphs, it became evident that some methods, measures, tasks, and designs are used more often than others. This dissertation will focus on the empirical investigation of German vowel processing using neural and behavioural measures in active and passive oddball paradigms. The methodology of the methods used and a broader introduction into the used design will be described in the next section (section 1.2).

1.2. Prominent methodology: Oddball-paradigm, Mismatch Negativity and Reaction Times

In this section, the design and measurements for the empirical investigations of directional asymmetries in German vowel processing of this dissertation will be introduced. Starting with a description of the oddball paradigm, the section also will explain the used ERP component Mismatch Negativity (MMN) as well as the relationship between the MMN as a neural measure and RT as a behavioural measurement.

The oddball paradigm is a design often used in the study of speech sound processing and, as described above, common in the investigation of directional asymmetries. In this paradigm, participants are confronted with a series of repetitive stimuli, i.e. standards, which are interspersed occasionally by a different stimulus, the so-called deviant (Duncan et al., 2009). In sound processing, the repetition of standards evokes a strong prediction and, furthermore, activates a memory trace and therefore the underlying representation of the speech sound in the mental lexicon (Dalebout & Fox, 2000; Deacon et al., 2000; Näätänen et al., 2007). The infrequently presented deviants are then compared on-line to the activated mental representations. Hence, deviants provide information about and are processed closer to the surface structure (Dalebout & Fox, 2000; Näätänen et al., 2007). Typically, the oddball paradigm uses one stimulus as standard and another one as deviant, which can vary by different dimensions from the standard. Here, single-item conditions are possible in contrasts to other experimental designs. Therefore, researchers commonly have a better control over physical and linguistic differences or confounding factors when using oddball designs. Furthermore, he major advantage of this design is that it can be used as a passive task. In a passive oddball design, participants do not need to focus on the stimulation and, therefore, no attention is needed. Participants typically watch a silent movie or read a book and are instructed to ignore the auditive stimulation. The passive design is for example useful when testing clinical populations. Nonetheless, it is still possible to use the oddball design in an active task where participants should indicate perceiving the deviants with a button press (Shtyrov & Pulvermüller, 2007). For an adequate signal-to-noise ratio, a large presentation rate of stimuli, especially the deviant, is needed (Duncan et al., 2009). Furthermore, the oddity and infrequency of the deviant plays a crucial role, i.e., the probability of standards and deviants. Usually, the deviant probability lies around 0.2 and the probability of standards around 0.8 respectively (Polich, 2007; Polich et al., 1994). When using the oddball paradigm for measuring neural effects, a stimulus-onset asynchrony (SOA) should be used. This jittered presentation supresses rhythmic processing best (Takegata, Syssoeva, et al., 2001). Studies showed that an unjittered presentation as well as very short interstimulus intervals (ISI) lead to a misgrouping of stimuli and reduced effects (Hartley et al., 2016; Ritter et al., 2006; Tervaniemi et al., 1994).

Although the use of the oddball paradigm has major advantages, there are some critical aspects. The passive oddball design is typically described as independent of the attention on the stimulation. Here, it is important to consider that attention control in this context is not synonymous with a guarantee that the participants really do not pay attention to the stimulation (c.f., Shtyrov & Pulvermüller, 2007, p. 184). Another critical point of the simplistic oddball design, i.e., using only one stimulus as standard and one as deviant, lies in the limited possibility of generalising the results since the stimulus set is typically very limited (Shtyrov & Pulvermüller, 2007).

While the first critical remark can hardly be avoided, it is possible to address the second point of criticism by adapting the design. By introducing some phonetic variation in standards and deviants, using for example several different recordings of the same speech sounds, a more complex and natural stimulus set can be obtained and therefore get a better generalisation of the results (Shtyrov & Pulvermüller, 2007). In this case, "only when all of the different tokens of each [stimulus] category are treated as identical it is possible to find a consistent difference between standards and deviants" (Phillips et al., 2000, p. 1043). Therefore, when introducing allophonic variation, the stimuli need to be grouped in phonological categories to maintain mental representations and to complete the discrimination task. Hence, it can be argued that allophonic variation activates a more abstract representation by blending out acoustic differences.

When investigating speech sound processing, the oddball paradigm is often used for collecting neural data by measuring EEG and analysing event-related potentials (ERPs). ERPs offer a means to investigate speech processing on a temporal axis with the accuracy of milliseconds. In combination with oddball tasks, the so-called Mismatch Negativity (MMN) is often found. The MMN has been first described by Näätänen and colleagues (1978) as a negative shift in the ERP wave form with high amplitudes and "that it reflects specific auditory stimulus-discrimination processes" (Näätänen et al., 1978, p. 327). Therefore, this component can be defined as a specific electrophysiological change detection response of the brain as a result of an on-line analysis of the differences between repetitive standards and infrequent, unpredictable deviants. The component, as the difference wave between standard and deviant, usually peaks at fronto-central and central electrode sites (Duncan et al., 2009; Näätänen et al., 2007) between 100 and 250 ms after the change onset and can be elicited by any discriminable change in the simulation (Fitzgerald & Todd, 2020; Näätänen, 2001). Typically, the amplitude of the MMN lies between 0.5 and 5 µV (Duncan et al., 2009). The amplitude and latency of this component can be influenced by stimulus-specific factors (Fitzgerald & Todd, 2020). Here, the main influence is the magnitude of deviation between standard and deviant: the MMN amplitude increases and latency decreases with increasing magnitude of deviation leading to easier discrimination (Duncan et al., 2009; Näätänen et al., 2007).

Since the component can be elicited without attention on the stimulation, the MMN reflects preattentive and automatic speech processing and, therefore, enables researchers to differentiate the neural responses to the stimuli without confounding attention effects and other perceptual and cognitive processes (for a review of the component, see Fitzgerald & Todd, 2020; and also Näätänen et al., 2007). The MMN is influenced by the stimulus-driven, bottom-up factor of discriminability of the differences between standard and deviant (Näätänen, 2001; Näätänen et al., 2007; Sams et al., 1985) but also by top-down factors mainly associated with prediction and prediction errors (Bendixen et al., 2012; Fitzgerald & Todd, 2020). In the context of predictive coding, the MMN can be interpreted as an index of prediction error because of the discrepancy between current input – the deviant - and the predictions made based on prior exposure – the standards (Fitzgerald & Todd, 2020).

MMN effects have been found in a variety of stimuli, for example in pure tones (Sams et al., 1985) with sensitivity for changes in frequency (Takegata & Morotomi, 1999; Tervaniemi et al., 2000), intensity and duration (Corcoran et al., 2018; Näätänen et al., 1989; Paavilainen et al., 1991), but also more complex stimuli like speech sounds (Dehaene-Lambertz, 1997; Dehaene-Lambertz et al., 2000). The complexity of the stimuli in the investigation of speech sound processing using the MMN range from isolated speech sounds like vowels (for example Deguchi et al., 2010; Grimaldi et al., 2014; Jacobsen, Schroger, et al., 2004; Näätänen et al., 1997; Rinker et al., 2010; Wottawa et al., 2022; Ylinen et al., 2006) or consonants (Wottawa et al., 2022), syllables (Bomba et al., 2011; Rivera-Gaxiola et al., 2000; Sittiprapaporn et al., 2005) to pseudowords or words (e.g., Aleksandrov et al., 2017; Henrich & Scharinger, 2022; Jacobsen, Horváth, et al., 2004; Pulvermüller et al., 2006; Riedinger et al., 2021).

The oddball task in combination with MMN measurements is suitable for and used in different populations like infants and children (Chládková et al., 2021; Dehaene-Lambertz & Baillet, 1998; Korpilahti et al., 2001; Rinker et al., 2010), as well as adults (for example Aaltonen et al., 1994; Deguchi et al., 2010; García & Froud, 2018; Kirmse et al., 2007; Lidji et al., 2010; M. S. Peltola & Aaltonen, 2005). But the component is also beneficial in clinical populations like aphasia patients (Aaltonen et al., 1993; Ilvonen et al., 2004), schizophrenia (Baldeweg et al., 2004; Corcoran et al., 2018) and hearing-impaired patients (Garrapa et al., 2013).

As described above, it is also possible to use this design as an active task measuring reaction times. Reaction times are associated with processing difficulties, i.e., with increasing task difficulty the reaction times usually increase es well. This means, the easier the discrimination, the faster the reaction times get. (Cutler et al., 1987; Pisoni & Tash, 1974; Savela et al., 2003; Zhang et al., 2019). Hence, reaction times indirectly reflect processing costs.

In the synopsis of the described methods, a link between MMN and RT can now be identified: With easier the discrimination between speech sounds, MMN amplitudes increase and MMN latency as well as reaction times decrease. This link between neural measures (MMN) and behavioural measures (RT) has been found in several studies (Bouwer & Honing, 2015; Brunellière et al., 2009; Deguchi et al., 2010; Dehaene-Lambertz et al., 2000; Frenck-Mestre et al., 2005; Kujala et al., 2001; Mager et al., 2005; G. Novak et al., 1992; G. P. Novak et al., 1990; Novitski et al., 2004; Savela et al., 2003; Tiitinen et al., 1994; Wang et al., 2012).

2. Models of asymmetric vowel processing

Different models exist for explaining asymmetric speech sound processing. Two models are prominent in the research literature regarding asymmetric vowel processing in healthy adults, namely the Featurally Underspecified Lexicon (FUL) model (Lahiri & Reetz, 2002a, 2010a) and the Natural Referent Vowel (NRV) framework (Polka & Bohn, 2011a). Both models have in common that they specifically predict directional asymmetries in speech sound discrimination and processing. However, the approaches in explaining asymmetries differ. Additionally, the hypotheses made by each model regarding neural and behavioural asymmetries are often contradicting. What separates the models furthermore are the substantially different assumptions about the features involved in discrimination processes and therefore in the mental representation of speech sounds. In this chapter, both models will be described in-depth starting with explaining the theoretical basis of the frameworks and the assumed features involved in speech sound processing. Afterwards, their approaches for predicting and explaining directional asymmetries will be explained, before existing empirical evidence for both models is pointed out.

2.1 Featurally Underspecified Lexicon

The first, for this dissertation important, model is the so called Featurally Underspecified Lexicon (FUL) model first proposed by Lahiri and Reetz (2002a). This model specifically predicts directional asymmetries in speech recognition on the basis of abstract, phonological features and through underspecification in the mental lexicon (Lahiri, 2018; Lahiri & Reetz, 2002a, 2010a).
Underspecification is a core concept in phonological theory. A key notion in most underspecification theories is the assumption that phonological features, as opposed to segments, are the smallest units in phonology (Archangeli, 1988, p. 193). Furthermore, especially in generative phonological approaches, two representations within the phonological system are assumed: on one hand the underlying representation – the phonological form – and on the other hand the surface structure – the phonetic form – which is the output of derivational processes from the underlying representation (Archangeli, 1988, p. 183; Monahan, 2018, p. 23; Steriade, 1995, p. 114). In most phonological theories, it is assumed that there is underspecification of phonological features in the underlying representation. Underspecification means that only necessary features are retained in the underlying representation while redundant or predictable feature are omitted and are inserted in the derivation of the surface form (Monahan, 2018, p. 23;

Steriade, 1995, p. 114). The basis of underspecification of features in the underlying representation lies in the concept of markedness (Archangeli, 1988; Mohanan, 1991; Steriade, 1995). Within the markedness concept, it is assumed that some feature values are for example universally more common, simpler, more predictable (Hume, 2004, p. 182). These features or feature values are then described as the unmarked value. Returning to the link between markedness and underspecification, it is assumed that unmarked values are omitted in the underlying representation because they are the most predictable. Marked values, on the other hand, are retained in the underlying phonological form (Archangeli, 1988, pp. 190, 192; Steriade, 1995, pp. 118–119).

The FUL model shares key concepts like the existence of abstract, phonological features as well as the notion of underspecification and minimalism with other underspecification theories. But in contrast to other theories like Radical Underspecification (Archangeli, 1988), the underspecification approach within FUL is not only a theoretical means to describe linguistic phenomena (e.g. assimilation) but it plays also a crucial role in speech sound processing and speech recognition. Therefore, FUL can be seen as a combination of phonological theory and cognitive implementation (Lahiri & Reetz, 2010a; Monahan, 2018). Starting with the assumption that unmarked values are underspecified not only in underlying representations but also in the mental representations of speech sounds as well as context-independent underspecification (Lahiri & Reetz, 2010a), the FUL model is characterised by the following additional basic assumptions regarding phonological features: Within FUL only monovalent features, as opposed to binary features, are used to describe speech sounds and their processing (Lahiri & Reetz, 2002a, 2010a). Additionally, the same features are used to describe vowels and consonants (Kotzor et al., 2017, p. 86; Lahiri & Reetz, 2010a). Last but not least, within the FUL feature hierarchy no feature dependencies are assumed, i.e., no feature dominates other features. Features are only dominated by their respective nodes (Kotzor et al., 2017, p. 88; Lahiri, 2018, p. 235). For vowel perception and processing, it is assumed that the place node and the subsequent articulator and tongue height nodes are important. The articulator node refers to the horizontal dimension of the vocal tract and contains the features [LABIAL], [CORONAL], and [DORSAL] (abbreviated in future as [LAB], [COR], and [DOR]). [LAB] is attributed to rounded vowels like [y], [COR] is used for front vowels, for example [i], and [DOR] for back vowels, e.g., [u] (Lahiri & Reetz, 2010a). While [COR] and [DOR] are mutually exclusive, [LAB] can be seen as an additive feature which is either present or absent and does not stand in conflict with the other features of the articulator node (Lahiri, 2018, p. 235). The tongue height node, on the other hand, refers to the vertical dimension of the vocal tract. This node dominates the features [HIGH] for high vowel, e.g., [i], and [LOW] for low vowels, for example [a].

Additionally, it is important to note regarding the underspecification of FUL, that there is universal as well as language-dependent underspecificatio. For instance, [COR] and [PLOSIVE] are assumed to be the only universally underspecified features (Lahiri, 2018, p. 236) since both feature are available in all languages and therefore can be described as "default" features for language production. Which features are additionally underspecified in the underlying representation and in the mental lexicon is language-dependent - depending on whether they are contrastive in a given language (Kotzor et al., 2017, p. 87). There is evidence for the language-dependent underspecification of FUL for different languages as well as dialects (Hestvik et al., 2020; Lipski et al., 2007; Scharinger & Idsardi, 2014; Scharinger & Lahiri, 2010).

Turning to the prediction and explanation of directional asymmetries within FUL, it is important to note that featural underspecification is assumed only in the underlying and mental representation. Therefore, there can be a discrepancy between the features contained in the signal and those stored in the mental lexicon, because the underspecified mental representation does not contain all possible features. Hence, the framework proposes a specific ternary mapping processes between incoming speech signal, which can contain all possible features, and mental representations, where underspecified features are omitted. Due to the underspecification of redundant features, there are three possible outcomes of the mapping process: a match occurs if the feature extracted from the speech signal has the same equivalent feature in the mental representation (e.g., [DOR] – [DOR]: [u] - [o]). A mismatch between the features in the signal and those stored in the mental lexicon occurs if they are complementary and exclude each other (for example [HIGH] – [LOW]: [i] – [a]). Finally, a no-mismatch occurs if the feature extracted from the signal neither mismatches nor matches with an activated feature on the mental lexicon. No-mismatches are crucial in the elicitation of directional asymmetries. For instance, if [COR] is extracted from the incoming speech signal, it produces a mismatch with the feature [DOR], if it is activated in the mental lexicon. The other way round, if [DOR] is extracted from the signal, the result is a no-mismatch due to the underspecification of the coronal place of articulation in the mental lexicon ([–]). These different results usually become apparent when the discrimination of two speech sounds is tested bidirectionally ([i] – [u] vs. [u] – [i]) (Kotzor et al., 2017; Lahiri & Reetz, 2002a, 2010a). Hence, directional asymmetries are a fundamental principle, according to FUL, for the speech recognition process. The incorporation of asymmetric mapping processes is assumed to lead to more effective and easier sound discrimination and speech recognition (Kotzor et al., 2017, p. 94). These asymmetric mapping processes should also be noticeable in neural and behavioural measures. FUL, therefore, makes specific hypotheses regarding asymmetric MMN and RT effects: A mismatch result is accompanied by stronger effects leading to larger MMN amplitudes, faster MMN latencies and faster RTs than matches or no-mismatches (Kotzor et al., 2017, p. 94).

These links between featural mapping processes proposed by FUL and asymmetric neural and behavioural effects have been shown in several studies. Regarding German vowel processing, Eulitz and Lahiri (2004) conducted a MMN study with German vowels [o], [ø], and [e]. These vowels differ mainly in place of articulation (PoA). When discriminating [o] – [ø], a larger and faster MMN occurred because of mismatching features [DOR] and [COR]. In the reverse presentation order, the effects were diminished due to the underspecification of [COR] in the mental lexicon. Similar results were obtained by Scharinger and colleagues (Scharinger, Monahan, et al., 2012) for tongue height oppositions in American English vowels. Here, the mid of tongue is believed to be underspecified. They found larger effects if the mid vowel [ɛ], as the second presented sound, had to be discriminated from low vowel [æ] due to a mismatch of features [LOW] and [MID], as opposed to the reversed direction where a no-mismatch occurs. Similar evidence, in favour of the FUL approach has been found not only for vowels (e.g., de Jonge & Boersma, 2015; de Rue et al., 2021; Nudga et al., 2021; Scharinger, Idsardi, et al., 2011b; Scharinger, Monahan, et al., 2016; Yu & Shafer, 2021) but also consonants (Cummings et al., 2017, 2020, p. 2; Fu & Monahan, 2021; Hestvik et al., 2020; Hestvik & Durvasula, 2016; Højlund et al., 2019; Hwang et al., 2010; Scharinger, Bendixen, et al., 2012; Schluter et al., 2017), words or pseudowords (Cornell et al., 2013; Kotzor et al., 2020; Lahiri & Reetz, 2010a; Riedinger et al., 2021; Scharinger, 2006; Tavabi et al., 2009) as well as lexical tones (Law et al., 2013; Li & Chen, 2015; Politzer-Ahles et al., 2016).

Additionally, evidence for speech processing using sparse, abstract phonological features in accordance with FUL has been found in a variety of languages: German (for example Eulitz et al., 2003; Eulitz & Lahiri, 2004; Obleser et al., 2004; Riedinger et al., 2021; Scharinger, Domahs, et al., 2016; Scharinger et al., 2010; Schild et al., 2012), English (e.g., Cummings et al., 2020; Hestvik & Durvasula, 2016; Kotzor et al., 2020; Scharinger, Merickel, et al., 2011; Scharinger, Monahan, et al., 2012; Scharinger & Idsardi, 2014; Scharinger & Lahiri, 2010), Turkish (Lipski et al., 2007, p. 20; Scharinger, Idsardi, et al., 2011a, 2011b), Chinese languages (Fu & Monahan, 2021; Li & Chen, 2015; Meng et al., 2021b, 2021a; Politzer-Ahles et al., 2016) as well as French (de Jonge & Boersma, 2015), Dutch (de Rue et al., 2021), Danish (Højlund et al., 2019), Arabic (Schluter et al., 2017), Russian (Schluter et al., 2017) and Japanese (Hestvik et al., 2020).

2.2. Natural Referent Vowel

The second, for this dissertation important, model is the Natural Referent Vowel (NRV) framework proposed by Polka and Bohn (2011a). This framework predicts directional asymmetries, especially in vowel processing, based on phonetic features of the involved vowels leading to hypotheses that often contradict those postulated by the FUL model.

The NRV framework is rooted in two key concepts: dispersion and focalisation. First, within the dispersion theory, it is assumed that quantal vowels /i/, /a/ and /u/ are universally favoured in vowel inventories because of their intrinsic properties, i.e., they are maximally perceptually distant to each other and perceptually prominent within vowel systems. This perceptual prominence can be interpreted as higher salience of the quantal vowels (Lindblom & Engstrand, 1989). On the basis of dispersion within vowel systems Schwartz and colleagues (Schwartz et al., 1997a) proposed the addition of focalisation leading to the dispersion-focalisation theory. Here, it is stated that quantal vowels are perceptually more salient due to the convergence of adjacent formants leading to more energy in a spectrally narrower space. Therefore, they assume that focal vowels are the most prominent points in vowel spaces leading to a higher salience of focal vowels (Schwartz et al., 1997a, 2005). While /i/, /a/ and /u/ usually are described as quantal vowels (Lindblom & Engstrand, 1989; Pisoni, 1980), /i/, /a/, /u/ and /y/ have been characterised as the most focal vowels within vowel systems (Masapollo et al., 2019; Schwartz et al., 1997a, 2005).

These two characterisations of vowel prominence together with observations of directional asymmetries in vowel discrimination are combined in the NRV model. The first proposition of this models refers to the perceptual salience of the vowel space periphery leading to easier discrimination towards more peripheral vowels (/e/ - /i/) than vice versa (/i/ - /e/) (Bohn & Polka, 2001; Karypidis, 2007; Kriengwatana & Escudero, 2017; Polka & Bohn, 1996, 2003a, 2011a; Pons et al., 2012; Tsuji & Cristia, 2017; Tyler et al., 2014). The second proposition incorporates focalisation, claiming higher salience of focal vowels. Here, discriminations towards more focal vowels are easier than the other way round (Y. Liu et al., 2021; Masapollo et al., 2018; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019, 2021; Zhao et al., 2019). However, it remains unclear how focalisation should be computed: “To advance in this direction we need an applicable metric for quantifying vowel focalization that reveals how this parameter is computed and processed at the behavioural and neural level” (Polka & Bohn, 2011a, p. 475; see also Simon et al., 2014). Furthermore, directional asymmetries have been mostly found when investigating within-category vowel processing (e.g., different instance of

/u/) (Y. Liu et al., 2021; Masapollo et al., 2015, 2018; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019, 2021). Whereas, when investigating across-category discrimination, the found directional asymmetries were usually based on the peripherality approach (Kriengwatana & Escudero, 2017; Polka & Bohn, 1996, 2003a, 2011a, 2011a; Pons et al., 2012; Tsuji & Cristia, 2017; Tyler et al., 2014). Therefore, the question arises if peripherality is used to discriminate across-category contrasts and whether focality only plays a crucial role when discriminating fine-grained within-category differences. Although this question is certainly important, it needs to be explored in future research and is therefore not the focus of this dissertation.

Nevertheless, within the NRV framework, discrimination towards peripheral or more focal vowels leads to easier discrimination accompanied by stronger effects, e.g., larger MMN amplitudes, faster MMNs and faster RT (Polka & Bohn, 2011a).

The general salience of the vowels space's periphery and focal vowels can be influenced by language experience. Initially, the NRV model was developed on basis of infants' easier discrimination towards peripheral vowels (Polka & Bohn, 2003a) while in adults only non-native contrasts evoked such directional asymmetries. In native contrast, the universal bias of vowel space periphery can be overridden due to language specific requirements and adaptations (Polka & Bohn, 2011a). However, there is evidence that directional asymmetries proposed by NRV are also present in native contrasts in adults (Masapollo, Polka, Molnar, et al., 2017) leading to an interpretation as universal bias in vowel processing.

Empirical evidence in favour of the NRV framework stems mostly from (isolated) vowel processing (Y. Liu et al., 2021; Masapollo et al., 2015, 2018; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019, 2021; Tsuji & Cristia, 2017; Zhao et al., 2019). However, there is some evidence for phonetic directional asymmetries in consonants (Nam & Polka, 2016; Polka, 1991), words (Simon et al., 2014) and even lexical tones (Masapollo et al., 2019; Wayland et al., 2019). Here, several studies showed directional asymmetries towards peripheral and/ or focal vowels in infants (Bohn & Polka, 2001; Nam & Polka, 2016; Polka, 1991; Polka & Bohn, 1996, 2003a; Pons et al., 2012; Tsuji & Cristia, 2017) and adults (Karypidis, 2007; Y. Liu et al., 2021; Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Masapollo et al., 2018; Polka et al., 2021; Polka & Bohn, 2011a; Tyler et al., 2014; Zhao et al., 2019). Furthermore, NRV based directional asymmetries have been found in different languages, mostly in English (e.g., Y. Liu et al., 2021; Masapollo, Polka, Molnar, et al., 2017; Nam & Polka, 2016; Polka & Bohn, 1996, 2011a; Zhao et al., 2019), French (for example Y. Liu et al., 2021; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka,

& Ménard, 2017; Tyler et al., 2014) and German (Bohn & Polka, 2001; Polka & Bohn, 1996, 2011a). But there is also evidence form Thai and Norwegian (Tyler et al., 2014), Catalan and Spanish (Pons et al., 2012), Dutch (Kriengwatana & Escudero, 2017) as well as Mandarin (Wayland et al., 2019).

In summary, it has been shown that directional asymmetries are evident in different languages, populations, and measures. Both models described in this chapter propose directional asymmetries as important mechanisms in vowel processing. However, the models presented use different approaches to explain directional asymmetries and the hypotheses are often contradicting. While the FUL model employs sparse and abstract phonological features, the NRV framework is based on phonetic, and often more fine-grained, cues. This dissertation, therefore, sets out to empirically test both models in German vowel processing to draw conclusions on whether vowel processing is completed with either phonetic or phonological features.

III Empirical investigations of asymmetric effects in vowel processing

The first part of the chapter (section 1) is in large parts already published as: Riedinger M., Nagels A., Werth A. and Scharinger M. (2021): Asymmetries in Accessing Vowel Representations Are Driven by Phonological and Acoustic Properties: Neural and Behavioral Evidence From Natural German Minimal Pairs, *Frontiers in Human Neuroscience*, 15:612345. A later addition to this published series of studies is the latency analysis of the neural data. While the first set of experiments, described in section 1, investigated vowel processing when embedded in real German words, the second part of the chapter focuses on vowel processing of isolated German long vowels. This part, section 2, contains unpublished data.

1. Neural and behavioural asymmetries with German minimal pairs

In this series of studies, we tested both competing models (see chapter I) with German vowels to investigate which model can best explain directional asymmetries. Consequently, we tested the predictions of both models on a large stimulus set (five German long vowel contrasts) in a more natural listening situation by using real minimal pairs. The study was based on the following notions:

The use of real words in MMN investigations is associated with obstacles due to interference, such as lexical status, familiarity, or other confounding factors. In several studies, it has been shown that MMN responses for real word deviants are enhanced in comparison to pseudowords: It is believed that the enhancement of the lexical MMN is due to stronger memory trace activation for real meaningful words (Endrass et al., 2004; Pettigrew, Murdoch, Ponton, et al., 2004; Pulvermüller et al., 2001, 2004; Shtyrov et al., 2008; Shtyrov & Pulvermüller, 2002). Another known influential factor on speech processing is the lexical frequency of the used real words. This influence can even be present when testing real words in an passive oddball paradigm and can lead to a stronger MMN response for words with higher lexical frequency in opposition to deviants with lower or intermediate frequency of occurrence (Aleksandrov et al., 2011; Shtyrov et al., 2011). Furthermore, it has been shown that phonotactic probabilities (sequential order of phonemes in words) have an influence on MMN results with higher probability, accompanied by enhanced MMN effects (Bonte et al., 2005; Emmendorfer et al., 2020, 2021; Yasin, 2007). With respect to vowel perception, acoustic properties, like for example fundamental frequency, vowel duration or intensity, (Aaltonen et al., 1994; Kirmse et al., 2007; Partanen et al., 2011; Peter et al., 2010), have an impact on neural effects. While some of the mentioned influential factors can be controlled for when developing stimulus materials, others are not avoidable. For instance, various acoustic differences in vowels stem from collinearities between vowel identity and acoustic consequences. Changes in vowel identity simultaneously lead to changes in the spectral frequency structure (mainly F1 and F2) of the stimuli. Moreover, vowel features used in theoretical frameworks are based largely on articulatory-acoustic properties – mainly formants – and therefore, it could also be possible that there is more of an acoustical influence, especially in MMN effects, than proposed by operationalizing more abstract and theoretical derived features. For instance, the feature opposition of [HIGH] and [LOW] is the more abstract representation of the articulatory and acoustic properties of those vowels with regard to the first formant: High vowels have a low F1, while low vowels have a high F1 (Lahiri & Reetz, 2010a). Also, the abstract description of vowels referring to focality which were used in the NRV framework is based on articulatory-acoustic properties, since focalization stems from the convergence of adjacent formants (Schwartz et al., 1997b). While the common contributing articulatory-based factors of vowel perception have often been investigated, the influence of perceptual and psychoacoustic parameters (e.g., perceived loudness) on vowel perception has hardly been studied. Thus, we wanted to additionally investigate which were the influential factors on vowel discrimination, including not only theoretical and acoustical factors but also perceptual factors beyond the well-known.

Hence, the following research questions shall be investigated: (1) which model accommodates directional asymmetries in processing of natural and unmanipulated German long vowels in the best way (**model comparison**), and (2) which factors are influencing vowel discrimination in natural German minimal pairs (**influential factors**)? The first question has been addressed on an electrophysiological level through measurement of MMN (experiment 1a) and on a behavioural level in means of reaction times (RT) (experiment 1b). The second aim of identifying influential factors on vowel discrimination, pursued via multiple regressions on both datasets, should shed more light on factors that co-determine MMN effects.

1.1 Experiment 1a: identity MMN

To test both models, we first conducted an MMN study with a large stimulus set, testing five German long vowel contrasts embedded in natural minimal pairs, which almost mapped the entire German (long) vowel space. The vowels chosen for investigation were among the most frequent long vowels in the German language (Aichert et al., 2005).

1.1.1 Participants

Nineteen participants (9 female, mean age 24.7, SD 3.4), graduate and undergraduate students at the Philipps University of Marburg, participated in two sessions for monetary compensation. They were all right-handed and reported no hearing or neurological impairments. All participants were monolingual German native and German Standard speakers without being able to speak any German dialect actively. They were all born and socialized in Hesse, Germany, with Standard German. The information about the participants' dialect and Standard German competence was retrieved by questionnaire. Informed written consent was obtained from each participant prior to the experiment. One subject had to be excluded because the participant missed the second session. Another subject had to be excluded due to excessive contamination with artifacts in the EEG data (movement artifacts). In total, we assessed and analysed the complete data of seventeen participants.

1.1.2 Materials

To test the hypotheses of the aforementioned models, we chose the five German long vowel contrasts /i:/ – /e:/, /e:/ – /a:/, /y:/ – /u:/, /i:/ – /u:/, and /i:/ – /a:/. They differed with respect to place of articulation, vowel height as well as rounding. To ensure more phonological processing, we embedded these vowels in German monosyllabic minimal pairs. We tried to keep the phonetic context between pairs as similar as possible. We also controlled for frequency of occurrence with SUBTlex (Brysbaert et al., 2011), as seen in Table 1.

Table 1: Phonetic and lexical parameters of the vowels.
Mean f0, F1, F2, and F3 values are given in Hertz for vowels per word category. Mean Intensity for the vowels within the words are given in dB. Mean duration measures referring to the vowels within the words. Frequency of occurrence (as log-values) for each word is given in the last column.

Vowel contrast	Words	Mean f0 (SD)	Mean F1 (SD)	Mean F2 (SD)	Mean F3 (SD)	Mean Intensity (SD)	Mean duration (ms)	Word frequency (log-values)
/a:/ - /i:/	*Zahl* (“number”)	169 (24)	875 (19)	1488 (59)	3046 (77)	72.85 (0.27)	292	2,861
	Ziel (“target”)	203 (5)	244 (28)	2445 (61)	3446 (74)	73.79 (0.92)	159	3,358
/e:/ - /i:/	*Steg* (“bridge”)	179 (10)	359 (6)	2483 (58)	3171 (113)	73,96 (0.67)	270	1,255
	Stieg (“climbed”)	190 (3)	286 (11)	2479 (51)	3528 (21)	74.92 (0.31)	228	2,352
/a:/ - /e:/	*Mahl* (“meal”)	174 (5)	913 (10)	1484 (47)	2966 (44)	71.68 (0.22)	225	1,672
	Mehl (“flour”)	191 (3)	341 (8)	2566 (39)	3377 (209)	71.74 (0.26)	195	1,857
/u:/ - /i:/	*Stuhl* (“chair”)	200 (8)	294 (37)	1974 (212)	2647 (99)	73.88 (0.44)	186	2,892
	Stiel (“handle”)	199 (6)	275 (22)	2475 (75)	3579 (59)	73.67 (0.44)	176	1,756
/u:/ - /y:/	*Sud* (“brew”)	194 (2)	296 (40)	1145 (140)	2628 (93)	74.75 (0.20)	194	0,845
	Süd (“south”)	180 (5)	292 (14)	2085 (120)	2633 (146)	74.43 (0.43)	223	1,771

Twenty natural exemplars of each word were recorded in a sound shielded booth by a female German Standard speaker who was phonetically trained. All tokens were spoken with neutral pronunciation. All sounds have been analysed for f0, F1, F2, F3, as well as vowel duration and were all scaled to an intensity level of 70 dB within Praat (Boersma & Weenink, 2016). The five best tokens per word have been chosen as experimental stimuli. Phonetic parameters of the word categories are displayed in Table 1. Note that we reported here only mean values per word category since MMN and RT data are also averaged measures. There can be seen that our stimuli had some variance regarding, for example, vowel duration. Since we wanted to test natural spoken words, no manipulation was applied. All vowels should be perceived as long vowels despite the length differences since the phonological category is additionally supported by the lexical context. Therefore, the focus in processing lies on categorical differences regarding vowel height and place of articulation. All experimental stimuli were found to sound natural by two different persons. All tokens were also assessed as being distinct for their category (see Figure 1).

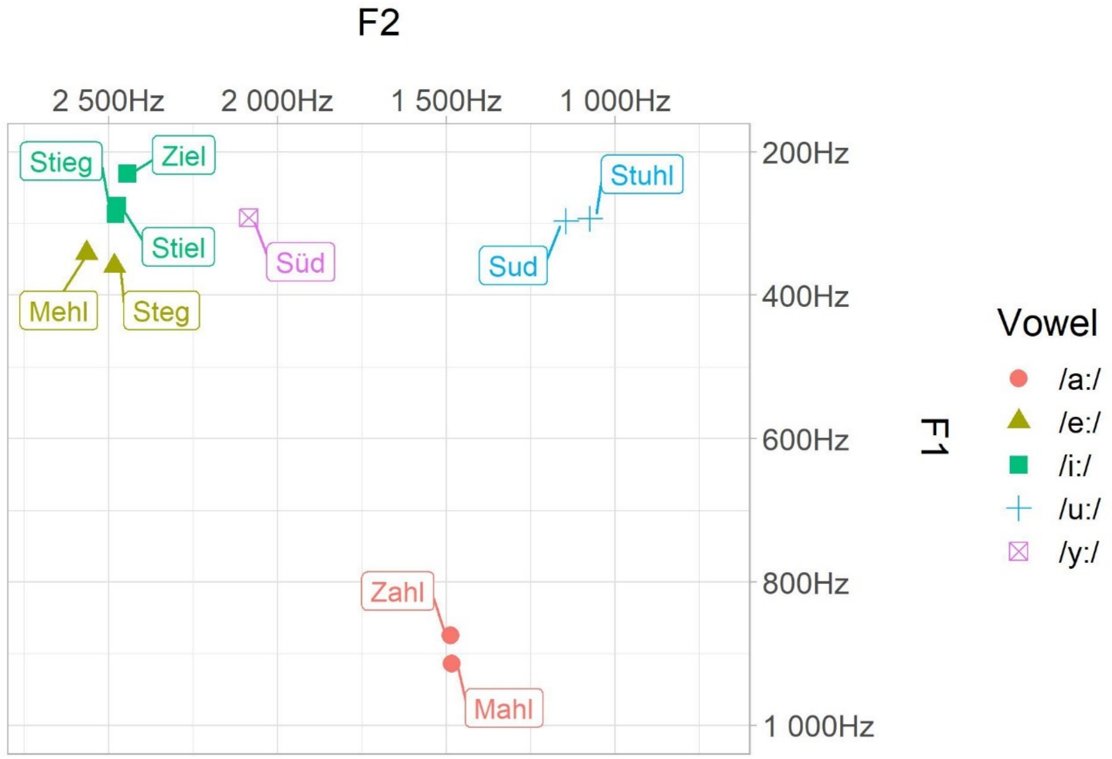

Figure 1: Acoustic characteristic of the stimuli.
Mean values of the first (F1) and the second (F2) formant are given per word category in Hertz.

We compared the formant values (F1, F2) to the ones of Sendlmeier & Seebode (2006) to ensure that they will be perceived as Standard German. We chose to introduce inter-token variation to obtain a more natural listening situation and to ensure a more phonological approach since participants are forced to map the incoming variable acoustic signals onto a unified and more abstract representations in order to cope with inter-token variability (Eulitz & Lahiri, 2004; Jacobsen, Horváth, et al., 2004; Phillips et al., 2000). This is an important design feature since it mediates the likely collinearity between formant frequencies and acoustic- or articulatory-phonetic features.

1.1.3 Task and Procedure

The stimuli were embedded in a passive oddball design. In this paradigm, the participants were presented with a series of repetitive stimuli (standards) that were interspersed occasionally by a deviant varying only in vowel quality while they were watching a silent movie. The frequently presented standards were assumed to activate the memory trace and therefore the representation in the mental lexicon, whereas the infrequently presented deviants provided information about and are processed closer to the surface structure. Each vowel contrast was tested bidirectionally. Because we investigated five contrasts in both directions, all subjects were tested in two sessions (with testing times per session approximating two hours) within fifteen to twenty days. Thus, each word served as standard and as deviant in different blocks and sessions.
Each contrast direction was presented in two blocks containing 425 standards and 75 deviants each. In total, we presented 850 standards and 150 deviants per contrast direction. Thus, we presented 2000 stimuli for each vowel contrast. Within the blocks, stimuli were randomized, and the interval between two deviants randomly consisted of four to eleven standards. Blocks were randomized for both sessions. Blocks of the same condition never succeeded each other. The fixed ISI was 1000 ms, while the stimuli varied in duration. Therefore, we still obtained a jittered presentation to suppress rhythmic processing and habituation to synchronously presented stimuli.
Subjects were seated comfortably in a sound insulated and electromagnetically shielded chamber in front of a screen. Sounds were presented binaurally at a comfortable listening level via two loudspeakers on the left and the right of the screen, using the open source software OpenSesame (Mathôt, et al., 2012). The listening level was set prior to the experiment and was kept equal across all subjects (based on the intensity level of 70 dB as manipulated in PRAAT).

1.1.4 Hypotheses

Since this paper aims at comparing the two aforementioned models, hypotheses have been made on the assumptions of vowel discrimination in accordance with FUL as well as NRV for identity MMN amplitudes and latency data. For NRV, we proposed the hypothesis based on the universal assumptions of the framework that vowels /i:/, /y:/, /u:/, and /a:/ are reference vowels (Polka & Bohn, 2011a). The basic assumptions of both models regarding feature specifications and position in the vowel space for each investigated contrast are displayed in Table 2.

Table 2: Hypotheses.
Assumptions for feature specifications (FUL) and location in the vowel space (NRV) for each vowel contrast and hypotheses for MMN effects according to both models.

Vowel contrast	Presentation order	Features (FUL)	Mapping result (FUL)	Expectations MMN (FUL)	Classification (NRV)	Expectations MMN (NRV)
/a:/ - /i:/	*Zahl – Ziel*	[DOR] – [COR] [LOW] – [HIGH]	Mismatch Mismatch	Stronger effect	Both peripheral	Symmetrical effect
	Ziel – Zahl	[–] – [DOR] [HIGH] – [LOW]	No-mismatch Mismatch	Weaker effect		
/e:/ - /i:/	*Steg – Stieg*	[–] – [HIGH]	No-mismatch	Weaker effect	Central – peripheral	Stronger effect
	Stieg – Steg	[HIGH] – [MID]	Mismatch	Stronger effect	Peripheral – central	Weaker effect
/a:/ - /e:/	*Mahl – Mehl*	[DOR] – [COR] [LOW] – [MID]	Mismatch Mismatch	Stronger effect	Peripheral – central	Weaker effect
	Mehl – Mahl	[–] – [DOR] [–] – [LOW]	No-mismatch No-mismatch	Weaker effect	Central – peripheral	Stronger effect
/u:/ - /i:/	*Stuhl – Stiel*	[DOR] – [COR]	Mismatch	Stronger effect	Both peripheral	Symmetrical effect
	Stiel – Stuhl	[–] – [DOR]	No-mismatch	Weaker effect		
/u:/ - /y:/	*Sud – Süd*	[DOR] – [COR]	Mismatch	Stronger effect	Both peripheral	Symmetrical effect
	Süd – Sud	[–] – [DOR]	No-mismatch	Weaker effect		

In the amplitude data, the hypotheses are the following (see also Table 2): Within FUL, the effects should be stronger for mismatching presentation orders /a:/ - /i:/, /u:/ - /y:/ and /u:/ - /i:/ (because of the mismatching features [DOR] and [COR]), /i:/ - /e:/ (due to mismatch of [HIGH] and [MID]) as well as /a:/ - /e:/ (mismatch of [DOR] – [COR] and, [HIGH] – [MID]). If NRV holds true, MMN effects should be stronger when *Stieg* and *Mahl* are deviants since they are referent vowels in this model. In the other three contrasts, a symmetry should occur since both vowels are peripheral and act as referents within the framework.

Regarding the latency data, we follow the general assumption in the literature that stronger amplitudes are usually accompanied by shorter latencies whether asymmetries occur due to underspecification (Eulitz & Lahiri, 2004) or peripherality (Verá-Constán & Sebastián-Gallés, 2008, as cited in Polka & Bohn, 2011). Accordingly to FUL, shorter latencies should occur for mismatching presentation orders /a:/ - /i:/, /u:/ - /y:/, /u:/ - /i:/, /i:/ - /e:/, and /a:/ - /e:/.

1.1.5 EEG Recording and Analysis

EEG was recorded with 28 Ag/AgCl passive electrodes connected to a BrainAmp amplifier (Brainproducts GmbH). Electrodes were arranged on an EasyCap in 10-20 positions. AFz served as Ground electrode, and the online reference was placed on the nose tip. Four additional electrodes measured the electrooculogram (EOG) for the identification of artifacts caused by eye movements (e.g., blinks). Two electrodes were placed left and right of the eye canthi to measure lateral eye movements. Two electrodes above and under the right eye measured vertical eye movements. For all electrodes, impedances were kept below 5 kΩ and the sampling rate was 500 Hz.

EEG analysis was done with the Matlab toolbox fieldtrip. Raw data were filtered with 0.16 Hz and 30 Hz high- and lowpass filters. Data were re-referenced offline to linked mastoids. After segmentation, EEG data were automatically corrected for muscle artifacts. Eye movements were automatically corrected through correlation of EOG channels and ICA components. The calculation of the MMN component was based in relation to the onset of the vowel, i.e., epochs beginnings were aligned with vowel beginnings. Thereby, consonant onset clusters in the stimuli should play no role in the MMN effects. Additionally, ERP data were baseline corrected using the 100-ms prestimulus epoch.

For averaging, the first ten standards of a block and the first standard after a deviant were excluded from data analysis. To maintain ERP results without the influence of pure acoustic influences, we calculated and plotted the MMN as identity MMN (iMMN). Here, the standard and the deviant of the same word are compared to each other (Pulvermüller et al., 2006).

1.1.6 Results

The results of the iMMN study are plotted in Figure 2.

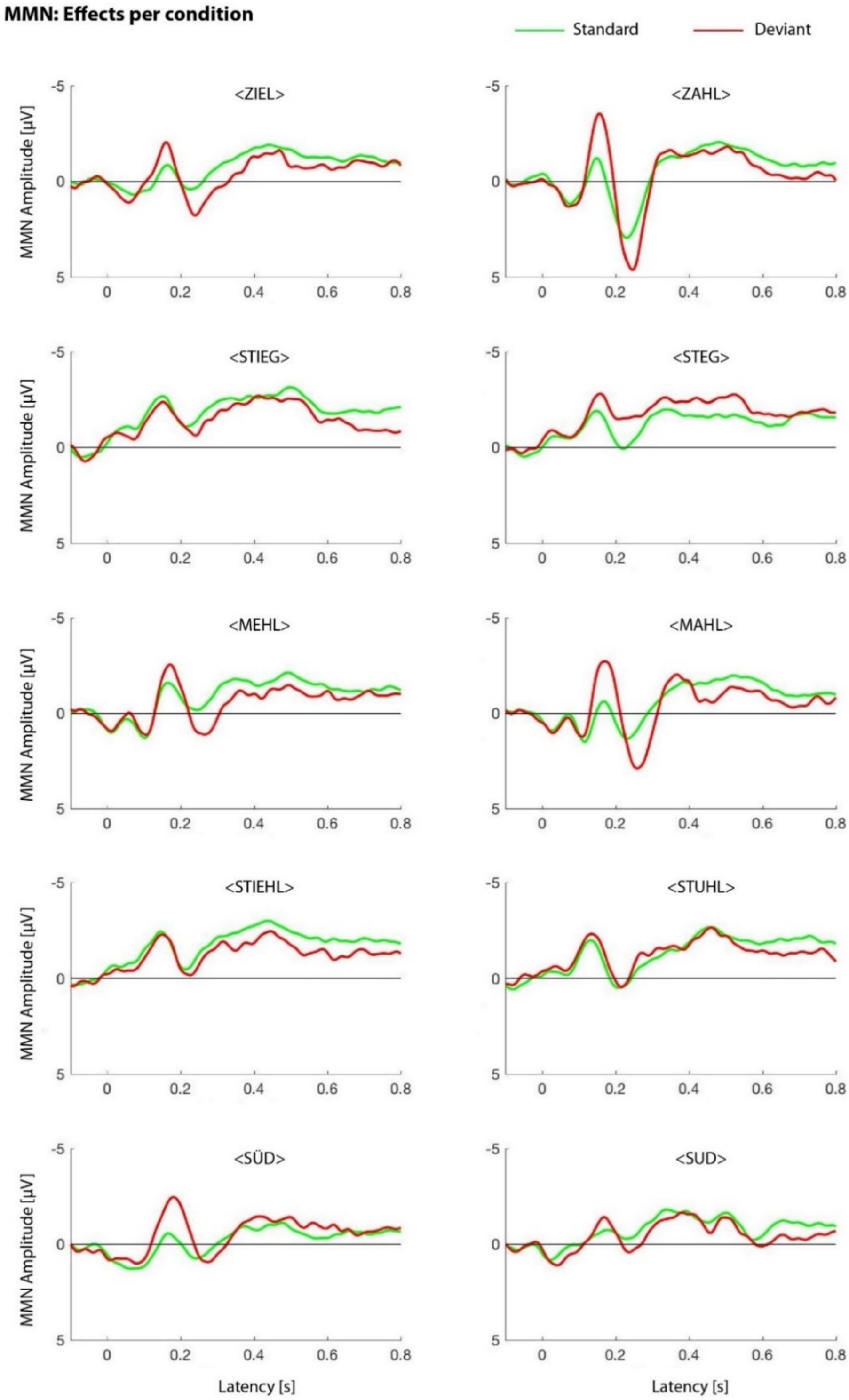

Figure 2: Identity MMN effects per condition.
MMN waveforms for all word pairs, in both presentation orders, are shown.

In a first step, we were interested in significant standard-deviant differences in the auditory evoked potentials. To this end, we employed a conservative measure of amplitude contrasts without prior assumptions of regions of interest and followed a multilevel statistical approach (e.g., Henry & Obleser, 2012; Strauß et al., 2014). At the first level, we calculated independent-samples t-tests between the single-trial amplitude values of standards and deviants. Uncorrected by-participant t-values were obtained for all time-amplitude bins of all electrodes. At the second level, t-values were tested against 0 with dependent-sample t-tests. Taking into consideration the problem of multiple comparisons, a Monte-Carlo nonparametric permutation method with 1000 randomizations, as implemented in fieldtrip (Oostenveld et al., 2011), estimated type I-error controlled cluster significance probabilities (at $p < .05$). In an electrode x time cluster (with Fz, Cz, CPz, between 130 and 200 ms post vowel onset), deviants elicited a significantly more negative response than standards (see Fig. 3).

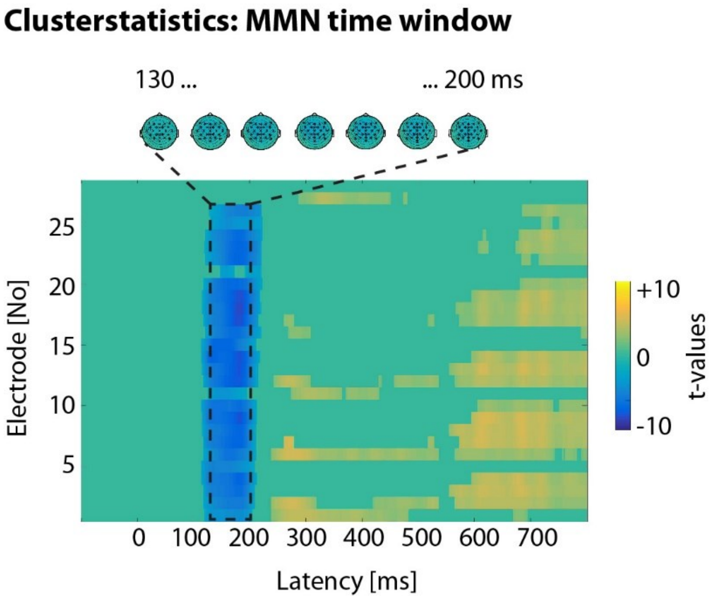

Figure 3: Clusterstatistics.
In an electrode x time cluster, deviants elicited more negative responses than standards in the time window between 130 and 200 ms post vowel onset.

To analyse the EEG data for directional asymmetries, we calculated the iMMN as difference waves (deviant minus standard of the same words) in this aforementioned time window. Then, we calculated repeated-measures and Bonferroni corrected ANOVAs for each contrast with factors *word* (e.g., *Ziel* vs. *Zahl*) and *electrode* (Fz, Cz, CPz) for identity MMN amplitudes. Electrodes were chosen by cluster statistics.

In the /i:/ – /a:/ contrast, we found both main effects for *word* ($F(1,16) = 7.286$, $p = .016$) with a larger MMN for *Zahl* (M = -1,721, SEM = 0.36; Ziel: M = -0.586, SEM = 0.253) and for *electrode*

(F(2,32) = 14.634, p < .001) with strongest effect at Cz (F(1,16) = 5.890, p < .05). In the vowel contrast /e:/ – /a:/, there was not only a highly significant main effect for *electrode* (F(2,32) = 12.307, p < .001) but also an interaction *word* x *electrode* (F(2,32) = 4.942, p = .013). Post-hoc analysis of the interaction showed a significant effect on Cz (F(1,16) = 5.039, p < .05). Hence, we found asymmetries in visual inspection as well as in statistical analysis in both contrasts. The comparison of /u:/ – /y:/ only revealed a main effect of factor *electrode* (F (2,32) = 12.349, p < .001) with a marginal effect on CPz (F(1,16) = 4.265, p = .055). Therefore, comparing *Süd* and *Sud*, we found an asymmetry in the visual inspection, which did not hold out statistical analysis. Hence, statistically we found a symmetrical effect.

The vowel contrast /i:/ – /e:/ shows a symmetrical pattern in visual inspection and statistics with both main effects insignificant (word: F(1,16) = 1.687, p =.212, electrode: F (2,32) = 2.367, p = .110). The same is also true for the comparison of /i:/ – /u:/ (word: F (1,16) = 0.294, p = .595, electrode: F (2,32) = 0.725, p = .492).

Additionally, we conducted after publication an analysis of the peak latency of the identity MMN to test whether asymmetric effects due to underspecification or peripherality can be seen in this measure. There is especially evidence for underspecification effects in latency data (Eulitz & Lahiri, 2004). A linear mixed model, calculated in jamovi with the package gamlj (Gallucci, 2019; The jamovi project, 2021), with factors *contrast* and *reversal* produced only a significant main effect for *reversal* (F = 4.52, p < .05), systematically occurred asymmetries should have resulted in an interaction effect (see figure 4). Therefore, we found no systematic effect of the FUL or NRV model in our latency data.

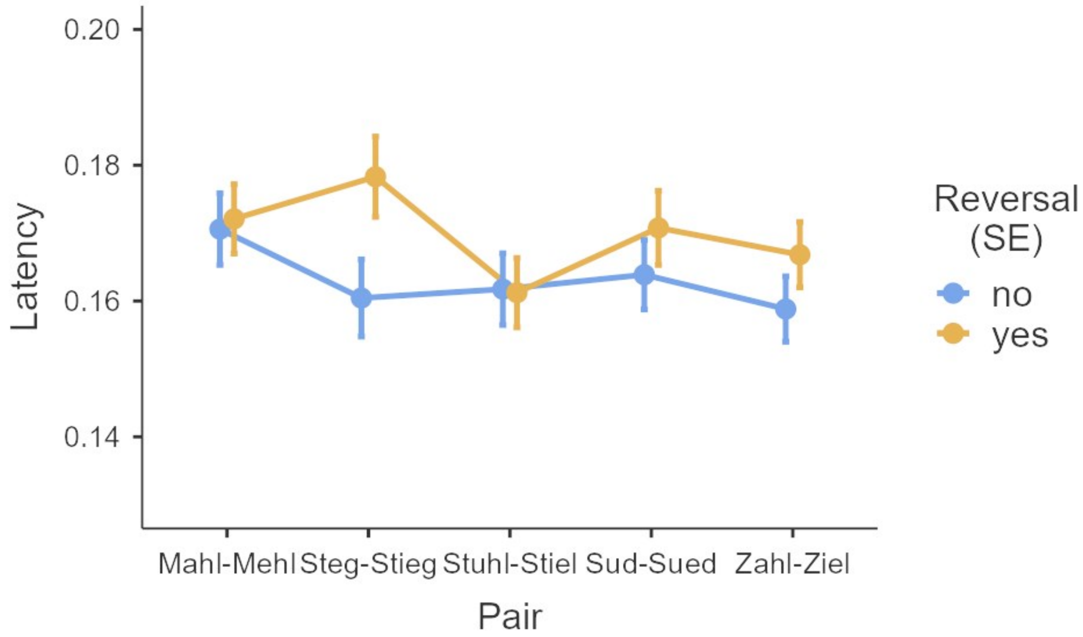

Figure 4: Effect plot for the linear mixed model of iMMN latency with factors pair and reversal.
The x-axis represents the investigated minimal pairs, while the y-axis measures the iMMN latency (in seconds). The presentation order is shown as different error bars in blue and yellow.

The figure shows that especially in the pair *Steg – Stieg* but also in pairs *Sud – Süd* and *Zahl – Ziel* difference between both presentation orders occur. This result is also reflected in the main effect of *reversal*. In order to investigate these visually seen differences further, we conducted a regression analysis of latency data with word frequency of occurrence since it has been shown that word frequency is a key influence on MMN latencies (e.g., Aleksandrov et al., 2017) with shorter latencies for higher frequent words. For this purpose, differences between words were established as the difference between the log-values (log lexical frequency of the deviant minus log lexical frequency standard). In the linear regression we found a marginal effect of lexical frequency (F = 3.75, p = .055) on the latency data with a tendency of shorter latencies for higher frequent deviants (see figure 5).

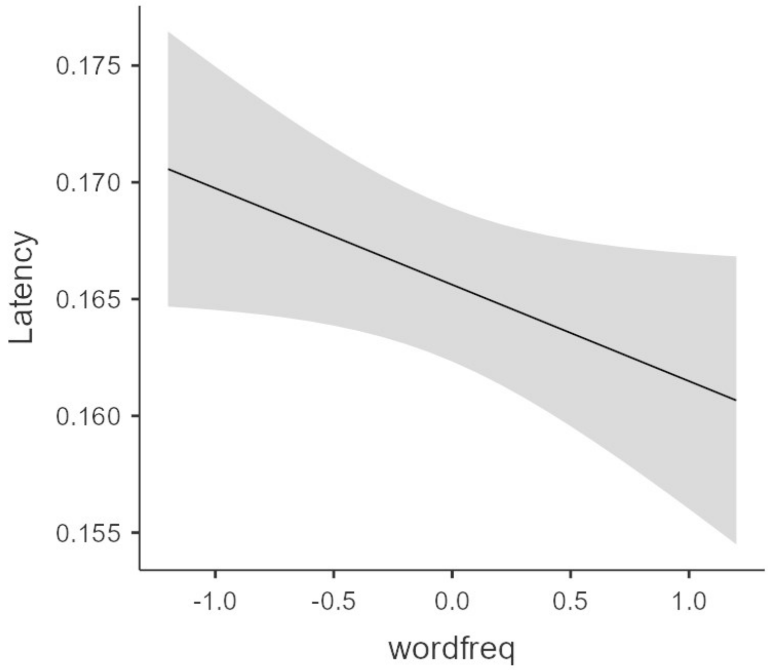

Figure 5: Effect plots for linear regression between latency and lexical frequency.
With increasing frequency (deviant more frequent than standard) latency of iMMN is decreasing.

1.1.7 Discussion

In summary, we did not find clear evidence in latency measures for any of the operationalised models since there was no systematic effect of feature mismatch and/ or peripherality. We could detect a tendency in latency data to decrease with increasing difference between standard and deviant regarding word frequency of occurrence. When the deviant has a higher lexical frequency than the standard iMMN latency decreases. This finding is in line with the research literature (e.g., Aleksandrov et al., 2017; Shtyrov et al., 2011).
To summarize the effects in the iMMN mean amplitudes, no clear evidence was found for neural asymmetries due to underspecification (FUL) but evidence for vowel discrimination based on phonetic salience of referent vowels (NRV). Furthermore, there were asymmetric as well as symmetric patterns in the MMN.
The asymmetric pattern of the comparison between /e:/ – /a:/ was along the hypothesis of NRV. Here, /a:/ is a referent vowel in addition to being more peripheral, and discrimination of /e:/ – /a:/ is therefore easier and comes with a stronger MMN effect than vice versa. Additionally, the symmetric effect in the contrast /y:/ – /u:/ can also be explained with this model since both vowels are referents within this framework. The same holds true for the comparison of MMN

effects between presentation orders of /i:/ – /u:/. But in the latter contrast, there could be also a phonological explanation within the underspecification approach. The phonological variation in morphological processes can lead to different specifications of segments within words and therefore to effects that are on first sight not compatible within the FUL paradigm (Lawyer & Corina, 2018). The same is true for German umlauting back vowels. In our case, when deriving plural of German word *Stuhl*, the stem vowel is umlauting and fronting (*Stühle*). It can be assumed that umlaut is only possible if the stem vowel /u:/ is not specified for place of articulation features and is therefore underspecified for backness (Scharinger, 2009; Scharinger et al., 2010). If the stem vowel of *Stuhl* is underspecified for place of articulation information, a symmetry must occur when it is compared to /i:/, which is also underspecified.

Contrary to this, the results of the remaining two contrasts are somehow challenging since none of the previous operationalized models can explain the effects given in the data. Comparing the presentation orders in the contrast /i:/ - /e:/, a symmetric MMN pattern occurred. This is challenging the predictions of FUL as well as NRV since both models predict an asymmetry. According to the underspecification approach, MMN effects for *Steg* should be stronger (underspecification of mid vowel height), while NRV predicted that neural effects for *Stieg* should be stronger (/i:/ should act as focal referent here). An explanation for the symmetric effect could lie in the close phonetic distance of the vowels involved. There has been evidence from previous MMN studies that effects diminished or failed due to only small acoustic deviances in speech stimuli (Pettigrew, Murdoch, Kei, et al., 2004; Pettigrew, Murdoch, Ponton, et al., 2004).

The most challenging results were obtained in the vowel contrast /i:/ - /a:/. Although there is an asymmetric effect, both models failed to predict the direction of the found asymmetry: FUL predicted stronger effects for *Ziel* as coronal deviant due to mismatching place of articulation (PoA) information. In comparison with *Zahl* as standard, which is classified as a dorsal vowel (Scharinger, 2009), the extracted feature [COR] from the acoustic signal of /i:/ should evoke a mismatching stronger MMN. Also, the mismatching height features of those two vowels cannot have evoked the asymmetry. Since both height features ([HIGH] and [LOW]) involved are specified in the underlying representation, a mismatch occurs regardless of the presentation order. Since a mismatch of those features occurs in both presentation orders, they should not evoke an asymmetric effect. Additionally, the NRV model cannot explain the found asymmetric pattern either. According to NRV, a symmetry should have occurred since both vowels act as focal referents within this framework. The explanation for these results is still unclear. We argue that since the more abstract feature representations are based on acoustic properties (mainly formants), the effects could be more driven by changes in the acoustics than in feature representations.

Because this is the contrast with the largest difference in terms of F1 or degree of openness, spectral characteristics (e.g., changes in F1) of the vowels could have been more involved in eliciting the surprising effects on an automatic and preattentive level. Additionally, changes in vowel quality do not only lead to changes of formants but also result in changes of other perceptual and psychoacoustic parameters. There is evidence that, for example, perceived loudness of speech stimuli varies for vowel quality. That is, lower front vowels are perceived louder despite equal intensity (Glave & Rietveld, 1975, 1979) and vocal effort (Eriksson & Traunmüller, 1999, 2002). Thus, we hypothesize that psychoacoustic and perceptual parameters such as perceived loudness could have played a crucial role. This possibility is explored in greater detail by means of multiple regression analyses in section 1.3.

1.2 Experiment 1b: Reaction Time

Since our MMN results present some evidence that effects were not only driven by phonemic factors but also by acoustic differences, we decided to conduct a reaction times study in an attended listening task. It has been shown that the MMN evoked by unattended processing is sensitive to a great variety of different dimensions between standard and deviant. Here, preattentive processing has been proven to be sensitive also for acoustic information like variations in duration and intensity (Jacobsen et al., 2003; Näätänen et al., 1989; Paavilainen et al., 1991; Schröger, 1996) or acoustic distance of stimuli (Deguchi et al., 2010; Savela et al., 2003). Since this component is highly sensitive to small low-level information differences (i.e. changes in frequency), higher-order information (for example phonemic identity) may be ignored or overridden in preattentive processing, for example, by acoustic proximity (Pettigrew, Murdoch, Kei, et al., 2004). Therefore, reaction times in an active discrimination task might reflect more cognitive, decision-based processing in which higher and more abstract effects like phonemic discrimination might surface better and more clearly. For this study, we thus propose the same hypotheses regarding potential asymmetries for both models as in experiment 1a.

1.2.1 Participants and Materials

Twenty-six participants (17 females, mean age 24.43, SD 4.23) were recruited, all of whom were graduate or undergraduate students at Johannes Gutenberg University Mainz. They received monetary compensation for their efforts. All participants were right-handed monolingual German speakers with no active dialect competence and were socialized with Standard German. No participant reported neurological, psychological, or hearing impairments. Written informed consent was obtained from each participant prior to the experiment.

The stimuli used in Experiment 1b were the same as in Experiment 1a. In contrast to the prior experiment, we had to reduce the amount of tested vowel contrasts to shorten the session length (approximately testing 45 minutes). Therefore, we tested only the vowel contrasts /i:/ – /a:/, /i:/ – /e:/, and /i:/ – /u:/. Contrasts were chosen as followed: /i:/ - /e:/ and /i:/ - /u:/ obtained in the MMN investigation symmetrical patterns and /i:/ - /a:/ evoked an asymmetrical pattern. The symmetrical pattern of /i:/ - /u:/ could be explainable with NRV and will therefore serve as control contrast for the remaining two vowel oppositions. Here, the iMMN results were not explainable by either of the models.

1.2.2 Task and Procedure

Stimuli were presented in an active oddball setup, in which participants had to press a button as soon as they perceived the deviant. They were told to perform a categorical, phonemic decision (and therefore ignoring the inter-token variability) (Johnson, 2015). During the experiment, subjects were seated comfortably in front of a screen in a sound shielded chamber. Sounds were presented with the Presentation software (http://www.neurobs.com, version 16.4) at a comfortable volume via two loudspeakers to the left and right of the screen. The volume was set prior to the experiment and was kept equal across all subjects. All written instructions were presented on screen. This way, participants were also informed about the beginning and end of the experiment as well as pauses. Each contrast direction contained 180 standards and 20 deviants divided into two blocks. In total, 12 blocks were presented. Stimuli within blocks were randomized with four to eleven standards between two deviants. Blocks of the same condition never followed each other.

1.2.3 Analysis and Results

The reaction time analysis was based on correct responses only (98% of data points included). Reaction times were corrected for the onset cluster of each stimulus. Thus, measurement of reaction times began on the vowel onset. Reaction times faster than 100 ms and slower than 1000 ms were excluded. The remaining data were log-transformed to obtain an approximately normal distribution (Ratcliff, 1993; Whelan, 2008). Outliers (+/- 2.5 SD) were removed before statistical analysis.

A repeated measures ANOVA with the factor word (e.g., *Ziel* vs. *Zahl*), controlled for multiple testing by applying Bonferroni correction, was calculated to reveal possible behavioural asymmetries. Here, we found a highly significant main effect ($F(5,2320) = 107.811$, $p < .001$). Post hoc analysis revealed asymmetric patterns in two of the tested vowel contrasts. /i:/ – /e:/ was significantly faster than vice versa ($F(1,464) = 22.234$, $p < .001$). The same was true for /i:/ – /u:/ ($F(1,464) = 13.550$, $p < .001$). However, in the vowel contrast /i:/ – /a:/ a symmetrical pattern of reaction times occurred. Here, the post hoc analysis shows no difference between presentation directions ($F(1,464) = 0.793$, $p =.374$) (see Fig. 6).

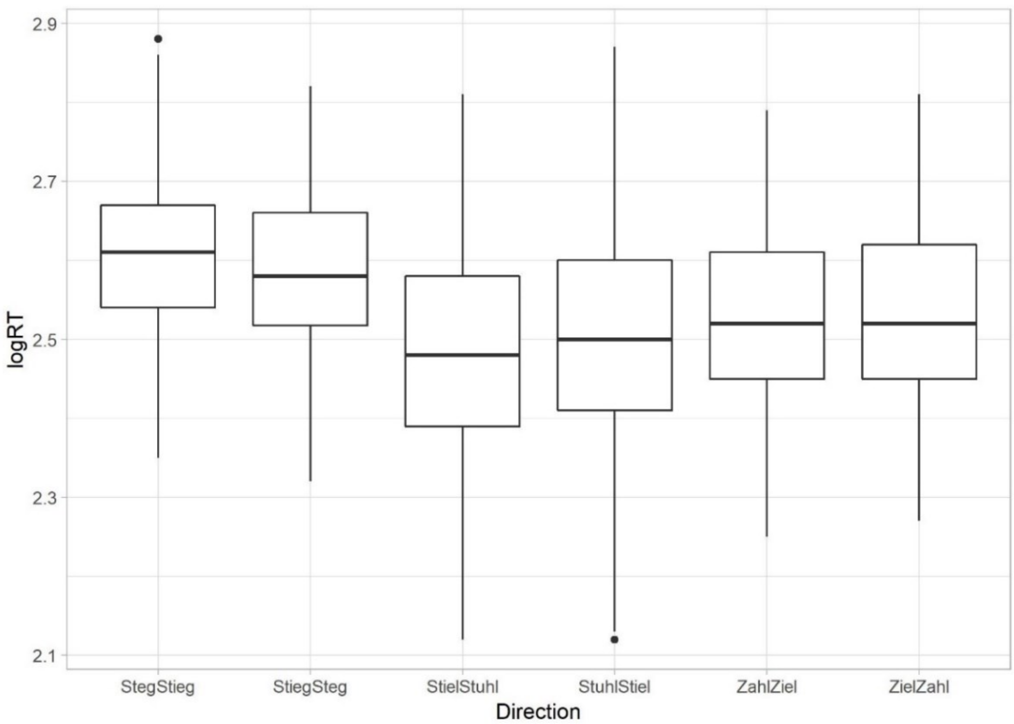

Figure 6: Reaction Time results per condition.
Reaction Time results are given as log values per presentation direction of words with whiskers indicating the variance of the data and small dots representing outliers (but not extreme values) which were beneath the +/- 2SD cut off.

1.2.4 Discussion

The behavioural study aimed to investigate the basis for some of the found electrophysiological effects in experiment 1a. There, both models could not provide a comprehensive explanation for our results, meaning that both models failed to explain all found effects. The vowel contrasts /i:/ – /a:/ and /i:/ – /e:/ were particularly challenging. Therefore, we conducted a reaction time experiment in an active oddball paradigm to investigate the previously found neural patterns in more detail. Overall, the reaction times indicate that in an active discrimination paradigm, German natural word stimuli were discriminated phonemically, based on higher-order abstract phonological features.

The reaction times for the vowel contrast /i:/ – /e:/ and the observed asymmetrical pattern match with the predictions of FUL. The faster reaction times obtained when /e:/ was deviant seem to be due to underspecification of [COR]. In this contrast, abstract representations may help to discriminate with respect to the close phonetic distance of the vowels. There is evidence from an fMRI study indicating that participants had to rely more on abstract feature representations while discriminating acoustically very close vowels (Scharinger, Domahs, et al., 2016).

In contrast, the directional asymmetry of the vowel contrasts /i:/ – /u:/ is, on first sight, more challenging for the underspecification approach since, following the theory, reaction times should be faster when subjects are presented with a fully specified vowel (e.g., /u:/) followed by an underspecified vowel (e.g., /i:/). But regarding the present study, we found the opposite effect for the obtained reaction times. The hypothesis of NRV for this contrast seems equally unsuitable: it states that a symmetric effect should occur because both vowels are reference vowels. In the case of /u:/ as deviant, one possible explanation could be that the additional labial feature drives the stronger effect. This additivity would then “overwrite” the feature mismatch. Several studies proved that the MMN is sensitive to an additivity effect correlated with the amount of deviating dimensions (Schröger, 1995; Takegata et al., 1999; Takegata, Paavilainen, et al., 2001; Takegata, Syssoeva, et al., 2001; Wolff & Schröger, 2001). Similar observations have been made in an fMRI study in which an increasing number of features led to stronger activation in the superior temporal sulcus (STS). In addition, the effect of stronger STS activations was also seen in reaction time measures whereby reaction time decreased with increasing feature number (Scharinger, Domahs, et al., 2016). Furthermore, in an MEG study it was shown that N1m amplitudes increased when feature number increased (Scharinger, Idsardi, et al., 2011a). More evidence for the additive effect has been brought to light in a MEG study with consonants, in which labial, specified glides produced stronger MMFs than coronal glides (Scharinger, Merickel, et al., 2011). Under the assumption of an additivity effect of the phonological feature [LAB], we argue that the underspecification approach still holds true since this model predicts effects based on sparse and abstract phonological features.

For the symmetrical effect in the contrast /i:/ – /a:/, there are two explanations we believe to be conceivable. The first one is that the hypothesis of NRV holds true. Since both vowels of this contrast are reference vowels in the framework, there is no discriminatory advantage in either direction. But why participants rely on phonetic features in these cases remains a question. The second more likely explanation argues within the underspecification approach: we classified Standard German /a:/ as a dorsal vowel. But there is articulatory evidence (see Fig. 1), evidence from theoretical analysis (Wiese, 2000), and also neurobiological evidence (Obleser et al., 2004) that Standard German /a:/ is likely not specified for place of articulation. Thus, there is no place feature mismatch anymore for /i:/ and /a:/. Since the remaining height features [LOW] and [HIGH] are both specified and mismatching regardless of the presentation order, a symmetry must occur.

In conclusion, it seems that participants use phonological and phonemic cues in vowel discrimination within natural German words. But the effects found in experiment 1a and 1b are different although the same experimental paradigm has been applied. The reason for different effect patterns in the electrophysiological and behavioural data could lie in the different attention requirements or differences of involved processing levels between the two tasks, but to this point, it is still not clear.

1.3 Identifying influential factors: perceived loudness

Because the interpretation of the MMN and RT data with common models is challenging, and because both models failed to explain the found patterns in a comprehensive way, we decided to test for additional influential factors in both datasets.

Vowel perception could be influenced not only by vowel identity but also by acoustic properties like intensity, duration, and fundamental frequency (Jacobsen et al., 2003; Näätänen et al., 1989; Paavilainen et al., 1991; Pakarinen et al., 2013; Peter et al., 2010; Schröger, 1996). Most of these factors researchers commonly try to exclude or control in the stimuli preparation procedure, but some acoustic factors cannot be avoided. For instance, since the phonological feature oppositions to distinguish different vowel qualities (i.e., high vowels vs. low vowels) are based on formants (Lahiri & Reetz, 2010a), they also automatically imply an acoustic difference. Moreover, when words are used as stimuli, lexical features like frequency of occurrence (Aleksandrov et al., 2011; Shtyrov et al., 2011) or phonotactic probability (Bonte et al., 2005; Emmendorfer et al., 2020, 2021; Yasin, 2007) are known to interfere in speech perception and vowel discrimination. In our approach, were we tested the hypotheses of the models by using natural German spoken words, those influences may contribute to patterns of results. Therefore, even though we here focused on the identity MMN, i.e., on electrophysiological responses to physically identical stimuli in different conditions, we decided to test whether and which of these factors have an influence on our electrophysiological and which affected the behavioural data. For this purpose, we operationalized different acoustic, phonological, and lexical factors. Furthermore, we also took acoustic and perceptual factors beyond the well-known ones (e.g., degree of openness and perceived loudness) into account to disentangle their contribution to the iMMN and RT data patterns. This in-depth analysis is explorative and has never been done this extensively before.

1.3.1 Preparation: Rating of Perceived Loudness

One possible additional influence beyond the well-known factors could be the perceived loudness of the stimuli. Here, loudness is referring to the magnitude of the auditory sensation (Fletcher & Munson, 1933) (not the physical intensity), but has been mainly taken as a perceptual correlate for sound intensity. Note that it has been shown that physical intensity of sounds and perceived loudness are measures on different auditory dimensions. While physical intensity is stimuli-inherent, the perceived loudness of stimuli is a perceptual phenomenon and therefore subject-dependent (Yanushevskaya et al., 2013). Moreover, while perceived loudness and sound intensity might be expected to be treated as equal, hearing research showed that two sounds of the same intensity can be rated with different perceived loudness levels due to various factors (e.g., spectral characteristics, bandwidth) (Moore, 2003, 2005). Additionally, perceived loudness levels could be correlated to gender differences since there is evidence that females perceive sounds louder than males despite the same sound pressure level (Hamamura & Iwamiya, 2016). Furthermore, there is evidence from sound processing that cortical activations are more likely driven by perceptual factors (e.g., perceived loudness) than physical characteristics (e.g., physical intensity) (Langers et al., 2007). Therefore, it could be possible in our study, although the stimuli words were normalized for the same average intensity and vowel intensity was approximately the same across words, that participants perceived the two words in a minimal pair as strongly different in terms of perceptual or sensational loudness.

1.3.1.1 Materials, Subjects, and Procedure

To test the word stimuli of experiments 1a and 1b for differences in the perceived (or implicit) loudness, we conducted a rating study. Here, 10 subjects (7 females, 3 males) participated, all of them students or employees at Johannes Gutenberg-University of Mainz who reported normal hearing. They were all monolingual German speakers (with a mean age of 32.6 years, SD 9.6) and gave written consent prior to the rating.

The word stimuli were arranged in the same minimal pairs as in the previous experiments. We tested all five minimal pairs in both presentation orders. Because we had five tokens per word in each presentation order, there were 25 possible combinations per presentation order, resulting in 250 trials overall. The trials were randomly arranged in ten blocks with 25 trials per block. Block order differed between subjects. The study was conducted via Presentation (http://www.neurobs.com, version 16.4), and auditory stimuli were delivered via headphones on the same listening level for all participants. For the experiment, all participants were seated in a quiet room.

At the beginning of the experiment, the instructions were presented on the computer screen. Afterwards, each trial started with a 1500 ms blank screen. Following a fixation star to keep participants engaged with the experiment, both words were then presented (ISI: 800 ms). After the presentation of the second word, a short blank screen (600 ms) was presented before two question marks with a timeout of 2500 ms appeared. The question marks were used as an indication for participants to give their answer via button press. Participants were instructed to rate the perceived loudness of the two words of each minimal pair in comparison to one another. Three answers were possible: first word louder, second word louder, or both words equally loud.

1.3.1.2 Analysis and Results

Having collected the responses of all participants, frequency values of the three answer categories (first, second, equal) for each minimal pair per presentation order were calculated. Timeouts were not included in the analysis. The distributions of answers for each direction can be seen in Figure 7. The Pearson chi square test, calculated in IBM SSPS (version 21) with variables direction (10) and answer (3), showed that the relationship between both variables is highly significant ($\chi^2_{(18)}$ = 998.986, $p < .001$).

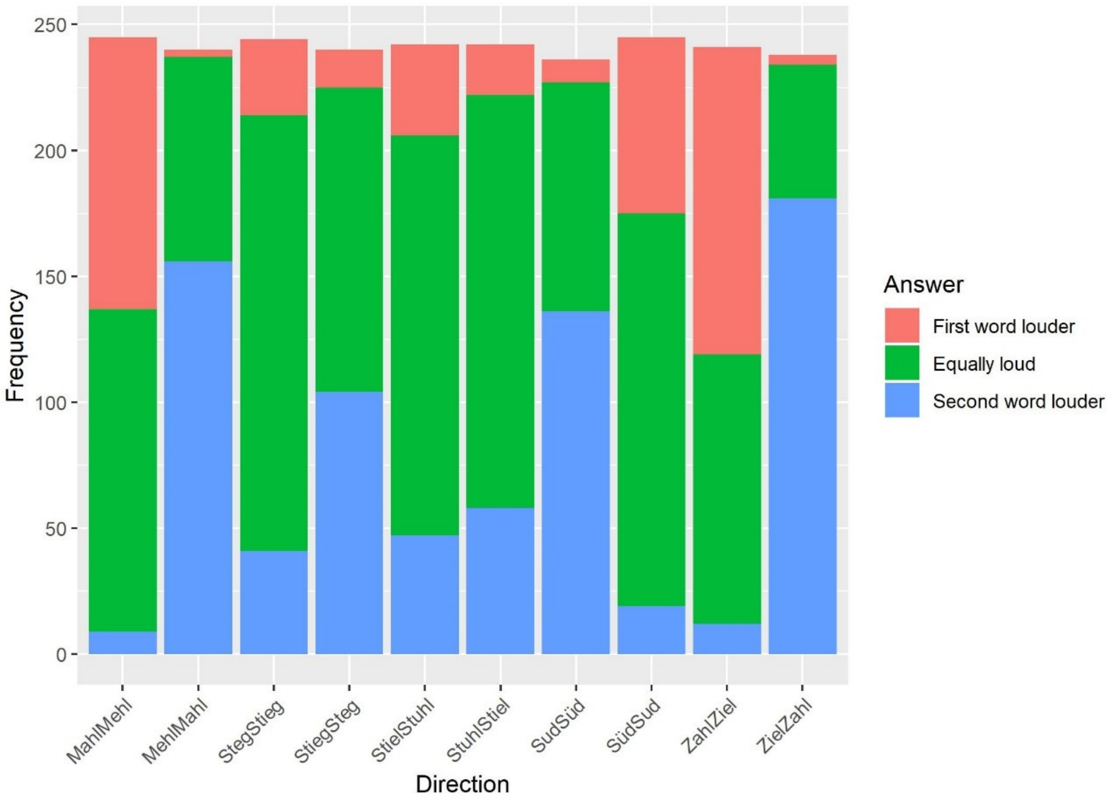

Figure 7: Results of the perceived loudness rating.
The results are plotted for each presentation direction (x-axis) in relation to the frequency of the given responses (y-axis).

In preparation for the operationalization of the factor of perceived loudness for the multiple regression analysis, frequency values, transformed in percentage with the highest given answer, will be considered in the next step of analysis. The percentages of answers per direction are given in Table 3.

Table 3: Distribution of the answers for the perceived loudness of words (in percent).

Direction/ condition	Equal	First	Second
Mahl – Mehl	**52.2**	44.1	3.7
Mehl – Mahl	33.8	1.3	**65.0**
Steg – Stieg	**70.9**	12.3	16.8
Stieg – Steg	**50.4**	6.3	43.3
Stiel – Stuhl	**65.7**	14.9	19.4
Stuhl – Stiel	**67.8**	8.3	24.0
Sud – Süd	38.6	3.8	**57.6**
Süd – Sud	**63.7**	28.6	7.8
Zahl – Ziel	44.4	**50.6**	5.0
Ziel – Zahl	22.3	1.7	**76.1**

The descriptive results are indicating a possible influence on the MMN data: In the contrast /e:/ – /a:/ (Mahl – Mehl, Mehl – Mahl), participants rated the word pair as equally loud by a higher percentage if *Mehl* was the second word (52.2%). In the reverse direction, the second presented word *Mahl* was more likely perceived as louder than *Mehl* (65%). Since, in this contrast, the MMN effect of Mahl (as deviant) was greater than in the reverse direction, perceived loudness could have a potential influence on preattentive processing of words.

In the contrast /i:/ – /a:/ (Zahl – Ziel, Ziel – Zahl), the word *Zahl* was perceived more often as louder regardless of the presentation order. In the first presentation order (Zahl – Ziel), *Zahl* was perceived in 50.6% as louder, and in the reverse direction, with *Zahl* as second word, it was also rated as louder (76.1%). The neural data showed clear MMN effects in both directions, with an asymmetric, because stronger, result for *Zahl* as deviant. It may be that the higher perceived loudness of *Zahl* has driven (Ziel – Zahl) or reduced (Zahl – Ziel) the neural effects.

In the next contrast with words *Sud* and *Süd*, similar patterns to the first one can be observed. When *Süd* was presented as the second word, participants perceived it as being louder (57.6%). When *Süd* was presented as first word, both words were rated more often as equally loud (63.7%). Taking the MMN results into account, it might again be possible that perceived loudness influenced the neural data. While the MMN effects are statistically symmetric, there is a slightly stronger effect for *Süd* as deviant (than in the reverse direction), when the plotted data are inspected.

In the last two contrasts (*Steg* & *Stieg*, *Stuhl* & *Stiel*), both words were described more often as equally loud within both presentation orders (Steg – Stieg: 70.9%, Stieg – Steg: 50.4%; Stiel – Stuhl: 65.7%, Stuhl – Stiel: 67.8%). Since the MMN data showed a symmetrical pattern in the statistical analysis, it could be stated that the perceived loudness might have had an influence on the neural effects once more.

Additionally, and in accordance with feedback of participants, it can be hypothesized perceived loudness could be correlated with the degree of openness of the long vowels. Especially with larger openness differences between vowels (/i:/ – /a:/, /e:/ – /a:/), the more open vowel /a:/ was rated as louder than the closer counterparts. Regarding the openness difference of /i:/ and /e:/, it can be stated that, phonetically, the difference here is smaller than between /i:/ – /a:/ and /e:/ – /a:/, and therefore the loudness effect could be perceptually reduced or inhibited. Moreover, the different MMN results, despite equal loudness rating patterns of /e:/ – /a:/ (asymmetric MMN) and /y:/ – /u:/ (symmetric MMN), could also support the hypothesis that the perceived loudness is correlated with degree of openness because of the latter contrast's lack of height difference. Contrary, the first mentioned contrast differs in vowel height and openness, and therefore the influence of perceived loudness could lead to stronger neural effects.

1.3.2 Explorative Analysis via Multiple Regressions

Because of the challenging and unexpected electrophysiological effects that do not match the behavioural results in addition to the descriptive identification of a potential influence of the perceived loudness, we decided to also investigate the possible influences of several additional factors on the neural (iMMN difference values of mean voltages between standard and deviant of the same stimulus) as well as the behavioural level (log RTs).

1.3.2.1 Defining Factors

Fifteen potential influential factors were defined, based on theoretical (*specificity*, *peripherality*, *focality*) and empirical input (*implicit loudness*, *electrodes*, *degree of openness*) as well as stimulus-inherent characteristics (*F1*, *F2*, *F3*, *frequency of words*, *bigram frequency*, *f0*, *vowel duration*, *intensity*). Additionally, one control factor (*contrast*) has been considered. All factors were operationalized and calculated with mean values per word category (i.e., mean F1 for *Mahl*) since the iMMN data as well as log RT data were also obtained as averaged data (for example in iMMN data, all tokens for *Mahl* as deviant or standard are collapsed in respectively one mean amplitude value).

The theoretical factors of *specificity* and *peripherality* have been operationalized in reference to both evaluated models in experiment 1a and 1b. While *specificity* (difference value between the number abstract features in the deviant minus the number of features in the standard) refers to FUL and takes the additivity effect discussed in the RT data into account, *peripherality* has been operationalized according the assumption of NRV that peripheral vowels maybe more salient and act as referents in vowel discrimination as a categorical variable (discrimination towards a more peripheral vowel, towards a more central vowel, or no referent/equal position in the vowel space). Additionally, *focality* has been operationalized according to the notion in NRV that the universal preference of referent vowels maybe alternated in adults due to language experience and is therefore taking the formant convergences in our stimulus set into account. *Focality* was again operationalized as categorical variable (discrimination from a less to a more focal vowel or from a more focal to a less focal vowel).

The empirically motivated factors have been chosen from input of the loudness rating. As mentioned before, the results of the rating study suggest an influence of *implicit loudness* on the neural data. Therefore, this factor has been included in further analysis as a function of the answer category with the highest percentage per presentation order (first word louder, second word louder, equally loud). Because there could be a possible relationship between loudness and the openness of vowels, *degree of openness* (increasing openness of the mouth, decreasing openness, equal openness) was also considered.

Since we tested natural (and mostly unmanipulated) German words in this study, there were stimuli-inherent differences between words that were controlled but could not be excluded in the preparation of the stimuli. The three first factors of this category are the differences between standard and deviant in terms of a change in *F1*, *F2*, and *F3*. To display the presentation orders of the stimuli in this factor, difference values (e.g., mean F1 of deviant minus mean F1 of standard) were calculated. Another possible stimulus-inherent influence is the *frequency* of occurrence of the words used. Since we wanted to test a large set of long vowels, we had to choose monosyllabic minimal pairs to reduce testing time. Being restricted by the German lexicon, we were not able to perfectly balance the lexical frequency of words; therefore, there are frequency differences between the words making up the minimal pairs. To test the influence of the *frequency* of occurrence on the electrophysiological and behavioural patterns, we included this factor in the explorative analysis in the form of difference values (log lexical frequency of the deviant minus log lexical frequency standard). The same was true for the factor *bigram* *frequency* (bigram frequency deviant minus bigram frequency standard). The same holds true for the factors *f0*, *vowel duration* and *intensity*. To evaluate the influence of the stimuli-inherent differences in fundamental frequency, vowel duration and intensity, we operationalized those factors also as difference values between deviant and standard (f0: difference between mean f0 of the deviant minus mean f0 of the standard; vowel duration: difference between vowel duration of the deviant minus the vowel duration of the standard; intensity: difference between mean vowel intensity of the deviant minus mean vowel intensity of the standard).

Finally, *contrast* has been included as control factor, since the iMMN and RT effects were different for the tested vowel oppositions. Here, both presentation orders of each minimal pair are combined.

1.3.2.2 Correlation and Single Linear Regressions

In preparation for the multiple regression analysis, we conducted first Kendall's Tau correlation for all previously defined factors. Additionally, we calculated for each factor a single regression model on the MMN data to identify reasonably factors to be included in the final multiple regression analysis.

Correlation analysis showed very strong correlation between the factors F2 and specificity (τ = -.907, $p < .001$) and F1 and degree of openness (τ = .866, $p < .001$), vowel duration and degree of openness (τ = .856, $p < .001$), as well as F2 and F3 (τ = .867, $p < .001$). Because of that, we only included specificity, F1, and F3 as theoretically implied factors. Additionally, vowel duration was included in the analysis since there is evidence that sound duration is influencing the perceived loudness (Todd & Michie, 2000). The exclusion of the other factors was necessary to avoid collinearities.

Additionally, strong, but in respect with collinearity uncritical correlations, were found between the following factors: F1 and f0 (τ = -.764, $p < .001$), F3 and specificity (τ = -.760, $p < .001$), F1 and vowel duration (τ = .764, $p < .001$), vowel duration and f0 (τ = -.778, $p < .001$), specificity and peripherality (τ = .559, $p < .001$), bigram frequency and peripherality (τ = .676, $p < .001$), vowel duration and intensity (τ = -.689, $p < .001$), f0 and intensity (τ = .556, $p < .001$), focality and F1 (τ = -.603, $p < .001$), F2 (τ = .507, $p < .001$) as well as F3 (τ = .686, $p < .001$), focality and f0 (τ = .686, $p < .001$), focality and vowel duration (τ = -.745, $p < .001$), with intensity (τ = .566, $p < .001$), with degree of openness (τ = -.731, $p < .001$) and implicit loudness (τ = -.654, $p < .001$), implicit loudness and f0 (τ = -.775, $p < .001$), implicit loudness and F1 (τ = .667, $p < .001$), implicit loudness and vowel duration (τ = .660, p < .001), implicit loudness and degree of openness (τ = .603, p < .001), and implicit loudness and intensity (τ = -.488, p < .001).

Single factor regression (with MMN data) revealed the influence of only five significant factors with reasonable R2 and adjusted R2 values: implicit loudness (ΔR2 = .023, p < .001), contrast (ΔR2 = .019, p < .001), intensity (ΔR2 = .016, p < .001), f0 (ΔR2 = .012, p = < .001), vowel duration (ΔR2 = .013, p < .001).

1.3.2.3 Hierarchical Multiple Regressions for MMN and log RT Data

The five previously identified factors were applied to hierarchical multiple regression with separate calculations for the MMN and the RT datasets in five steps with the following order: implicit loudness (model 1), contrast (model 2), f0 (model 3), vowel duration (model 4) and intensity (Model 5).

Results for the MMN dataset, for which the relevant key figures are displayed in Table 4, indicate that only the first two factors (implicit loudness and vowel contrast) are contributing to account for variance and that implicit loudness and contrast are approximately an equal fit for the explanation of results. In this study, implicit loudness seems to have an influence on the neural results of the MMN study (see Fig. 8).

Table 4: Multiple regression models for the MMN dataset.

Model#		b	SE B	β	p
1	Constant	-.363 (-0.592, -0.133)	.117		.002
	Loudness	-.436 (-0.607, -0.265)	.087	-.155	.000
R^2 = .024, ΔR^2 = .023, p < .001					
2	Constant	-.703 (-0.971, -0.434)	.137		.000
	Loudness	-.436 (-0.605, -0.267)	.086	-.155	.002
	Contrast	.170 (0.098, 0.242)	.037	.143	.000
R^2 = .044, ΔR^2 = .043, p < .001					

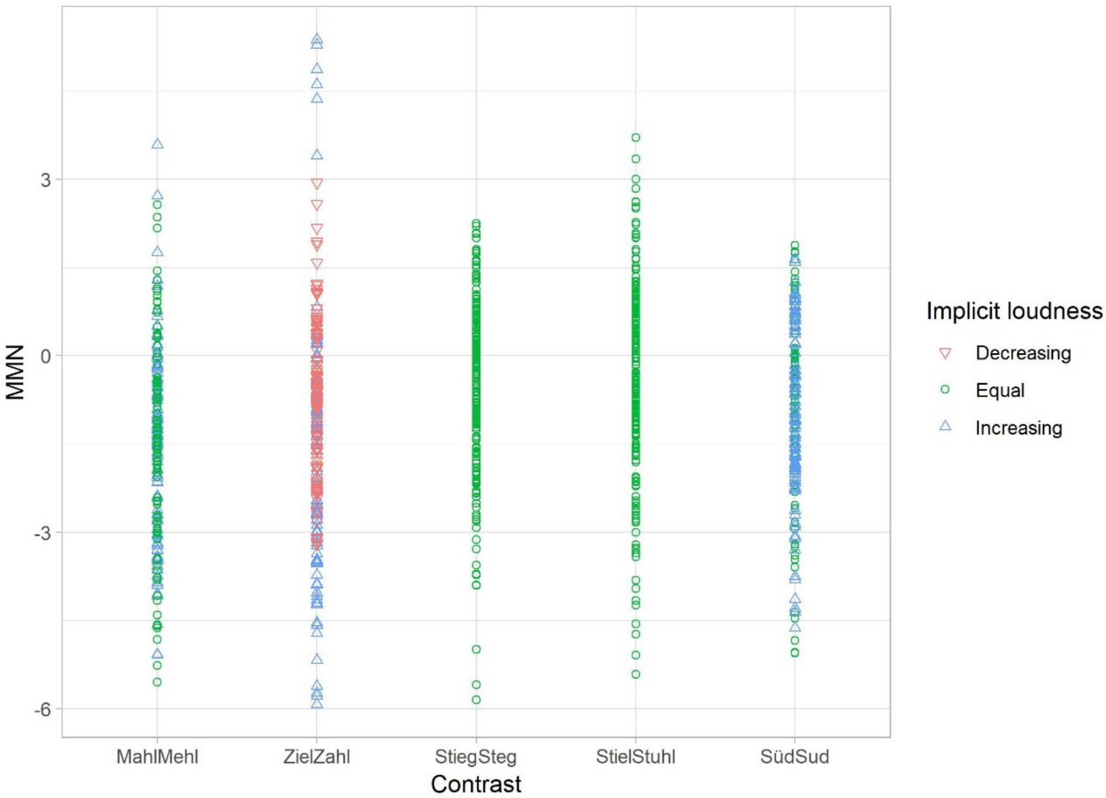

Figure 8: Scatterplot for the regression analysis of the given iMMN data from experiment 1.
MMN difference values of each participant (y-axis) per vowel contrast (x-axis) in relation to implicit loudness. Increasing loudness (deviant louder than standard) is shown as a blue triangle, decreasing loudness (standard louder than deviant) as a red triangle and equal perceived loudness as a green dot. MMN difference values are scaled with perceiving the stimuli as equally loud (most clearly seen in vowel contrasts StiegSteg and StielStuhl).

In contrast, results of the second multiple regression model in the log RT dataset implicate that the factors influencing the neural results are not contributing to the explanation of behavioural patterns. Once again, Model 2 (implicit loudness and contrast) is fitting best with contrast as the only significant contributing regression coefficient (Table 5). Thus, implicit loudness is not contributing to the found behavioural pattern in the reaction time experiment (see Fig. 9).

Table 5: Multiple regression models for the RT dataset.

Model#		b	SE B	β	p
1	Constant	2.547 (2.522, 2.573)	.013		.000
	Loudness	.001 (-0.021, .024)	.011	.010	n. s.
R^2 = .000, ΔR^2 = -.006, n. s.					
2	Constant	2.558 (2.535, 2.595)	.015		.000
	Loudness	.001 (-0.020, 0.023)	.011	.127	n. s.
	Contrast	-.021 (-0.033´6, -0.005)	.008	-.204	.011
R^2 = .041, ΔR^2 = .029, $p < .05$					

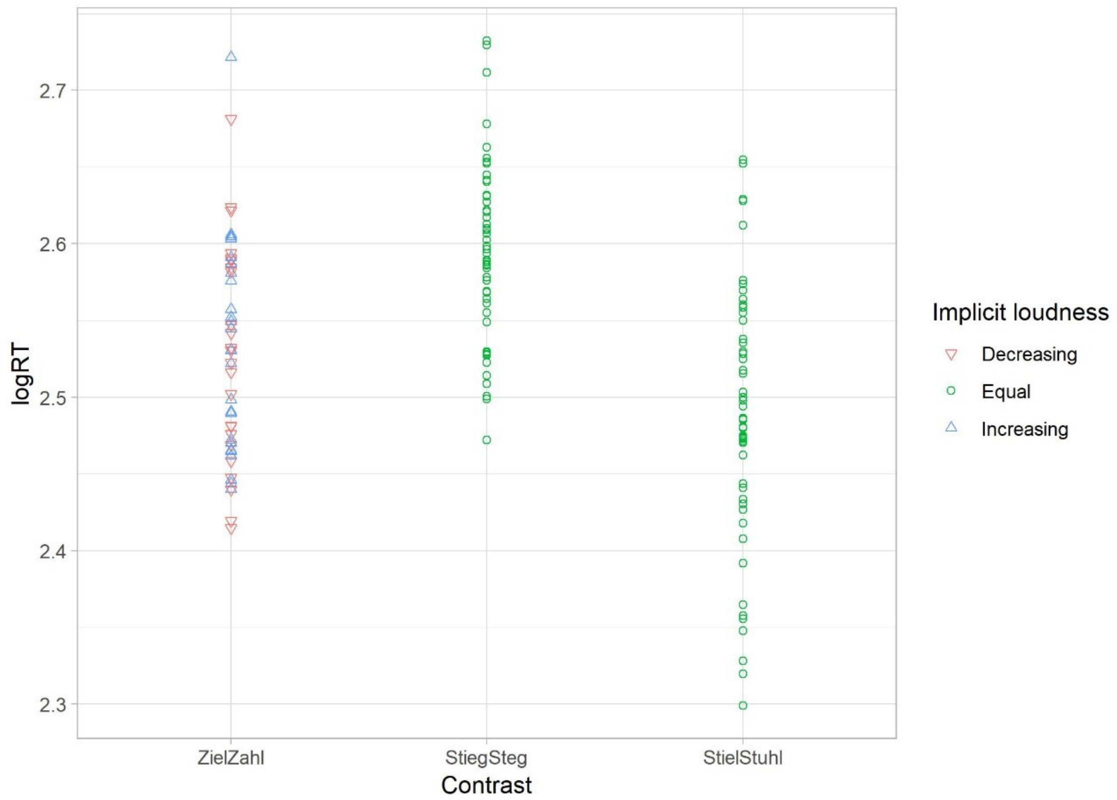

Figure 9: Scatterplot for the regression of the obtained RT data from experiment 2.
Mean log RTs of each subject (y-axis) are depicted per vowel contrast (x-axis) in relation to implicit loudness. RT results (log-values) are not scaled by the perceived loudness of the stimuli.

1.3.3 Discussion

Multiple regression analysis on both datasets revealed that the perceptual factor of perceived loudness had only an influence on neural effects. Here, it can be stated that regarding the MMN data, the phonological and phonetic status of the vowels presented in the minimal pairs played a role in the elicitation of MMN effects (factor contrast), but–crucially–effects were simultaneously driven by the implicit loudness of the presented words. Overall, it seems that the neural effect was scaled with perceiving the stimuli as equally loud. Put differently: the more strongly one word was perceived as being louder, the further the MMN difference values deviated from zero. Therefore, it can be argued that perceived, or implicit, loudness seems to be an important factor in interpreting the found neural patterns.

This is especially true for those contrasts for which both models (NRV and FUL) failed to explain the effects. Especially in the symmetrical contrasts of Stieg – Steg and Stiel – Stuhl, implicit loudness seems to drive the symmetry as in this contrast, both words were more often perceived as equally loud regardless the presentation order. The small visible (but statistically not significant effect) in the contrast Süd – Sud could also be explained by this factor since there were more increasing judgments if *Süd* was the second word. Missing statistical significance could be a result of the correlation of loudness with degree of openness. Since this contrast did not differ in terms of vowel height (and therefore openness), the influence of perceived loudness could be weaker than in vowel oppositions with height differences. Turning to the asymmetrical patterns in the MMN data, it can be stated that for the pattern of Ziel – Zahl, which was especially challenging because both models could not explain the found asymmetry, implicit loudness once more seems to drive the neural effects since *Zahl* was more often perceived as louder regardless of the presentation order. If *Zahl* served as deviant, the greater perceived loudness has led to a stronger effect than in the reverse direction. Here, the greater implicit loudness of the standard could have reduced MMN effects. In the last contrast (i.e., Mehl – Mahl), differences in perceived loudness and degree of openness could have led to the stronger effect for *Mahl* (as deviant). *Mahl* as the second presented word was perceived as louder than in the reverse direction. In this direction, the degree of openness is also increasing. In the reverse direction, equally perceived loudness might have elicited smaller effects.

Turning to the RT data, multiple regression analysis revealed that the perceived loudness had no influence on the behavioural patterns; therefore, the effects are more likely driven by the traditional models (namely FUL and NRV) discussed before.

1.4 General Discussion

In this paper, we reported the results of an electrophysiological and a behavioural study as well as an explorative analysis of influential factors on vowel discrimination. While most studies investigating speech sound discrimination only tested two or three vowel oppositions, we conducted our MMN study on a much larger stimulus set. Here, we investigated preattentive vowel processing with five different vowel contrasts covering the most important German long vowels embedded in real and natural German words. To obtain an even more natural listening situation, we used five tokens per word, which resulted in a large stimulus set. The purpose of the investigations reported here were two-fold: First, we wanted to compare two often discussed models for vowel discrimination to investigate which model can explain the found effects in German in the best way. Second, we wanted to shed further light on factors influencing vowel discrimination on the neural and behavioural level. For this purpose, we conducted an in-depth analysis delving into possible confounds to a degree which has not been investigated so far.

To summarize the results of the electrophysiological experiment with respect to the first research question, we found MMN evidence for discrimination and perceptual asymmetries (or symmetries) in vowel perception according to the NRV model in the amplitude data. Three contrasts showed facilitated and asymmetric discrimination on presenting a less peripheral vowel as standard and a more peripheral vowel as deviant. These results are in line with other behavioural and electrophysiological studies (Y. Liu et al., 2021; Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019) that also report easier discrimination from a more central to a more peripheral vowel. Only one contrast could be explained within the underspecification approach. By contrast, both models failed to explain the found symmetric neural patterns. For those vowel oppositions, it could be possible that the phonemic discrimination was overridden in preattentive processing through acoustic proximity (Pettigrew, Murdoch, Kei, et al., 2004) or sensational interferences caused by perceived loudness. The lack of phonemic discrimination and weighting of sensational influences in the MMN experiment could be due to experimental protocol since subject were not instructed to perform phonemic discrimination, but to ignore the stimulation (Johnson, 2015). Additionally, we analysed the MMN latency with respect to FUL and NRV to investigate, whether asymmetric effect due to underspecification or higher salience of the periphery of the vowel space emerge in this measure. Analysis revealed no systematic effect of either of both models. Instead, we could find a marginal effect of word frequency which is in line with other MMN studies (Aleksandrov et al., 2017; Shtyrov et al., 2011).

To test those challenging results in the amplitude data, we investigated these critical, symmetric contrasts, namely /i:/ - /e:/ and /i:/ - /u:/, with a behavioural active oddball paradigm and instructed participants here to perform a phonemic decision. Thus, we assumed that in the active oddball paradigm subjects had to activate more abstract mental representation more strongly due to allophonic variance in the stimuli; therefore, blending out simple acoustic differences in the decision making. The behavioural results, in contrast to the MMN effects, showed that participants were able of phonemic discrimination based on abstract representations. Here, the found patterns can only be fully explained by the underspecification approach, in line with previous studies delivering evidence for speech sound discrimination with the help of sparse and abstract features (e.g., Eulitz & Lahiri, 2004; Lahiri & Reetz, 2010a; Scharinger, Bendixen, et al., 2012; Scharinger, Monahan, et al., 2012).

In summary, the results of experiment 1a and 1b are challenging in two ways. First, the neural pattern cannot be explained in a comprehensive way by either of the two models. Secondly, the neural and behavioural patterns do not match. The lack of compliance between electrophysiological and behavioural results can be interpreted in terms of an attention shift and cue weighting as a function of task dependency. Differences in cue weighting due to attention shifts have been reported in several studies. Szymanski et al. (1999) conducted an MMN study with and without attention on the stimulation and found differences in the neural responses. They interpreted these findings in terms of a modulation of the memory trace in the attended condition. Here, attention leads to the activation of more accurate and precise representations of the standards, which in turn generate larger responses of the deviant. Therefore, it can be argued that in attended stimulation, more information is accessible for discrimination than in unattended conditions. Similar results with respect to the richness of mental representation accessible during discrimination as function of attention was also found by Tuomainen and colleagues (2013). In an MMN study with an active go/no-go task they found that in the attentive task participants were able to use more spectral attributes in vowel discrimination than when they listened passively to the stimulation. The authors interpreted the results as a change in perceptual discrimination strategy due to the attention shift. Furthermore, Savela et al. (2003) found, in a study combining MMN (passive oddball) and RT (active oddball), that subjects discriminated the used Finnish and Komi vowels differently depending on the task. While the behavioural results indicated a phonemic discrimination of the vowels, the preattentive MMN patterns were more driven by acoustic differences than phonemic representations. With respect to our electrophysiological and behavioural results, it can be assumed that the attention shift between the passive and the active oddball has led to differences in cue weighting in vowel discrimination. We argue that in the active experiment, participants were able to discriminate phonemically, which led to patterns explainable by common models. In contrast to this–but in accordance with previous studies–the passive MMN patterns are based not only on phonemic but also on acoustic or perceptual differences.

To address this issue further, we conducted an explorative analysis of influencing factors of the electrophysiological as well the behavioural datasets. We included several theoretical, lexical, and phonotactical factors that are known to influence results. While other studies found an influence on neural data, for example, of phonotactic probabilities (Bonte et al., 2005; Emmendorfer et al., 2020, 2021; Yasin, 2007) or the lexical frequency of words (Aleksandrov et al., 2011, 2017; Shtyrov et al., 2011), we cannot provide evidence for those factors neither on the electrophysiological nor on the behavioural data. On the contrary, we have identified a new influencing factor on MMN data: we found that neural effects were not only driven by phonemic features but also by the perceptual and psychoacoustic differences in perceived loudness in the stimuli. In contrast, no such influence of this factor could be found in the behavioural data. Therefore, the multiple regression analyses on both datasets support the aforementioned interpretation on different discrimination strategies since we found that the influence of perceived loudness of the word stimuli only mattered in the neural but not the behavioural data. Once again, these results can be interpreted as evidence that in preattentive processing, more perceptual and acoustic features are responsible for elicitation of effects. But when attention was shifted towards the stimulation (like in the active oddball paradigm of the RT experiment), these perceptual factors receded into the background, and discrimination was based on phonemic representations of the perceived vowels only.

Although perceived loudness is related to (and heavily determined by) sound intensity, two sounds of equal perceived loudness may well have different levels of sound intensity (Yanushevskaya et al., 2013). This is due to the processing of auditory stimuli in the cochlea (Moore, 2003, 2005), which depends not only on the characteristics of the stimuli, such as bandwidth, but on the listener as well. We found evidence that speech signals of approximately equal intensity could still be perceived to be of different loudness. Additionally, we could show that within our datasets', perceived loudness was highly positively correlated with degree of openness of the vowels and changes in F1. We conclude that perceived loudness differences could be guided by differences in degree of openness since with increasing openness of the tested vowels, the perceived loudness of words increased as well and since increasing loudness elicited larger MMN effects. These results add evidence to the hypothesis that perceived loudness of vowel stimuli is also linked to vowel quality (Glave & Rietveld, 1975, 1979). Additionally, we found correlations of perceived loudness and changes in f0, intensity and vowel duration. But the multiple regression analysis showed that those additional factors did not contribute to the found neural asymmetries. Here, only the differences in perceived loudness can explain the found patterns. However, since there is evidence that perceived loudness can be influenced by vowel duration (Todd & Michie, 2000) and changes in fundamental frequency (Hsu et al., 2015), more studies are needed to disentangle all the factors contributing to differences in the sensational perceived loudness of stimuli and influencing natural vowel processing.

To our knowledge, we are the first to find evidence for the influence of perceived loudness on the perception of German long vowels and on MMN data regarding natural vowel processing. We propose that the perceptual, or implicit, loudness of stimuli can act as an intermediate representation level between stimuli-inherent acoustics and abstract phonological features. The exact influence of perceptual and psychoacoustic factors, like perceived loudness in speech processing, is still under investigated and more research is needed. But for the time being, our results provide evidence that studies should include more factors beyond the well-known (and theoretically driven) when analysing and interpreting neural and behavioural data.

2. Neural and behavioural asymmetries with German long vowels

In the first series of experiments, we found patterns of cue weighting (attention and task dependent) as well as novel evidence for the influence of perceived loudness on neural effects (see also Riedinger et al., 2021). We conducted the first series of experiments with tokens of natural German minimal pairs since we wanted to test the often-implemented models of vowel processing (FUL and NRV) in a more natural listening situation. As described above, with the use of words and naturally spoken tokens of words, there are possible confounding factors. Albeit these confounding factors showed no significant influences in the multiple regression analyses, differences in MMN effects and perceived loudness ratings could be additionally influenced by (extra-)linguistic factors and not only due to the vowel quality. Therefore, we decided to conduct a second series of experiments, following the rationale of the experiment series described above, on stimuli with reduced linguistic complexity using isolated long vowels to test if effects and processes previously found are truly related to vowel processing.

Therefore, we tested both models (FUL and NRV) in isolated long vowels as well as the influence of perceived loudness on vowel processing. Again, we used isolated long vowels instead of words to isolate the effects of vowel processing from morpho-lexical influences. Hence, the aims and the research questions of the second series of experiments were threefold:

(1) which model can explain directional asymmetries the best (**model comparison**),

(2) can the before found **influence of perceived loudness** in unattended vowel processing be seen also in a setting where linguistic context is reduced. Therefore, it should be possible to differentiate influences of vowel processing from other linguistic effects, and

(3) does the utilizing of features differentiate with attention and acoustic distance/ feature dimension (**cue weighting**).

Thus, to test the research questions, we conducted an iMMN and a reaction time study with isolated, but naturally produced German long vowels /e:/, /o:/, and /u:/. The investigated vowel oppositions varied in feature dimension (opposition of PoA or Tongue Height) and acoustic distance (large or small). An additionally conducted rating of perceived loudness is used to investigate the influence of this psychoacoustic factor on vowel processing via multiple regressions.

2.1 Experiment 2a: identity MMN

2.1.1 Participants

For this EEG experiment we tested 30 participants between 20 and 33 years (mean age: 24.1, 14 male/ 16 female). All participants were right-handed monolingual German speakers with no active dialect competence. They were socialised with standard German. Additionally, none of them reported any neurological, psychological, or hearing impairment. All participants were undergraduate or graduate students at the university of Mainz. Every participator was informed about the used method and gave written consent for their participation prior to the experiment.

2.1.2 Materials

Stimuli were natural, spoken isolated German long vowels /e:/, /o:/ and /u:/. Stimuli were recorded from a female, phonetically trained standard German speaker. In the recording session, 20 exemplars for each vowel quality were spoken with neutral pronunciation. Afterwards all recordings were controlled for f0, F1, F2, and F3. All recordings were cut to an equal length of 400 ms with a 50 ms fade-out and the mean intensity was normalised to 85 dB in PRAAT (Boersma & Weenink, 2016). We then chose the five best recordings for each vowel quality as experimental stimuli. Phonetic measurements can be seen in table 6.

Table 6: Phonetic parameters of the vowels.
Mean f0, F1, F2, and F3 values are given in Hertz. Mean Intensity are given in dB.

Vowel	Mean f0 (SD)	Mean F1 (SD)	Mean F2 (SD)	Mean F3 (SD)	Mean Intensity (SD)	Mean Duration (SD)
/e:/	223.9 (0.7)	434.6 (4)	2535.2 (65.7)	3034.8 (53.3)	85.0 (0.7)	400.3 (0.3)
/o:/	223 (1)	434.6 (3.3)	708.2 (28.6)	3147.4 (113.1)	86.8 (0.3)	400.1 (0.3)
/u:/	223.7 (1.3)	326.4 (23.9)	657.4 (25.4)	2862 (73.5)	86.2 (1.3)	400.1 (0.5)

We compared formant frequencies with those of Sendlmeier & Seebode (2006) to ensure that the recorded long vowels were perceived as German standard phonemes. Additionally, three independent listeners assessed the recordings distinct for each vowel quality category. In figure 10 vowels of each quality category used as experimental stimuli are displayed.

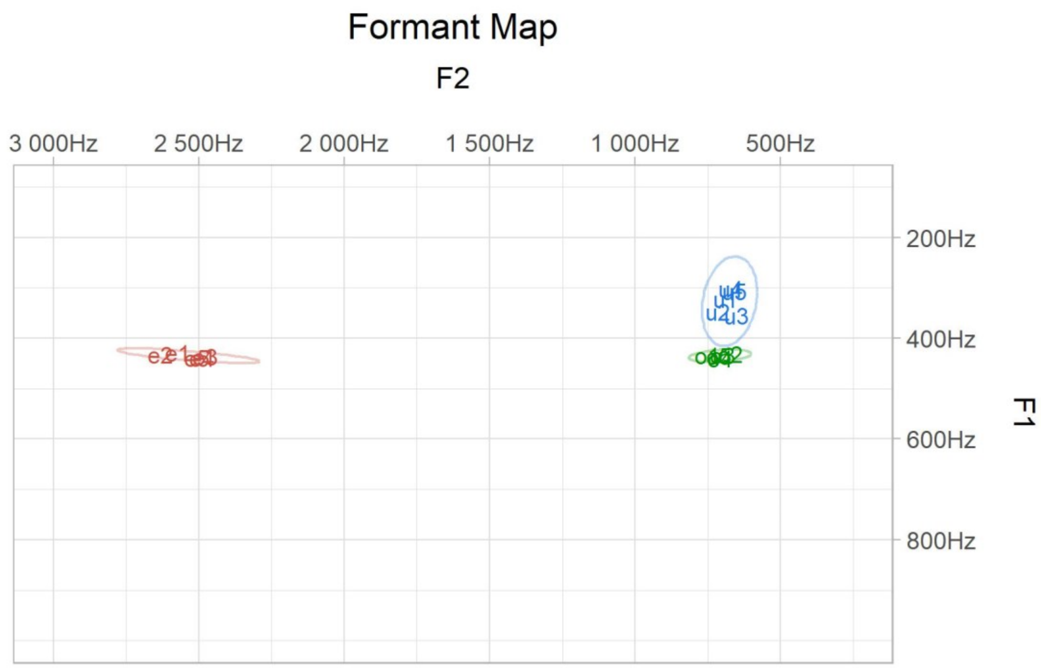

Figure 10: Acoustic characteristic of the isolated vowel stimuli.
Mean values of the first (F1) and the second (F2) formant are given for each vowel in Hertz.

2.1.3 Task and Procedure

To elicit the MMN in the EEG recordings, we used a passive oddball design as already described above. While the stimuli were presented, participants watched a silent movie.
The three used vowel categories were arranged in two vowel contrasts (/o:/ - /e:/, /o:/ - /u:/). The contrasts differ in terms of PoA and labiality (/o:/ - /e:/) or vowel height (/o:/ - /u:/). Additionally, there are differences in terms of acoustic distance (operationalised as Euclidean distance in bark based on F1 and F2): /o:/ - /e:/ (8.01 bark) has a larger acoustic distance than /o:/ - /u:/ (1.23 bark). We tested these vowel contrasts here, since they were not included in the before described series of studies and thus can shed light on whether the found influences on vowel processing can be generalized for (German) long vowels. Both vowel oppositions were tested bidirectionally resulting in four vowel directions. For each presentation order, 800 standards and 120 deviants were presented. The stimuli were arranged in two blocks per presentation

order, resulting over all in eight blocks. Stimuli were pseudorandomised and it was ensured that no more than 2 identical tokens were adjacent of one another. The interval between two deviants was kept randomly between four and eleven standards. Additionally, we implemented a random ISI between 600 – 1100 ms to guarantee a jittered presentation. Blocks were arranged in five different lists with pseudorandomised block order, where blocks of the same presentation direction never followed each other.

During the experiment, participants were seated in a sound-insulated booth in front of a screen. Stimuli were presented with Presentation (version 16.4, neurobs.com) via two loudspeakers left and right of the screen at a comfortable listening level, which was set prior to the experiment and was kept equal throughout the study.

2.1.4 Hypotheses

Since we wanted to compare the models FUL and NRV in this study, we formulated hypotheses regarding MMN amplitudes for each presentation direction based on the assumptions for both models. The made assumptions and derived hypotheses are also displayed in table 7. We operationalised both assumptions of NRV – universal salience of the vowel space periphery and stimuli-dependent salience of focalisation. Since it remains unclear in the literature how focalisation should be computed and implemented (Polka & Bohn, 2011a; Simon et al., 2014), we defined focality as the distance between F1 and F2 in bark (calculated as difference of F2 minus F1), with lower values pointing to stronger formants convergence. In our stimulus set, we found the following focality order: /o:/ is the most focal vowel in the stimulus set (2.24 bark), followed by /u:/ (3.04) and /e:/ (10.22).

Table 7: Hypotheses.
Assumptions for feature specifications (FUL), location in the vowel space (NRV - Peripherality), and formant focalization (NRV – Focality) for each vowel contrast and hypotheses for MMN effects according to both models.

Vowel direction	**Features (FUL)**	**Mapping result (FUL)**	**Expectations MMN (FUL)**	**Classification (NRV - Peripherality)**	**Expectations MMN (NRV - Peripherality)**	**Classification (NRV - Focality)**	**Expectations MMN (NRV – Focality)**
/o:/ - /e:/	[DOR] – [COR]	Mismatch	***Stronger effect***	Both peripheral	***Symmetrical effect***	More – less focal	Weaker effect
/e:/ - /o:/	[–] – [DOR]	No-mismatch	Weaker effect			Less – more focal	***Stronger effect***
/o:/ - /u:/	[–] – [HIGH]	No-mismatch	Weaker effect	Central – peripheral	***Stronger effect***	More – less focal	Weaker effect
/u:/ - /o:/	[HIGH] – [MID]	Mismatch	***Stronger effect***	Peripheral – central	Weaker effect	Less – more focal	***Stronger effect***

As can be seen in table 7, the different operationalised models lead to different hypotheses regarding potential asymmetries.

According to FUL, presentation orders /o:/ - /e:/ and /u:/ - /o:/ should obtain stronger MMN effects. In contrast, the NRV peripherality hypotheses are different. Here, the first vowel contrast (/o:/ - /e:/ vs. /e:/ - /o:/) should result in a symmetrical effect since both vowels are peripheral in vowel space. In the second contrast, /o:/ - /u:/ should lead to a stronger MMN. In difference to the first NRV assumptions, the focality hypotheses of NRV predict asymmetrical MMN effects in both vowel contrasts. Due to the focalisation of F1 and F2, vowel directions /e:/ - /o:/ and /u:/ - /o:/ should obtain higher MMN amplitudes.

Again, we expect that stronger MMN amplitudes are accompanied by shorter latencies (de Rue et al., 2021; Eulitz & Lahiri, 2004; Verá-Constán & Sebastián-Gallés, 2008, as cited in Polka & Bohn, 2011).

2.1.5 EEG Recording and Analysis

EEG was recorded in a sound-shielded booth with BrainVision Recorder (version 1.10, Brain Products GmbH) while participants watched a silent movie. For the measurement of the EEG signal we used in total 34 Ag/AgCl passive electrodes connected to a BrainAmp Amplifier (Brain Products GmbH). Electrodes were mounted on an elastic cap using the 10-20 system. Ground electrode was AFz. An electrode on the nose tip was used as online reference. Four electrodes were used to measure the electrooculogram (EOG) to identify artifacts of eye movement in the analysis. Therefore, two electrodes were placed left and right of the eye canthi to record lateral eye movements, while electrodes above and beyond the right eye measured blinks. For all electrodes, impedance was kept below 5 kΩ throughout the recording session.

For the analysis of the neural data, we used Matlab (version 2021b) with the fieldtrip toolbox (Oostenveld et al., 2011). Data of 26 (out of 30) were analysed. Four participants had to be excluded due to excessive sweating and movements as well as errors during data acquisition (loss of trials). Raw data were filtered with 0.1 Hz high- and 30 Hz lowpass filters. The neural data were re-referenced offline to linked mastoids. After data segmentation, muscular artefacts were removed automatically, which lead to a trial exclusion between 2.3 and 18.9 %. Afterwards, eye movements were automatically corrected through correlations between the EOG channels and ICA components.

For averaging, the first ten standards and the standard that was presented immediately after a deviant, were excluded. Additionally, ERP data were baseline-corrected using the 100 ms prestimulus epoch. MMN was calculated as identity MMN (iMMN), comparing standard and deviant of the same stimulus, to reduce the influence of acoustic differences between standard and deviant and to tap into underlying processing mechanisms (Eulitz & Lahiri, 2004; Pulvermüller et al., 2006).

2.1.6 Results

ERP results are plotted in figure 11.

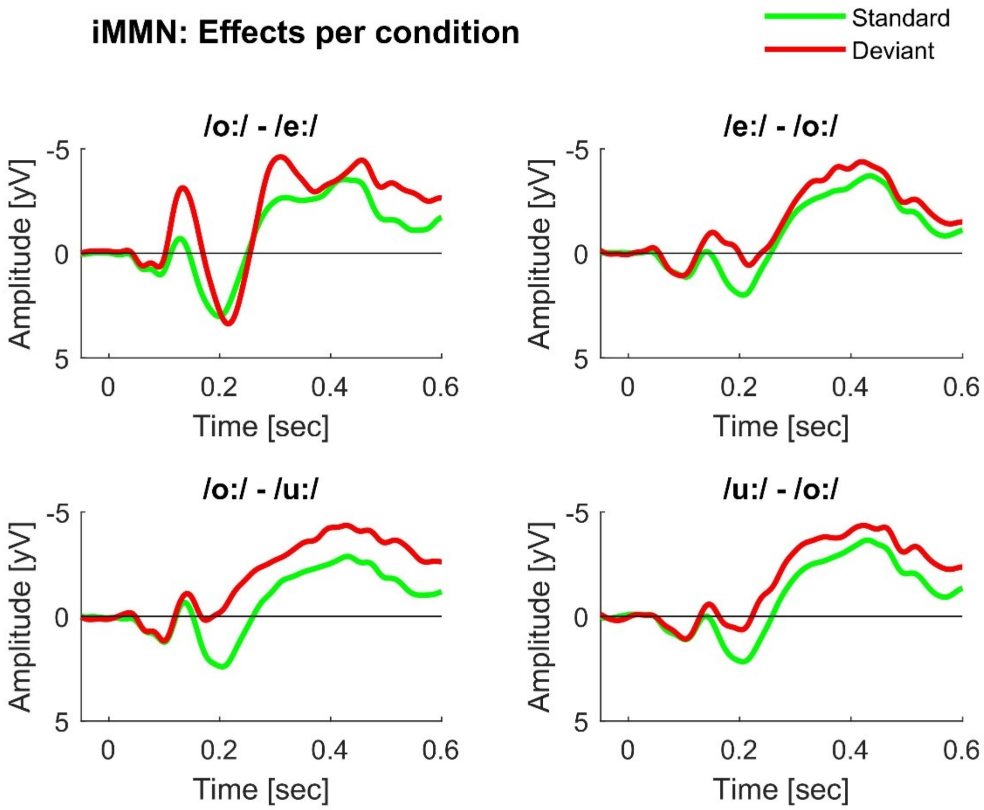

Figure 11: iMMN results per condition.
For each direction order standard (green line) and deviant (red line) of the same speech sound are plotted. For example, for the presentation order /o:/ - /e:/ deviant /e:/ is plotted against standard /e:/.

To identify the iMMN time window and significant electrode, we used the cluster-based permutation test implemented in fieldtrip (Oostenveld et al., 2011)[2]. As can be seen in figure 12, in the electrode (mainly midline electrodes) x time cluster (130 – 180 ms) deviants elicited a more negative response than standards.

[2] For a more detailed description of the procedure see chapter 1.1.6.

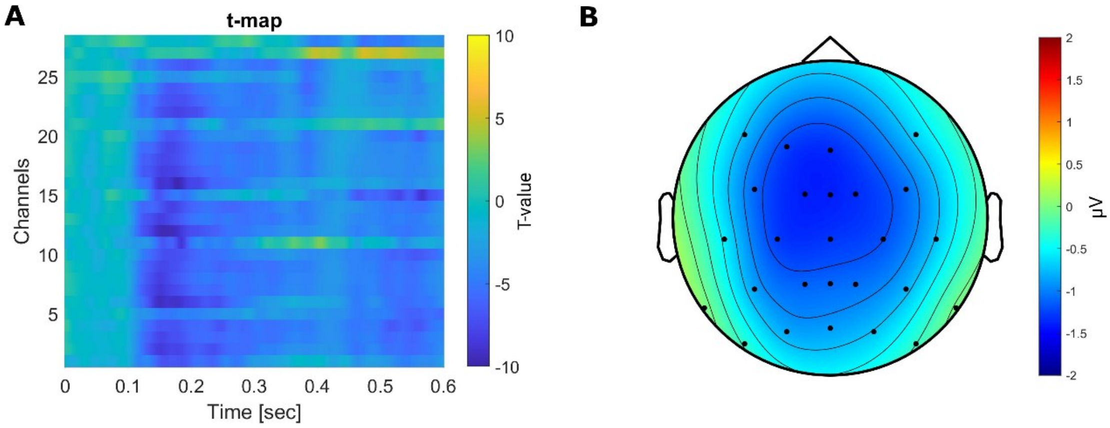

Figure 12: Clusterstatistics and effect topography.
(A) Clusterstatistics as t-map, (B) heat map in MMN window 130 – 180 ms with typical topographical distribution.

We analysed in the identified time window the amplitudes and peak latency of the iMMN for asymmetric or symmetric patterns. Inferential statistics are calculated with jamovi (The jamovi project, 2021).

For the amplitude analysis, we calculated a linear mixed model, using the gamlj package (Gallucci, 2019), with varying intercept with fixed factors *direction* and *electrode*, and *subjects* as a random factor. The dependent variable was the mean iMMN amplitude in the aforementioned time window. The following formula describes our model: diff ~ 1 + direction + electrode + direction : electrode+ (1|subject). We used a repeated contrast coding for *direction*, where each group is compared against the subsequent group. For the factor *electrode* a simple contrast coding (with Fz as basis of comparison) was implemented. On a descriptive level (see table 8), direction /o:/ - /e:/ had the most negative response and amplitudes were most negative at fronto-central electrode sites.

Table 8: Descriptives for amplitude measure for factors direction and electrode.
Mean amplitudes in the time window 130 – 180 ms are given in microvolt.

Direction	Mean amplitude (sd)	Electrode	Mean amplitude (sd)
oe	-2.18 (1.48)	Fz	-1.34 (1.56)
eo	-0.956 (1.32)	FCz	-1.45 (1.56)
ou	-0.603 (1.23)	Cz	-1.40 (1.46)
uo	-0.581 (1.29)	CPz	-1.10 (1.34)
		Pz	-0.869 (1.24)

16.8 percent of the data can be explained by the chosen fixed factors (R^2 marginal), and 40 percent when the random factor is additionally considered (R^2 conditional). Hence, our model was a good fit. The linear mixed model showed significant main effects for direction ($F = 42.96$, $p < .001$) and electrode ($F = 3.54$, $p < .01$). No interaction between fixed factors was found. In table 9, fixed effects parameter estimates are shown. Here, only direction /o:/ - /e:/ and electrode comparison between Fz and Pz were significant.

Table 9: Fixed Effects Parameter Estimates for iMMN amplitudes.

Names	Effect	Estimate	SE	df	t	p
Intercept	Intercept	-1.206 (-1.494, -0.917)	.147	25	-8.19	< .001
Direction 1	**oe – eo**	**-1.226 (-1.5, -0.952)**	**.140**	**487**	**-8.78**	**< .001**
Direction 2	ou - uo	-.090 (-0.364, 0.184)	.140	487	-0.64	n. s.
Electrode 1	Fz – FCz	-.055 (-0.361, -0.252)	.156	487	-0.35	n. s.
Electrode 2	Fz – Cz	.038 (-0.268, 0.344)	.156	487	0.24	n. s.
Electrode 3	Fz - CPz	0.237 (-0.069, 0.543)	.156	487	1.52	n. s.
Electrode 4	**Fz – Pz**	**0.455 (0.149, 0.761)**	**.156**	**487**	**2.91**	**< .01**

Post-hoc analysis with holm correction for multiple comparisons for the factor electrode showed that Fz ($t = -2.89$, $p < .05$) and FCz ($t = -3.23$, $p < .05$) had the strongest MMN effects, while for factor direction only the comparison between direction order /o:/ - /e:/ and /e:/ - /o:/ was significant (t = 7.18, p < .001). Hence, we found in the comparison of iMMN amplitudes between /o:/ - /e:/ and /e:/ - /o:/ an asymmetric effect, while the other important comparison between /o:/ - /u:/ and /u:/ - /o:/ showed a symmetric iMMN effect. The asymmetric effect can be clearly seen in figure 13.

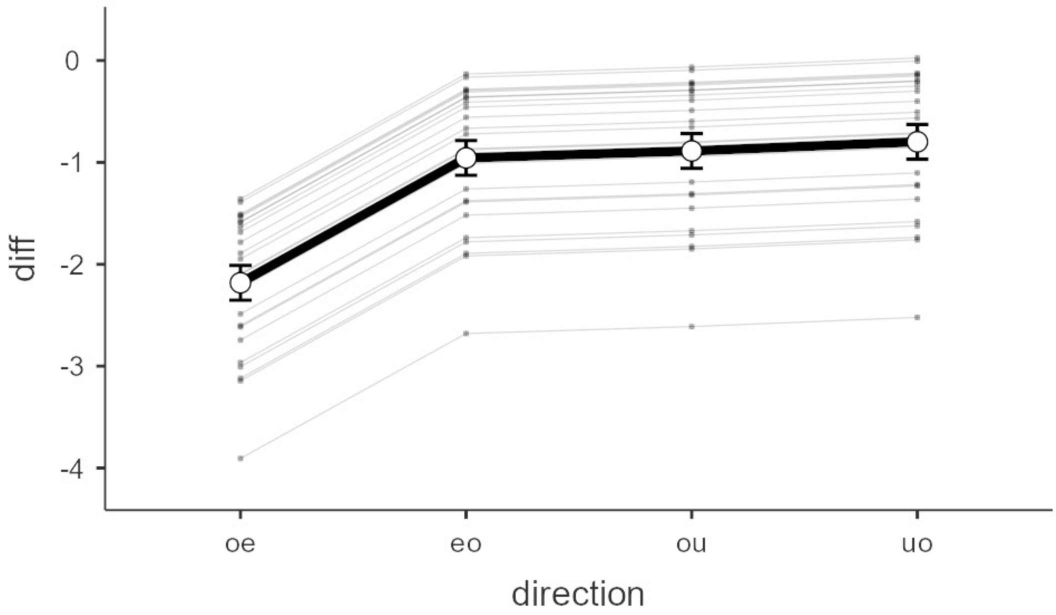

Figure 13: Effect plot of the linear mixed model for iMMN amplitudes.
Directions /o:/ - /e:/ and /e:/ - /o:/ show a clear asymmetric effect, while the comparison between the other directions reveals a symmetric pattern.

Additionally, we analysed peak latencies of iMMN within the time window in the four presentation orders to see whether asymmetric and symmetric effects can also be found in latency measures. We also calculated a linear mixed model with varying intercept using *contrast* and *reversal* as fixed factors and again *subjects* a random factor. Mean values of the iMMN latency are shown in table 10. The structure of the fixed factors can also be seen here.

Table 10: Descriptives for iMMN latency at Fz.

Contrast	Reversal	Mean latency in sec (sd)
OE	no	0.149 (.024)
	yes	0.165 (.024)
OU	no	0.180 (.022)
	yes	0.151 (.030)

In factor *contrast*, the two related presentation orders are collapsed into one contrast category, e.g., OE containing presentation orders /o:/ - /e:/ and /e:/ - /o:/. The factor *reversal* describes which presentation order within one contrast is meant, e.g., reversal "yes" in contrast OE points to the presentation order /e:/ - /o:/. Hence, asymmetric patterns should produce interaction effects. In this analysis, iMMN peak latency in the significant time window was the dependent variable. The model formula was the following: Latency ~ 1 + contrast + reversal + contrast : reversal + (1|subject). In this model, we used simple contrast coding for both fixed factors. The chosen model was a considerably good fit with a R^2 marginal and conditional of 16.7 percent. Overall, only a significant interaction contrast x reversal ($F = 13.75$, $p < .001$) was found. Table 11 shows the fixed effects parameter estimates.

Table 11: Fixed Effects Parameter Estimates for iMMN latencies.

Names	Effect	Estimate	SE	df	t	p
Intercept	Intercept	.161 (0.155, 0.167)	.003	72	53.54	< .001
Contrast	OU – OE	.009 (-0.003, -0.020)	.006	72	1.42	n. s.
Reversal	yes – no	-.006 (-0.018, 0.006)	.006	72	-0.97	n. s.
Contrast * Reversal	**OU – OE * yes – no**	**-.045 (-0.068, -0.021)**	**.012**	**72**	**-3.709**	**< .001**

Post-hoc analysis with holm correction for multiple comparisons in both contrasts were calculated. Latencies of iMMN in the comparison of presentation directions /o:/ - /e:/ and /e:/ - /o:/ show a marginal asymmetric pattern ($t = 2.23$, $p = .07$) with shorter latency for /o:/ - /e:/ as well as a clear asymmetric effect in the comparison between /o:/ - /u:/ and /u:/ - /o:/ ($t = 3.90$, p <

.01) with shorter latency for /u:/ - /o:/. The asymmetric patterns are clearly visible in the effect plots of our model (see fig. 14).

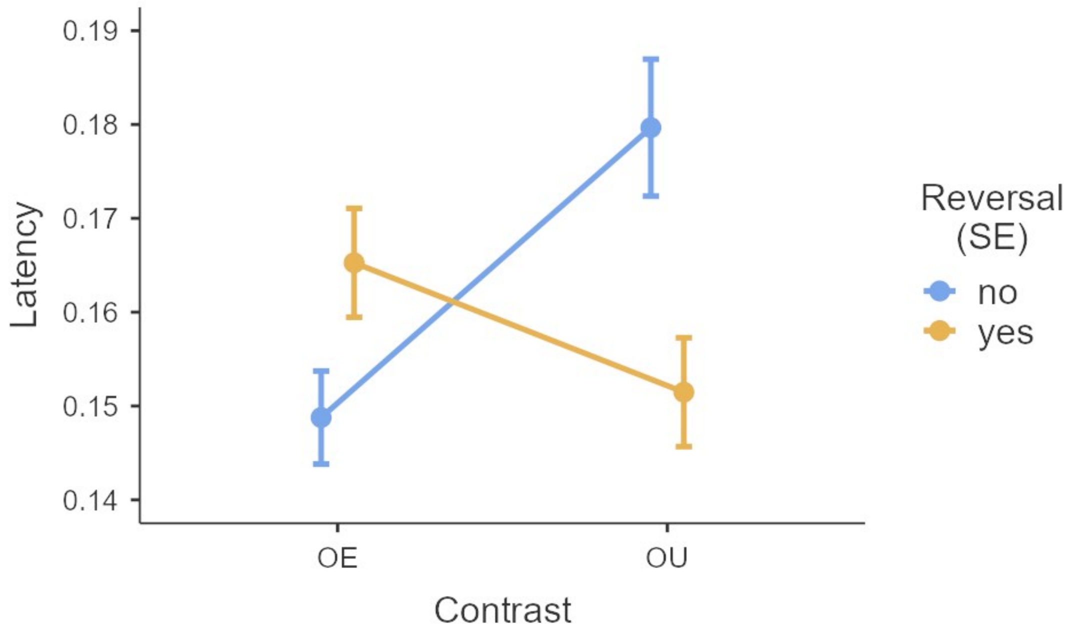

Figure 14: Effect plot for the iMMN latency analysis.
In both contrasts the two related presentation orders show asymmetric latency effects.

2.1.7. Discussion

As in experiment 1a of the previous chapter, we again found asymmetric as well as symmetric amplitude effects. In total, it can be stated that MMN responses to changes of place of articulation (and labiality), /o:/ - /e:/, were greater than to changes of vowel height. This finding is in line with other studies (de Jonge & Boersma, 2015; Nudga et al., 2021).
Regarding asymmetric and symmetric amplitude effects, we first found a stronger MMN effect in /o:/ - /e:/ than in the reversed direction. This result is in line with the hypotheses of FUL. The model predicted a stronger MMN amplitude in /o:/ - /e:/ since the feature [COR] in the surface structure of the deviant clashed with the, in the mental representation activated, feature [DOR] of the standard. Stronger effects due to underspecification are reported also in a series of other studies (Cummings et al., 2017; de Rue et al., 2021; Eulitz & Lahiri, 2004; Lahiri & Reetz, 2002a, 2010a; Nudga et al., 2021; Riedinger et al., 2021; Scharinger, Monahan, et al., 2012). In the second vowel contrast, we found a symmetric effect between presentation orders. But both models failed to predict this symmetric amplitude effect. Thus, both models cannot explain the found

effect. According to Riedinger et al. (2021), who also found symmetric amplitude effects (see also chapter 1.1 and 1.3), it is possible that the equalness in the perceived loudness of the involved vowels led to equal amplitudes also. This possible explanation will be examined in chapter 2.3.

To summarise, the found effects in the amplitude data could be evidence for cue weighting due to phonetic or acoustic distance (Scharinger, Domahs, et al., 2016). If the vowels have a larger acoustic distance, like /o:/ - /e:/, the discrimination process could be easier and therefore discrimination is possible with abstract and sparse phonological features (Nudga et al., 2021). Contrary, if acoustic distance between vowels is small (/o:/ - /u:/), the discrimination between vowels is more difficult (Flege et al., 1994) and therefore the abstract features are not sufficient. Hence, it could be hypothesised that in this case the participants resorted to other features, such as the psychoacoustic factor of perceived loudness, to resolve the harder discrimination situation. This possible influence will be examined later.

Regarding the found asymmetric latency effect, with shorter latencies for presentation orders /o:/ - /e:/ and /u:/ - /o:/, it must be stated that both effects were predicted by FUL but not by either of the NRV approaches. Therefore, in the latency data, the found effects can be fully explained by underspecification. To repeat, usually the stronger MMN amplitudes are accompanied by shorter MMN latencies. Hence, in mismatching presentation orders /o:/ - /e:/ ([DOR] – [COR]) and /u:/ - /o:/ ([HIGH] – [MID]) we found shorter latencies due to mismatching features between the activated underlying representation of the standard and the surface form of the deviant. Another possible explanation could lie in the cue weighting assumptions made for the amplitude effects. If we assume cue weighting due to acoustic distance of vowels in vowel discrimination, the following argumentation could be true: In the amplitude data we found for large acoustic distance, /o:/ - /e:/, discrimination effects according to FUL, hence we argued that in the easier discrimination situation sparse and abstract features are sufficient. The same applies to the found latency effect in this contrast. Here, the latency asymmetry, i.e. /o:/ - /e:/ shows faster latency than /e:/ - /o:/, is in line with the underspecification approach. Accordingly, we would expect the usage of other features (as in the amplitude data) in the case of the smaller acoustic distance of the vowel contrast /o:/ - /u:/. Here, presentation order /u:/ - /o:/ is explainable by the NRV focality hypothesis. Thus, it can be argued that in more difficult discrimination situations, MMN latency reflects vowel discrimination with phonetic (more acoustically motivated) features.

Why the effects of amplitude and latency do not match in total, remains unclear and needs further investigation. But it is possible that MMN amplitude and latency are influenced by different factors or processes of vowel discrimination.
Before we investigated the possible influence of perceived loudness on the neural data and tested for other influential factors of MMN amplitudes and latencies, we conducted a reaction time study with the same materials to investigate whether the found neural patterns are also present in behavioural data and if the found cue weighting pattern can be replicated.

2.2 Experiment 2b: Reaction Time

Analogous to the first series of experiments, we also conducted a reaction time study in a passive oddball paradigm after the MMN study. Once more, we wanted to test whether tendencies of cue weighting as seen in experiment 2a also occur behaviourally and therefore in a more decision-based dimension. Regarding possible asymmetric results between presentation orders, we assumed the same hypotheses as in experiment 2a. According to FUL, RTs should be faster for /o:/ - /e:/ and /u:/ - /o:/. With respect to NRV's peripherality assumption, a symmetric effect should occur between /o:/ - /e:/ and /e:/ - /o:/ while in the other vowel opposition /o:/ - /u:/ should lead to faster reaction times. Additionally, according to NRV's focality hypotheses, /e:/ - /o:/ and /u:/ - /o:/ should show faster reaction times.
Also, the same participants of the before described MMN study participated in the reaction time experiment. Likewise, we used the same materials and investigated the same vowel contrasts and presentation orders.

2.2.1 Task and Procedure

To investigate behavioural asymmetries, we implemented again an active oddball task where we presented both vowel contrasts bidirectionally. Since the same persons as in experiment 2a participated in the behavioural task, the reaction time study was conducted after the EEG recording but within the same session. After the EEG recording, participants had time to wash their hair and refresh. After this pause (usually between 15 and 20 minutes), participators were seated again in a sound-insulated booth in front of a screen. They were instructed to intentionally listen to a stream of speech sounds, where one sound would occur very often and the other one only few times. Participants had to press a button as soon as they perceived the deviant sound.

In total we presented 180 standards and 20 deviants per presentation direction divided into two blocks. Therefore, stimuli were presented in 8 blocks. Stimuli were pseudorandomised with four to eleven standards between the presentation of deviants. No more than two tokens of the same speech sound followed each other. Block orders were randomised across five lists. Blocks of the same presentation direction were never presented directly after one another. Participants had small pauses between blocks and could decide on their own when to begin with the new block. Stimuli were presented with a fixed ISI of 400 ms to ensure phonemic processing (Johnson, 2015). For stimulus presentation, the software Presentation (version 16.4, neurobs.com) was used. Stimuli were presented with a comfortable, fixed loudness via loudspeakers left and right of the monitor. Loudness of presentation was kept the same throughout the whole study. Reaction times and accuracy data were logged also in Presentation.

2.2.2 Analysis and Results

We included data of all 30 participants in the analysis. Before we analysed reaction times, we investigated the accuracy on subject and vowel contrast level, to ensure participants' active engagement in the study. On subject level, accuracies were between 91.5% and 100 %. On contrast level, the contrast /o:/ - /e:/ (both presentation orders together) reached slightly better accuracy (98.5 %) than in the contrast /o:/ - /u:/ (95.6 %). But on both subject and contrast level, accuracy shows a ceiling effect.

For the analysis of reaction time, data was first pre-processed in RStudio (RStudio Team, 2022). Here, we log-transformed the data to obtain a nearly normal distribution (Ratcliff, 1993; Whelan, 2008). Subsequently, outliers were removed by using the function 'identify_outliers'[3] of the package rstatix (version 0.3.0). Here, 50 data points (2.2 %) out of 2295 data points in total were excluded. The log-transformed and corrected results per direction are depicted in figure 15.

[3] Within this function, values above Q3 + 1.5 x IQR and below Q1 – 1.5 x IQR are excluded from further analysis.

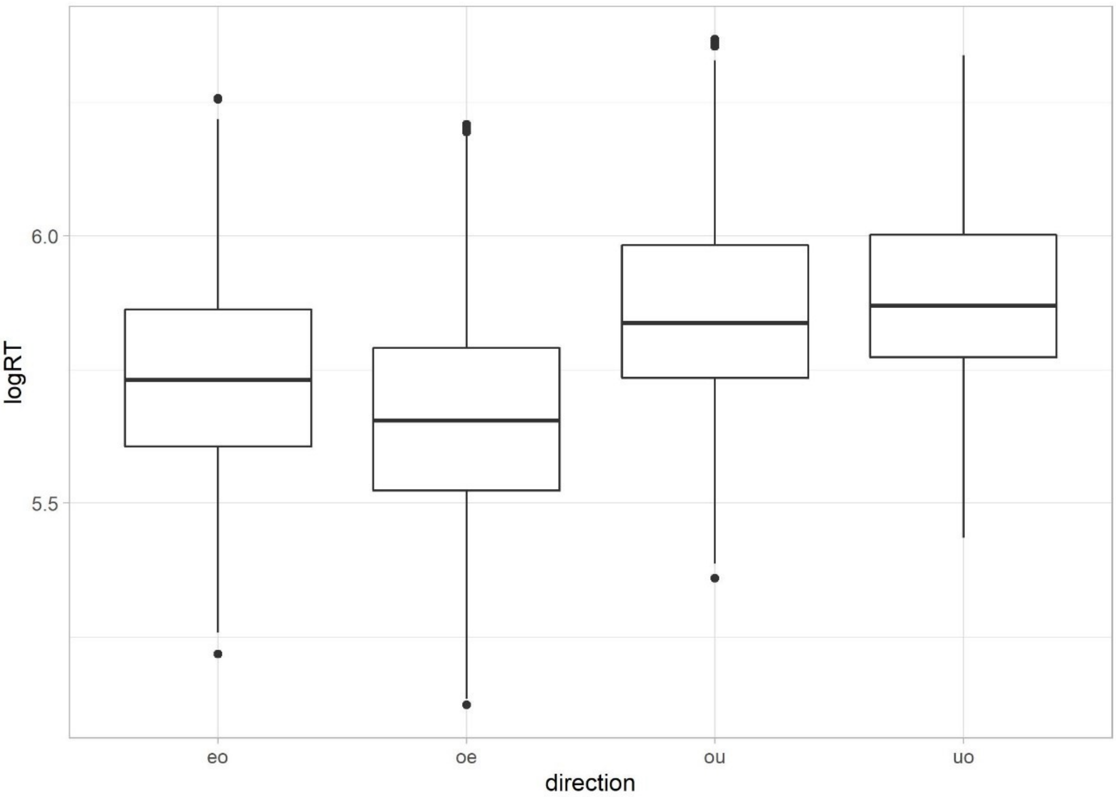

Figure 15: Box plots per direction of the log-transformed reaction times for each presentation order.

After pre-processing the data, statistical analysis was conducted in jamovi (The jamovi project, 2021). To investigate asymmetrical effect patterns, we build a linear mixed model (Gallucci, 2019) with fixed factor *direction* (repeated contrast coding), random factor *subject*, and *logRT* as dependent variable with varying intercept and varying slope. Hence, the following formula was used: logRT ~ 1 + direction + (1 + direction|subject). Our model was a good fit for the data with R^2 marginal of 18.1 percent and R^2 conditional of 46.3 percent.

We found a main effect for direction ($F = 155.34$, $p < .001$) with asymmetric results in both direction comparisons revealed by fixed effect parameter estimates (see table 12). Post-hoc testing with holm correction for multiple comparisons showed clear asymmetries in the behavioural data between /o:/ - /e:/ and /e:/ - /o:/ ($t = -4.64$, $p < .001$) with faster logRT for /o:/ - /e:/, as well as between /o:/ - /u:/ and /u:/ - /o:/ ($t = -2.31$, $p < .05$) with faster reaction times for /o:/ - /u:/. Those asymmetric effects are also plotted in figure 16.

Table 12: Fixed Effects Parameter Estimates for logRT.

Names	Effect	Estimate	SE	df	t	p
Intercept	Intercept	5.789 (5.749, 5.826)	.020	28	292.11	< .001
Direction 1	**oe - eo**	**-.076 (-0.107, -0.044)**	**.016**	**28**	**-4.64**	**< .001.**
Direction 2	**ou - uo**	**-.032 (-0.060, -0.005)**	**.014**	**28**	**-2.31**	**< .05**

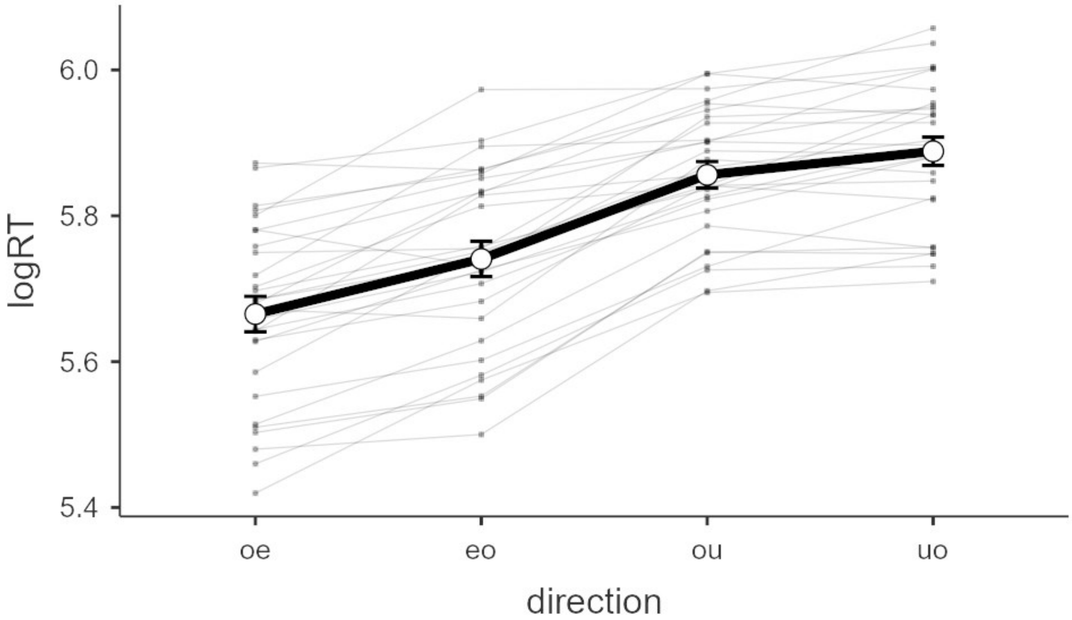

Figure 16: Effect plot of the linear mixed model for reaction times.
Grey lines depict the random variance of subjects modelled as random factor.

2.2.3 Discussion

In this reaction time study, we found asymmetric effects in both vowel oppositions. Presentation orders /o:/ - /e:/ and /o:/ - /u:/ produced faster reaction times than the corresponding inversions. Regarding the models, it must be stated that neither FUL nor both NRV assumptions of peripherality or focality can explain all found effects. For the latter contrast, the asymmetric effect for /o:/ - /e:/ is explained by the focality hypothesis, while the faster reaction times in /o:/ - /u:/ are in line with the peripherality approach. For FUL, only the asymmetry for /o:/ - /e:/ can be explained by this model but it fails to explain the faster reaction times of /o:/ - /u:/. Once more, we found in the behavioural data evidence for cue weighting due to phonetic distance. In principle, smaller acoustic distance can be defined as harder discrimination situation (Flege et al., 1994). Therefore, it has been shown that smaller phonetic distances result in slower reaction times compared to larger acoustic distances (Deguchi et al., 2010; M. Peltola, 2003; Scharinger, Domahs, et al., 2016). The same is true in our data set. Regarding cue weighting due to phonetic distance, there was not much evidence in the literature. Scharinger et al. (2016) found cue weighting regarding phonological features. In their study, they showed the usage of underspecified features in vowel oppositions with small acoustic distance. Our results are contrary to the results of Scharinger and colleagues. In our data set, it can be argued that in easier discrimination situations in vowel oppositions with large acoustic distance, sparse phonological and abstract features helped the discrimination process. Thus, the faster reaction times in /o:/ - /e:/ are explained by FUL. Smaller acoustic distance in /o:/ - /u:/, poses more difficulty in vowel processing and therefore the discrimination process relied on the phonetic and more detailed feature of focality. This interpretation that large acoustic distance leads to phonological vowel processing and small acoustic distance leads to phonetic vowel processing is further strengthened by the found cue weighting evidence of experiment 2a (MMN latencies and amplitudes).

But to be able to show a clearer picture of cue weighting in vowel processing and to investigate whether different cues appear in different measures as well as to see whether perceived loudness influenced iMMN amplitude data, we conducted first a rating study of the perceived loudness of the used vowels and then multiple regression analyses. These analyses are described in the next chapter.

2.3 Identifying the role of perceived loudness

2.3.1 Rating of perceived loudness

To investigate possible influences of perceived loudness on vowel processing, we conducted a rating study. Here, participants had to indicate whether they perceived a presented vowel quieter, louder, or as equal loud as the preceding sound. Hence, we adapted the procedure of Riedinger et al. (2021) (see also section 1.3.1 of this chapter).

Participants, Materials and Procedure

In this rating, we collected data of 23 participants (10 male, 13 female). They were between 21 to 35 years of age (mean age: 26.6). All subjects were monolingual German speakers with no hearing impairments and without neurological or psychological disorders.

We used the same experimental stimuli as in experiments 2a and 2b. Hence, all five tokens of German long vowels /e:/, /o:/ and /u:/ were presented in the same presentation orders as in the studies before. Vowels were presented in pairs. After each pair, participants had to indicate via button press which speech sound they perceived as louder (button 1 or button 2) or if they were equally loud (button 3). In total, the rating study was interested in the perceived loudness of vowel pairs of the two previous investigated vowel contrasts (/o:/ - /e:/, /o:/ - /u:/) á two presentation orders. Because we had five tokens per vowel in each presentation order, there were 25 possible combinations per presentation order, resulting in 100 critical trials overall. To get better randomisation, we included a filler contrast (/u:/ - /e:/) with both directions as filler items. Hence, we presented an additional 50 filler trials resulting in a total of 150 trials. Trials were arranged in six blocks across ten randomised stimuli lists, where not only block order but also stimuli varied randomly. Here, it was ensured that a maximum of two trials of the same presentation direction directly followed each other.

We used PsychoPy (Peirce et al., 2019) for programming the study. Due to COVID, the rating took place as an online study via pavlovia (www.pavlovia.org). Participants were informed before the start of the experiment that they had to use headphones and participation was possible only with a desktop computer or a laptop. Before the experiment started, subjects had consented and filled out a short questionnaire. The criteria having been met, the experiment started automatically. Since in online experiments fixed sound volume of the stimulus presentation cannot be directly ensured, the rating started with a short volume calibration phase. Here, we presented six pairs of vowel stimuli (German umlauts) of another study and participants were asked to set the volume to a comfortable listening level. After completion of the calibration phase, the rating study started. In table 13 the presentation specifications of one trial are displayed.

Table 13: Presentation specification of a trial in the online rating experiment of the perceived loudness.

Trial part	Blank	Vowel 1	ISI	Vowel 2	Blank	Answer
Time (in ms)	600	400	600	400	600	max. 2500
Display			*			??

Analysis and Results

After data collection, we first excluded timeouts. For the remaining data, frequency values of the answer categories (first vowel louder, second vowel louder, equally loud) were calculated. The distribution of each answer within each direction is plotted in figure 17. Note that the last two plotted directions (eu and ue) were filler directions and will be excluded from all further analyses.

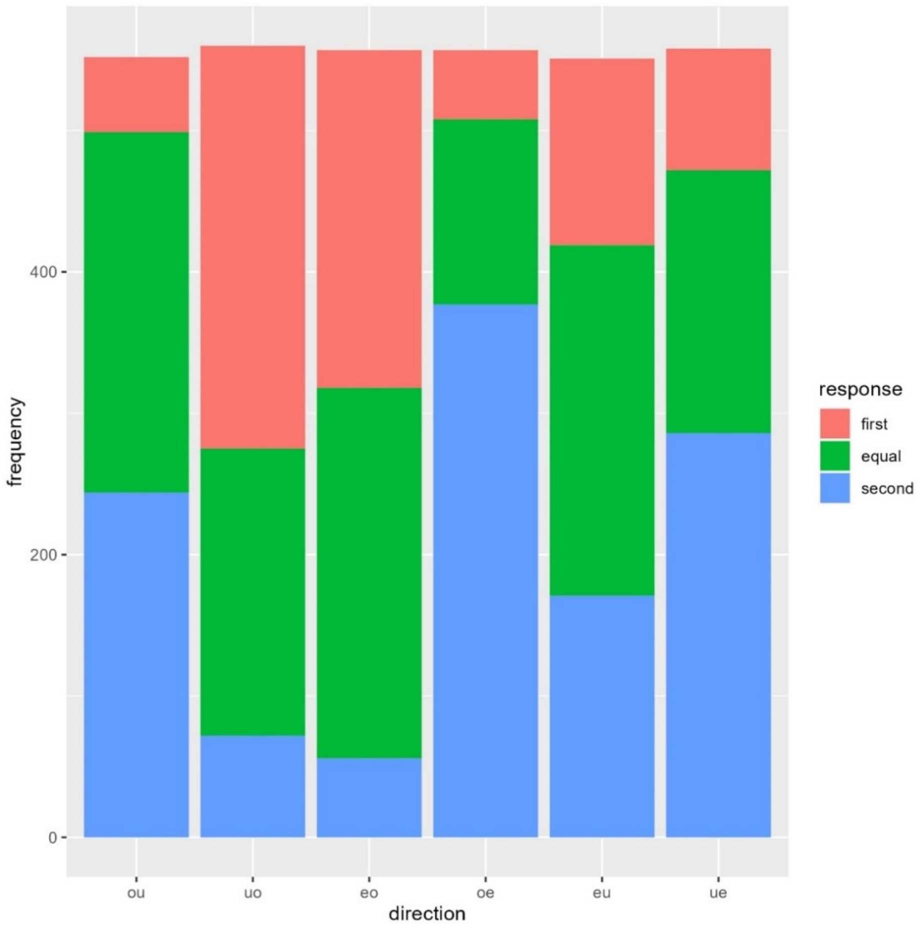

Figure 17: Frequency distribution of answer categories for each direction of the perceived loudness rating.

A Pearson's chi-square test was calculated in jamovi (The jamovi project, 2021) with variables direction (4) and answer (3), which showed a highly significant relationship between both variables ($\chi^2_{(6)} = 714.56$, $p < .001$).

To test later for an influence of the perceived loudness on amplitude and latency measurements of the iMMN data, the frequency data was transformed to percentages (see table 14). The answer category within each direction with the highest percentage will be used in all further analyses.

Table 14: Distribution of the answers for the perceived loudness of each vowel pair (in percent).

Direction	Equal	First	Second
/o:/ - /e:/	23.5	8.8	**67.7**
/e:/ - /o:/	**47.0**	43.0	10.0
/o:/ - /u:/	**46.2**	9.6	44.2
/u:/ - /o:/	36.2	**50.9**	12.9

As can be seen, in the direction /o:/ - /e:/ the second vowel /e:/ was mostly rated as louder. In the reverse order, both vowels were more often perceived as equally loud. The same was true for presentation order /o:/ - /u:/. In contrast, in the reverse direction /u:/ - /o:/ the first vowel /u:/ was mostly perceived as louder.

In the first vowel contrast, it can be hypothesized that a change from back to front vowels leads to greater perceived loudness differences and could therefore be perceptually more distinct than vice versa. In the second contrast, the same could be true for changes towards the upper tongue height dimension (e.g., change from [MID] to [HIGH]).

2.3.2 Influence of perceived loudness on iMMN amplitudes

Since the FUL as well as the NRV model failed to explain all effects found in the iMMN amplitude analysis, we decided to investigate this data set further by a multiple regression. Here, we should be able to identify influential factors on iMMN amplitude measures such as psychoacoustic perceived loudness (Riedinger et al., 2021). All following analyses are done with jamovi (The jamovi project, 2021)

Defining Factors

For the regression analysis, we operationalised one empirically driven *(perceived loudness*) and three theoretically driven factors (*specificity*, *peripherality*, *focality*).

The empirically motivated factor of *perceived loudness* has been structurally based on the before described results of the rating study. Here, the factor has been defined as a function of the answer with the highest percentage per presentation direction (first louder, second louder, equally loud).

The theoretically driven factors have been operationalised with regard to the models used in this dissertation. While *specificity* is related to the FUL model, *peripherality* and *focality* adapt the different assumptions made in NRV. *Specificity* (difference value between the number of abstract features in the deviant minus the number of features in the standard) incorporates assumptions of underspecification as well as additivity effects found in other studies (Riedinger et al., 2021; Scharinger, Domahs, et al., 2016; Scharinger, Idsardi, et al., 2011a; Scharinger, Merickel, et al., 2011). The remaining factors *peripherality* and *focality* incorporate the hypotheses made by NRV regarding the perceptive salience of the vowel space periphery (Polka & Bohn, 2011a) and convergence of adjacent formants (Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019), respectively.

Multiple Regression Analysis

First, we calculated Kendall's Tau correlation between all factors and the dependent variable iMMN amplitude. This analysis revealed only a significant correlation between the dependent variable and factors *perceived loudness* ($\tau = -0.255$, $p < .001$), *specificity* ($\tau = 0.184$, $p < .001$) and *focality* ($\tau = 0.168$, $p < .001$). Thus, peripherality will be excluded from the regression analysis.

Following, we conducted a hierarchical multiple regression analysis with three steps in the following order: perceived loudness (model 1), specificity (model 2), focality (model 3). Model fit measures and model comparison revealed that the best fit for our iMMN amplitude data set is model 2 (see table 15). Here, perceived loudness as well as specificity have influenced the iMMN amplitude results (14.7 percent variance explained), with greater influence of perceived loudness (11.3. percent). The analysis revealed in detail the following two arguments: (1) With increasing differences in perceived loudness iMMN effects get stronger; (2) with decreasing specificity and increasing feature distance iMMN effects increase. These tendencies are also displayed in figure 18.

Table 15: Multiple regression models for the iMMN amplitudes.
Model comparison showed that Model 2 has the best fit (F = 21.41, p < .001).

Model#		b	SE B	β	p
2	Intercept	-1.206 (-1.312, -1.091)	.059		< .001
	Loudness	-.604 (-0.771, -0.437)	.085	-.296	< .001
	Specificity	.176 (0.101, 0.250)	.038	.192	< .001
R^2 = .150, ΔR^2 = .147, p < .001					

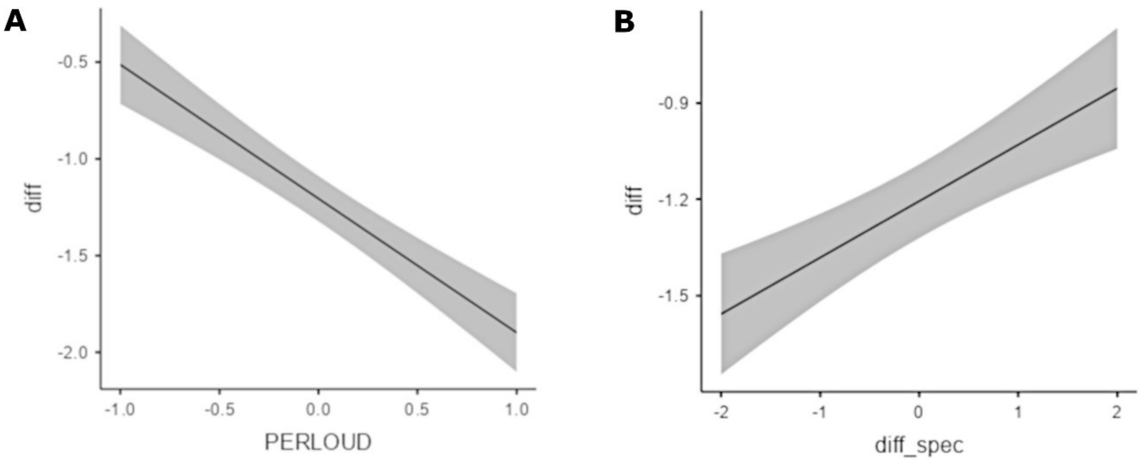

Figure 18: Regression plots for perceived loudness and specificity.
While asymmetric neural effects are increased with increasing perceived loudness, higher specificity differences lead to stronger asymmetric effects.

2.3.3 Influence of perceived loudness on iMMN latencies

The same procedure as in 2.3.2 was implemented for the iMMN latency dataset. Correlation analysis showed that only factors *specificity* (τ = 0.234, p < .01) and *peripherality* (τ = 0.273, p < .01) had a significant correlation with iMMN latency. Therefore, *perceived loudness* and *focality* are excluded from the regression analysis. Again, a hierarchical multiple regression with specificity (model 1) and peripherality (model 2) was calculated. Model comparison showed that model 2, with both factors had the better fit. Key values of model 2 are displayed in table 16. Hence, specificity and peripherality had influenced latencyeffects. In detail the following tendencies can be identified: (1) with decreasing specificity and increasing feature distance iMMN latency decreases; (2) with decreasing peripherality iMMN latency also decreases. These linear trends are displayed in figure 19.

Table 16: Multiple regression models for the iMMN latencies.
Model comparison showed that Model 2 has the better fit (F = 4.505, p < .05).

Model#		b	SE B	β	p
2	Intercept	.160 (0.155, 0.166)	.003		< .001
	Specificity	.004 (0.001, 0.008)	.002	.261	.023
	Peripherality	.010 (0.010, 0.020)	.005	.239	.037
R^2 = .172, ΔR^2 = .150, p < .001					

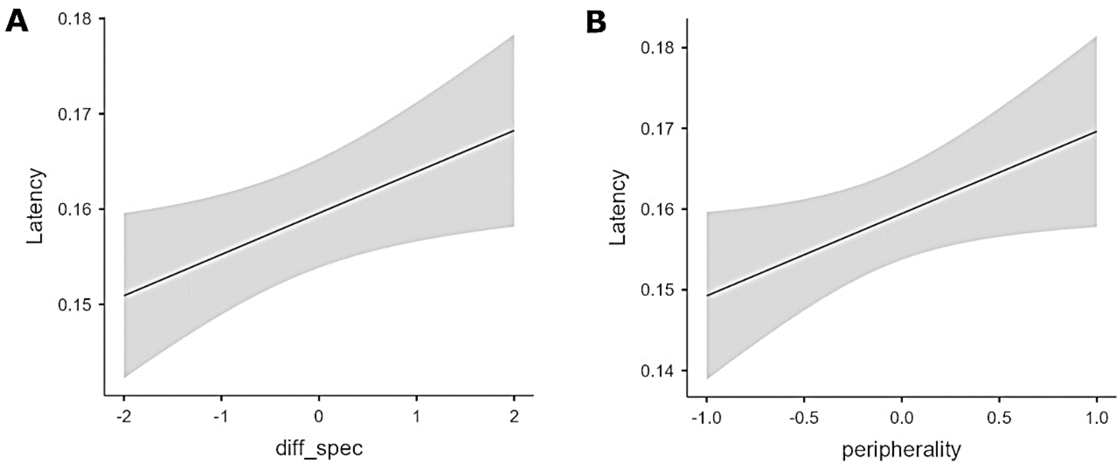

Figure 19: Linear trends of iMMN latency for specificity (A) and peripherality (B).
iMMN latencies increase with increasing specificity and peripherality.

Alternative Model

Since we found an influence of word frequency on latency values (see chapter 1.1.6) in the first iMMN study, we wanted to test whether frequency of occurrence of vowels had an impact in this study also. Kendall's Tau correlation revealed that *vowel frequency* is highly correlated with factors *specificity* (τ = -0.813, p < .001) and *peripherality* (τ = -0.607, p < .001) as well as a moderate correlation with the dependent variable of iMMN latency (τ = -0.377, p < .001). Thus, we had to build a separate regression model for the factor *vowel frequency* (see table 17). This model shows that vowel frequency of occurrence has a significant influence on MMN latency effects with 15 percent of variance explained by this factor. Figure 20 shows that with increasing vowel frequency the iMMN latency decreases.

Table 17: Alternative model for iMMN latencies.
Regression analysis of vowel frequencies on iMMN latencies.

Model#		b	SE B	β	p
1	Intercept	.160 (0.155, 0.166)	.003		< .001
	vowfreq	-.011 (-0.005, -3.845)	.003	-.401	< .001
R^2 = .161, ΔR^2 = .150, p < .001					

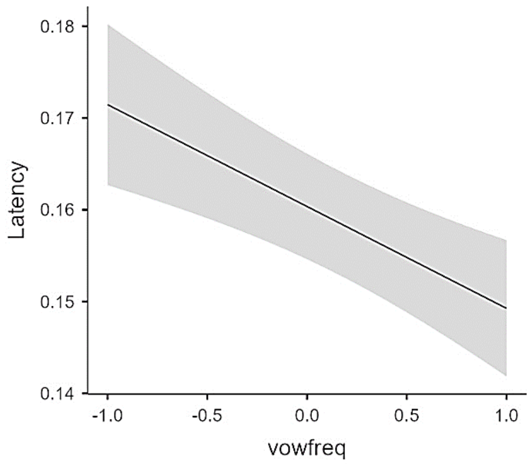

Figure 20: Linear Trend of Latency regarding vowel frequency.
Latencies of iMMN decrease with higher vowel frequency of occurence of the deviant.

Afterwards we evaluated both possible regression models (**model A**: *specificity* and *peripherality*, **model B**: *vowel frequency*) regarding a better fit for the dataset. Thus, we compared adjusted R^2 values and RMSE values but both key markers did not differ between model A and model B (ΔR^2 = .150, RMSE = .025). Therefore, we additionally compared AIC (Akaike information criterion) and BIC (Bayesian information criterion) values. Lower AIC and/ or BIC would indicate a statistically better fit. Here, we found a slight difference: Model A (AIC = -352, BIC = -342) seems to be a marginally better fit than model B (AIC = -353, BIC = -346). Since the differences were only minimal, both models have their justification in explaining the data and found effects.

2.3.4 Discussion

In this chapter, we investigated the role of perceived loudness and other factors on vowel processing especially when stimuli are ignored. Therefore, we conducted several regression analyses separately for MMN amplitudes and latencies on the neural data.
For the analyses of the amplitude and latency we operationalised four potentially influential factors: *perceived loudness* (Riedinger et al., 2021), *specificity* (Lahiri & Reetz, 2002a, 2010a; Scharinger, Domahs, et al., 2016) as well as *peripherality* (Polka & Bohn, 2011a) and *focality* (Y. Liu et al., 2021; Masapollo et al., 2015; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Polka et al., 2019).
The regression analysis on the amplitude data showed influences of *perceived loudness* as well as *specificity*. First, it was evident that with increasing perceived loudness, MMN effects get stronger. These results are in line with the influence of perceived loudness found in the previous series of studies (Riedinger et al., 2021). Note that in contrast to the studies with real words in chapter 1, we were better able to control for the known influence of perceived loudness like fundamental frequency (Hsu et al., 2015) and duration (Todd & Michie, 2000). Second, regarding the specificity of the vowels it became clear that with decreasing specificity (under the assumptions of underspecification) and feature distance, MMN amplitudes increase. This result confirms the important influence of underspecification accounts and the role of sparse, abstract phonological features in vowel processing (e.g., Eulitz & Lahiri, 2004; Lahiri & Reetz, 2010a; Scharinger, 2009; Scharinger, Monahan, et al., 2012). Different arguments regarding automatic vowel processing can be drawn from these results. As in Riedinger et al. (2021), we could show an influence of the psychoacoustic factor of perceived loudness in vowel processing. Therefore, we were able to replicate the results in isolated German long vowels as well. This strengthens the evidence that psychoacoustic mechanisms play a greater role in vowel processing than previously thought. Regarding the influence of specificity on MMN amplitudes, the results confirm the interpretation of the MMN amplitude asymmetry in /o:/ - /e:/ in line with FUL and therefore underspecification. Furthermore, it can be argued that not only feature mismatches but also featural distances seem to influence MMN amplitudes (Scharinger, Domahs, et al., 2016; Scharinger, Monahan, et al., 2016).
The regression analyses on the MMN latency data were conducted with the same factors (perceived loudness, specificity, peripherality and focality). Here, we found an influence of factors *specificity* and *peripherality*. Hence, perceived loudness seems to influence only MMN amplitudes (and interpretations) but not MMN latencies. Importantly, we found the same influence of specificity in the latency as well as in the amplitude data. The results therefore strengthen the role of sparse phonological features in vowel processing. Furthermore, the result of this

regression analysis revealed further evidence for the cue weighting argumentation proposed for the interpretation of MMN latencies. It can be argued, since only the vowels in /o:/ - /u:/ differed in terms of peripherality, that the factor of peripherality could influenced only this vowel opposition but the results were opposite to the initial hypothesis. Here, we found that with decreasing peripherality, i.e., discrimination towards a more central vowel, MMN latencies decreased as well. Reasons for this pattern could not be determined in these analyses and studies. Importantly, the vowel opposition of /o:/ - /e:/ was heavily influenced by specificity as has been shown in the MMN amplitude discussion before. Therefore, there is evidence from this regression analysis on latency data that once again cue weighting occurred in relation to the acoustic distance of vowels. With larger acoustic distance phonological features were preferred, while smaller acoustic distances led to favouring of phonetic features. This cue weighting pattern is also in line with the previous described pattern in the behavioural data.

Since we found an influence of word frequency of occurrence on MMN latencies in experiment 1a, we conducted an alternative regression analysis of latencies with vowel frequency of occurrence. There are reports that word frequency (Aleksandrov et al., 2017; Shtyrov et al., 2011) as well as phonotactic probabilities or biphone probabilities (Bonte et al., 2005; Emmendorfer et al., 2020, 2021) can influence neural effects, which show reduced latency effects with increasing frequency or probability. Most studies regarding vowel processing do not investigate specifically the influence of vowel frequency of occurrence. Therefore, it seemed reasonable to investigate (on an explorative level) whether vowel distributional frequencies influenced our latency data. The regression analysis shows clear evidence for an influence of vowel frequency of occurrence: With increasing vowel frequency, the MMN latency decreases. These results are in line with studies regarding word frequency and phonotactic probabilities (Aleksandrov et al., 2011, 2017; Bonte et al., 2005; Emmendorfer et al., 2020, 2021; Shtyrov et al., 2011; Yasin, 2007). Furthermore, correlation prior to regression revealed a strong collinearity between factors specificity and vowel frequency as well as between vowel frequency and peripherality. Regarding the latter collinearity, the analysis revealed that, mostly in contrast /o:/ - /u:/, with decreasing peripherality the vowel frequency of occurrence increased. It can be argued for a possible relationship between vowel frequency and peripherality within the NRV framework. Polka & Bohn (2011a) described that the natural vowels /i/, /a/ and /u/ are not only preferred in language processing but also in child language production. Furthermore, these anchor vowels are also typologically more frequent across vowel inventories. It is possible that these vowels are not only typologically preferred but are also used more commonly in languages. Regarding a relationship between underspecification and vowel frequency, we found that with decreasing specificity, i.e., deviant less specific than standard, and therefore containing a clashing feature in the surface structure, vowel frequency increased because the deviant was a more frequent vowel type than the standard. The reasons for this relationship between underspecification and frequency remain unclear. Additionally, it is unclear if the connections between vowel frequency and feature specifications in our models are limited to our stimuli set or if they can be generalised. There could be a possible connection between these factors since underspecification is in theory mostly motivated by the concept of markedness which is also linked with frequency (Haspelmath, 2006; Hume, 2008).

Within this analysis, it was not possible to investigate these correlations further. But different lines of argumentations regarding a general correlation between vowel frequency of occurrence in a given language and factors like peripherality and specificity will be discussed in chapter IV of this dissertation.

2.4 General Discussion

The set of experiments and analyses discussed in this chapter was conducted to investigate directional asymmetries in vowel processing. In this second set of studies, we decided to reduce linguistic context to see if we could replicate previously found influences and effect patterns in isolated German long vowels. Therefore, we conducted an iMMN and an RT experiment with the vowel contrasts /o:/ - /e:/ and /o:/ - /u:/. Both contrasts were tested in the experiments bidirectionally. While the iMMN experiment required no attention on the stimulation (passive oddball), the RT experiment needed attended discrimination of the vowels (active oddball). Additionally, the contrasts were chosen with regard to differences in feature dimensions and acoustic distance: While the vowels in the first contrast /o:/ - /e:/ differed mainly in the dimension of place of articulation and kept a larger acoustic distance, the vowels in the second investigated contrast varied in relation to tongue height and were distinguished by a small acoustic distance.

In respect to directional asymmetries in vowel processing, we found the following the results. In experiment 2a (iMMN), in amplitudes we found only asymmetric results in the contrast /o:/ - /e:/ with presentation order /o:/ - /e:/ producing larger iMMN amplitudes than in the reverse condition. In the contrast /o:/ - /u:/, amplitudes obtained a symmetric pattern. Regarding iMMN latencies, both contrasts showed asymmetric effects. /o:/ - /e:/ and /u:/ - /o:/ produced shorter latencies then the respective inversions. In study 2b (RT), both contrasts again displayed asymmetric effect patterns: While reaction times in the presentation order /o:/ - /e:/ were faster than those for the inversion, /o:/ - /u:/ tested significantly faster than /u:/ - /o:/. Thus, we can summarise that we found directional asymmetries in most of the measures and experiments. Only in amplitude measures of contrast /o:/ - /u:/ a symmetric effect emerged.

The set of experiments aimed to investigate the possible reasons behind emerging directional asymmetries or symmetric effects. Therefore, we formulated three research aims. Firstly, we wanted to compare the models of FUL and NRV, being the most prominent models in explaining directional asymmetries, and therefore shed more light on features involved in vowel processing (**model comparison**). Secondly, we wanted to verify **the influence of psychoacoustic perceived loudness,** discovered in the first series of studies in this dissertation, in isolated long vowel processing. Since we found cue weighting patterns due to attention requirements related to the subjects' task in the experiments described in section 1 of this chapter, we wanted to test **cue weighting** patterns in more detail. Therefore, the third aim of the second series of experiments was to identify cue weighting patterns not only in relation to attention but also in relation to acoustic distance.

Model comparison

Overall, we found strong evidence for vowel processing utilizing sparse and abstract, phonological features and therefore vowel processing according to FUL in all experiments and analyses. In iMMN amplitudes, latencies, and as well as RT data, all effects found in contrast /o:/ - /e:/ can be solely explained by this model. In the regression analysis of the neural data, we could additionally show that asymmetric effects are not only driven by feature specification but also by the featural distance between the vowels. With decreasing feature specification and increasing featural distance, the neural effect increased, and processing time reflected by latencies decreased. The findings are in line with the vast corpus of research literature investigating underspecification and vowel processing with abstract, phonological features (e.g., de Rue et al., 2021; Eulitz & Lahiri, 2004; Obleser et al., 2004; Scharinger, 2009; Scharinger, Domahs, et al., 2016; Scharinger, Monahan, et al., 2012, 2016). With regard to the NRV model, we only found limited evidence for vowel processing due to focality or peripherality of vowels in the vowel space. In the RT data, the faster reaction times for /o:/ - /u:/ can only be explained by the peripherality hypothesis of NRV (but not the hypotheses of FUL) and are in line with other studies (Polka & Bohn, 2011a; Verá-Constán & Sebastián-Gallés, 2008). Additionally, it could be argued that the asymmetric iMMN latency in /u:/ - /o:/ could be due to a reversed peripherality effect and therefore be somewhat explainable by the NRV approach. But this result could also be explained by FUL.

According to the hypothesis of the underspecification approach, a feature mismatch of tongue heights [HIGH] and [MID] has driven shorter latencies.

To conclude, in this series of experiments the FUL model can explain found asymmetric effects the best. Only in the RT data in the second vowel contrast, this model fails to predict the found asymmetric pattern. Additionally, the found symmetric effect in iMMN amplitudes cannot be explained by either of the implemented models.

Influence of perceived loudness

In Riedinger et al. (2021), we found an influence of the psychoacoustic factor perceived loudness on iMMN amplitudes, i.e., the amplitudes are scaled with increasing perceived loudness differences, but not behavioural data. Therefore, we tested the neural data again in this second series of studies. In regression analyses we found in regression analyses an effect of perceived loudness on the iMMN amplitudes but not on latencies. Hence, perceived loudness had an influence on effect sizes or processing costs but not processing time. Again, iMMN amplitudes were scaled by differences in perceived loudness. With increasing perceived loudness, the neural effects increased as well. This can be interpreted by implication that with decreasing perceived loudness difference, neural effects get weaker. Therefore, we argue that psychoacoustic perceived loudness is the main driving factor behind the found symmetry in iMMN amplitudes in the vowel contrast /o:/ - /u:/. Additionally, it also could be a contributing factor behind the strong asymmetric effect found in the other contrast /o:/ - /e:/. Thus, we were able to replicate the effect of perceived loudness found in Riedinger et al. (2021). Importantly, since we were able to replicate an influence of perceived loudness in the neural data in isolated long vowels, evidence is accumulating that perceived loudness plays an important role in vowel processing. There is evidence of two studies regarding perceived loudness in relation to vowel quality describing that different vowel qualities covary with differences in perceived loudness (Glave & Rietveld, 1975, 1979). Additionally, more recent studies have provided evidence that with increasing vowel duration (Todd & Michie, 2000) as well as differences in fundamental frequency (Hsu et al., 2015) perceived loudness difference should increase as well. Here, we argue that durational differences and changes in fundamental frequency cannot explain the found perceived loudness differences. Since we used isolated long vowels which were also cut to equal length, we were able to control for duration and f0 differences which were minor. Therefore, we argue that perceived loudness differences stem from different vowel qualities and that perceived loudness differences therefore are utilized to discriminate vowels in an unattended experiment protocol. Hence, unattended vowel processing could be largely determined by psychoacoustic factors as opposed to linguistic influences. The connection behind vowel processing and perceived loudness or the reasons for differences in perceived loudness between different vowel qualities cannot be determined in these studies and need further exploration and investigation.

Cue weighting

The found influence of perceived loudness leads to a perspective of cue weighting in vowel processing. In our series of studies in this chapter, the vowels in the contrasts differed mainly in acoustic distance (as well as feature dimension) and experiment protocols varied regarding attention allocation on the stimulation. Therefore, it is possible to investigate found effects and influences in the data in relation to changes in acoustic distance and attention requirement and thus, identifying cue weighting patterns. We found different cue weighting patterns in German long vowel processing varying between the features used in discrimination (phonological or phonetic or psychoacoustic) despite ensuring phonological encoding in our experiment protocols. Johnson (2015) showed that short ISIs triggered auditory processing while long ISIs resulted in phonemic processing. He tested in an AX discrimination task the discrimination of German vowels [e], [o] and [ø] with either a short ISI of 100 ms or a long ISI of 700 ms and found different discrimination strategies regarding the manipulation of the length of the interstimulus interval. Here, we argue that with a random ISI between 600 and 1100 ms in the EEG experiment and an ISI of 400 ms in the RT study, we chose rather long interstimulus periods. Hence, our subjects discriminated the vowels phonemically. Additionally, we introduced allophonic variance in standards and deviants using five tokens of each vowel as experimental stimuli. This method has proven to ensure phonemic processing as well (Hestvik & Durvasula, 2016; Phillips et al., 2000). Despite these efforts, our data shows that in all experiments and measures there is evidence for cue weighting, often involving more acoustic or auditory based features than expected in phonemic processing.

The following cue weighting patterns can be identified:

(1) In MMN amplitudes the asymmetric pattern is best explained by feature specification according to FUL (phonological factor) while the symmetric results show the influence of perceived loudness (psychoacoustic factor).

(2) In MMN latencies it can be argued that the asymmetric pattern of /o:/ - /e:/ is due to phonological features, while the asymmetry of /o:/ - /u:/ is influenced by peripherality (a phonetic, more acoustically based factor).

(3) In RT data, the asymmetry of /o:/ - /e:/ is determined by featural mismatches as predicted by FUL (phonological factor) and in contrast, the asymmetric result in the second vowel contrast is driven by peripherality differences (phonetic factor).

These patterns could emerge because of the different acoustic distances inherent in the vowel contrasts as well as different attention requirements stemming from the experimental set up. In several other studies, it has been proven that attention modulates discrimination strategies and neurocognitive effects (Alain & Woods, 1997; Remijn & Kojima, 2010; Riedinger et al., 2021; Scharinger et al., 2017; Szymanski et al., 1999; Tuomainen et al., 2013; Vannest et al., 2009). In our experiments, we also found evidence for different discrimination strategies due to attention allocation on the stimulation. In the unattended experiment setting (passive oddball) vowel discrimination was driven by phonological but crucially also psychoacoustic factors. In contrast, with attention on the stimuli (active oddball), psychoacoustic factors became less important, and discrimination was successfully utilizing phonological and phonetic factors. Additionally, since perceived loudness showed only an influence in the MMN amplitude data, there is possible evidence that different measures within the same protocol reflect different cues involved in vowel processing.
Interestingly, we also found evidence for cue weighting involving different factors due to acoustic distance of the vowels within the vowel contrasts. Overall, we found faster reaction times, stronger iMMN amplitudes and faster iMMN latencies with increasing acoustic distance. These relationships between acoustic distance and neural and behavioural effects were also found in other studies (Deguchi et al., 2010; M. Peltola, 2003; Savela et al., 2003; Scharinger, Domahs, et al., 2016; Scharinger, Monahan, et al., 2016). For example, Scharinger, Monahan, and colleagues (2016) found differences in MMN amplitudes in an MMN study with English vowels /ɪ /, /ɛ/, and /æ/. In contrasts with large F1 distance MMN effects were stronger than in contrasts with small F1 effects. Savela et al. (2003) found similar effects in their MMN data but also in reaction times: "the RT decreased and the MMN amplitude increased when the acoustic difference between stimuli increased" (Savela et al., 2003, p. 250). On closer inspection of our results, we were able to identify vowel processing utilizing different features and factors due to acoustic distance. In the case of large acoustic distance in vowel contrast /o:/ - /e:/, vowel processing was determined by the hypotheses of the FUL model and hence had been successfully completed with sparse and abstract phonological features. In the opposite case, in the vowel contrast with small acoustic distance, investigations showed that either the psychoacoustic factor of perceived loudness or phonetic features of vowel space periphery were used. This cue weighting pattern could be

explained by the easiness or difficulty in vowel processing implicated by acoustic distance. In principle, smaller acoustic distances pose harder discrimination situations (Flege et al., 1994) and by implication larger acoustic distances lead to easier discrimination. Hence, in cases of easier discrimination with larger acoustic distances, sparse, abstract phonological features are sufficient to discriminate between two vowels. In situations where the acoustic distance is small, hence vowels are harder to discriminate, more detailed and acoustically based information are needed to successfully distinguish the vowels from one another.

To conclude, we found strong evidence for the involvement of underspecified and abstract, phonological features in vowel processing as well as cue weighting or different discrimination strategies due to attention allocation or acoustic distance of the vowels. Furthermore, we again found that psychoacoustic factors like perceived loudness have a great impact on iMMN amplitudes. Therefore in the future, it would be necessary to include perceived loudness not only into the testing of MMN effects but also to incorporate psychoacoustics more strongly into models of vowel processing.

IV Conclusion

This dissertation investigated the commonly found directional asymmetries in vowel processing in-depth. In the literature, two models which make specific predictions regarding directional asymmetries are often discussed: the Featurally Underspecified Lexicon (Lahiri & Reetz, 2002b, 2010b) and the Natural Referent Vowel framework (Polka & Bohn, 2011b). While the latter utilises phonetic features to explain vowel discrimination and directional asymmetries, the former model incorporates abstract phonological features. Hence, their explanatory approaches are different, and the hypotheses drawn from each model are often contradicting each other. Therefore, this dissertation set out to compare both models regarding directional asymmetries in German vowel processing and to investigate which model explains the found asymmetries the best. Thus, the dissertation tried to answer the following research question: Are either phonological or phonetic features involved in German vowel processing?

This chapter will draw a summary of the empirical investigations and will give a conclusion to the found results before ending with an outline of further research approaches and remaining research questions.

Summary

As the research question indicates, the methodological design of this empirical dissertation was based on a comparison of the two models and their specific hypotheses regarding neural and behavioural directional asymmetries.

In total, the empirical investigations were comprised of two experiment series with stimuli of varying complexity, whereby the methodological design was always the same: First, an iMMN study was conducted, followed by an RT experiment. Finally, the perceived loudness of the stimuli was rated, and regression models were calculated to pinpoint influences in German vowel processing. While the first series investigated directional asymmetries of vowel perception in word contexts, the second series focused on influences on vowel processing in isolated contexts. A summary of all results is given in table 18.

Table 18: Summary of the results and found influences in the investigations of German vowel processing.

Experiment	Measure	Result	Influence
1a	iMMN amplitude	symmetry & asymmetry	perceived loudness
	iMMN latency	symmetry & asymmetry	frequency
1b	RT	symmetry & asymmetry	phonological
2a	iMMN amplitude	symmetry & asymmetry	perceived loudness & phonological
	iMMN latency	asymmetry	phonological & phonetic
2b	RT	asymmetry	phonological & phonetic

In the first set of experiments, we found directional asymmetries as well as symmetries in iMMN amplitudes and latencies (experiment 1a) and RT results (experiment 2a). Together with the linear regression analysis, it was revealed that in iMMN amplitudes perceived loudness drove the asymmetric and symmetric effects whereas frequency of occurrence influenced iMMN latencies. Finally, the found effects in reaction time data are driven by phonological features and are therefore explainable within the FUL model. In the second experiment series, we again found asymmetries and symmetries in iMMN amplitudes but only asymmetric results in iMMN latencies (experiment 2a) and RTs (experiment 2b). Here, the linear regression analyses revealed a more complex pattern of influences. While the symmetric results in the iMMN amplitudes are driven by perceived loudness (rated more equal), the asymmetric results are influenced by phonological features and are, therefore, explainable by FUL. The asymmetric iMMN latency effects are influenced by phonological features as well as phonetic properties (peripherality). Here, regression analyses showed that frequency of occurrence could also be a possible explanation. But frequency and the cues investigated, phonological as well as phonetic ones, were highly colinear. The asymmetric results in the RT data are also influenced by phonological and phonetic features. Hence, the RT data as well as iMMN latencies are only explainable by the FUL model together with the NRV framework.

Conclusion

Overall in the investigations of asymmetric German vowel processing, we found more complex patterns of influence than proposed by the research question. It has been proven that none of the models can explain vowel processing in German alone. Thus, to answer the above-mentioned research question: In German vowel processing, phonetic as well as phonological features are involved and necessary.

Furthermore, the complex pattern of influences and the different empirical findings are related to the complexity of stimuli, attention or task, and chosen measure. Overall, there is strong evidence that discrimination strategies adapt, especially with regarding different attention allocation requirements and chosen stimuli. Therefore, different cue weighting patterns can be identified in the synopsis of our experimental investigations:

(1) **Chosen measures:** While neural amplitude measures represent psychoacoustic and phonological influences of vowel discrimination, neural latency measurements are influenced by the frequency of occurrence. In attended decision making (RTs) the influences of phonological and phonetic features become apparent.

(2) **Linguistic context:** While in isolated vowel discrimination phonological as well as phonetic features are used, in more complex linguistic contexts more abstract phonological features are preferred.

(3) **Level of difficulty (phonetic distance):** While in contrasts with large phonetic distance (and therefore in easier discrimination situations) sparse, abstract phonological features are utilised, in more difficult contrasts with smaller phonetic distance fine-grained phonetic cues are prioritised.

(4) **Attention allocation:** While in unattended tasks the psychoacoustic influence of perceived loudness is utilised for vowel discrimination, in attended discrimination phonological features play a crucial role.

Hence, it has been proven that the research question needs to be extended. Due to the various cue weighting patterns found in the described investigations, vowel processing, and therefore the features involved in vowel discrimination, is influenced by experimental setting, chosen stimuli, level of discrimination difficulty, i.e., phonetic distance, and attention allocation. The described cue weighting patterns should, therefore, additionally and carefully be considered in further investigations. Importantly, since we found a robust influence of psychoacoustic cues like perceived loudness, further research regarding directional asymmetries and speech sound processing should include this factor in investigations and model building leading to the following extended research question: In which contexts are either phonological, phonetic, or psychoacoustic features utilised in speech sound discrimination?

From this new research question, it becomes apparent that this research area still offers many open desiderata and approaches for further investigation. Therefore, the following section offers an outlook on further research.

Future research

In summary, different research desiderata can be identified and new research questions are still awaiting answering.

We already described a possible relationship between specificity, peripherality and frequency of occurrence in chapter III (section 2.3). In the stimulus set used in the second series of experiments, we found a high collinearity of factors specificity, which incorporates underspecification, peripherality and frequency of occurrence. These relationships between frequency of occurrence and phonological and phonetic features of vowels should be further investigated. Here, cross-linguistic and typological investigations could also be helpful. From a typological perspective, the collinearity between peripherality and frequency is cross-linguistically evident since across vowel inventories peripheral vowels, especially quantal vowels, are preferred (Polka & Bohn, 2011b; Schwartz et al., 1997a, 2005). But it remains unclear whether peripheral vowels, e.g., /i/, /a/, /u/, are also most frequently used within languages. By identifying the most common vowels within vowel inventories, it should be possible to match frequency and peripheral position in the vowel space. Also of typological interest is the connection between specificity together with underspecification and frequency. Within our second stimulus set, mismatching deviants, i.e., bearing underspecified features and being less specific, are more frequent within the German vowel inventory. In theory, these factors are usually connected via the concept of markedness. Even though the concept is controversial (Haspelmath, 2006), it could possibly be used as an approach for incorporating specificity and peripherality into linguistic theories and models of language processing. Given the traditional interpretations of markedness, that typical or more common patterns are unmarked and atypical or rare ones are marked (Haspelmath, 2006; Hume, 2004), markedness can also be interpreted in terms of salience (Rice, 2007). Here, the marked patterns are more salient. Therefore, it can be argued that more peripheral as well as more specific vowels are marked. Since Haspelmath (2006) as well as Hume (2004, 2008, 2006) promote a reinterpretation of the markedness concept in favour of frequency and predictability, frequency of occurrence (as indicator for predictability) could be also a factor in explaining differences in specificity and peripherality between speech sounds on a theoretical level. The idea of frequency and predictability as unifying concepts in explaining differences of specificity and peripheral positions in vowel spaces could be an interesting new research desideratum.

In connection with the above described research idea and to gather more evidence for the cue weighting patterns described in the conclusion, it would be interesting to test more German vowels using the same methodological approach. Here, it would also be beneficial to test vowel contrasts in other languages to compare cue weighting patterns across languages.

Another remaining research field is related to the FUL and NRV model. For both models, there is an ongoing debate whether feature specification or resulting directional asymmetries are influenced by language experience and other language dependent factors or if the approaches are universally consistent (Hestvik et al., 2020; Lahiri, 2018; Lahiri & Reetz, 2010b; Lipski et al., 2007; Masapollo, Polka, Molnar, et al., 2017; Polka & Bohn, 2003b, 2011b; Scharinger & Idsardi, 2014; Scharinger & Lahiri, 2010). Hence, cross-linguistic examinations could shed light on language-dependent asymmetries or the universality of directional asymmetries.

The largest remaining new research desideratum concerns the psychoacoustic factor of perceived loudness. In both iMMN experiments we found a robust influence of perceived loudness on German long vowel processing. The exact nature of this influence is not yet researched. From a theoretical perspective, vowel description in phonetics and phonology uses moslty articulatory features to categorise vowels. Here, it could be possible to incorporate perceptually driven aspects of vowel description to shift vowel descriptions towards a more speech perception-based description model. But before this question can be answered, there are other aspects of perceived loudness in speech sounds that need attention. Although we found robust differences between the different vowel qualities in our rating studies, the reasons for these differences remain unclear. There have been studies in the past that found an association between different vowel qualities and perceived loudness, i.e., lower vowels are accompanied by higher perceived loudness (Glave & Rietveld, 1975, 1979). Others found perceived loudness differences in speech sound perception related to vocal effort (Eriksson & Traunmüller, 1999, 2002), glottal excitation characteristics (Seshadri & Yegnanarayana, 2009), subglottal pressure (Ladefoged & McKinney, 1963) and spectral characteristics like bandwidth (Moore, 2003; Rennies et al., 2013). In our first rating study, there were some indications that perceived loudness differences were related to the degree of openness of the used vowels, i.e., higher degree of openness might have led to higher perceived loudness (Riedinger et al., 2021, p. 12). However, this aspect must be systematically investigated in the future to confirm or falsify the relationship between degree of openness and perceived loudness in vowels. Moreover, additional studies are required to investigate the association between vowel quality and perceived loudness. Furthermore, in our rating study we obtained only categorical decisions about the loudness of two consecutive presented stimuli. With our data, it was not possible to describe the magnitude of perceived loudness differences. It would be beneficial to obtain magnitude data, for example with magnitude estimations or loudness matching experiments (Marks & Florentine, 2011, pp. 32–45; Moore, 2003, pp. 127–128), to investigate whether differences in the magnitude of perceived loudness are also reflected in neural amplitude effects. Finally, the influence of perceived loudness on iMMN amplitude could only be calculated on an aggregated level in this dissertation project, since the neural and perceived loudness data came from different participants. In the future, it would be interesting to conduct an iMMN experiment using vowel stimuli as well as a perceived loudness rating with the same participants. The advantage with this setting would be two-fold: First, even in aggregated data the relationship could be described more accurately since it is known that especially the psychoacoustic measure of perceived loudness is to some extent subject-dependent (Litovsky, 2015; Marks & Florentine, 2011; Yanushevskaya et al., 2013). Additionally, when obtaining neural and perceived loudness data from the same participants, it would also be possible to investigate the correlation with greater depth on a subject level.

References

Aaltonen, O., Eerola, O., Lang, A. H., Uusipaikka, E., & Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *The Journal of the Acoustical Society of America*, *96*(3), 1489–1493. https://doi.org/10.1121/1.410291

Aaltonen, O., Tuomainen, J., Laine, M., & Niemi, P. (1993). Cortical Differences in Tonal versus Vowel Processing as Revealed by an ERP Component Called Mismatch Negativity (MMN). *Brain and Language*, *44*(2), 139–152.

Aichert, I., Marquardt, C., & Ziegler, W. (2005). Frequenzen sublexikalischer Einheiten des Deutschen: CELEX-basierte Datenbanken. *Neurolinguistik*, *19*(1–2), 55–81.

Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*, 534–546.

Aleksandrov, A. A., Boricheva, D. O., Pulvermüller, F., & Shtyrov, Y. (2011). Strength of Word-Specific Neural Memory Traces Assessed Electrophysiologically. *PLoS ONE*, *6*(8), e22999. https://doi.org/10.1371/journal.pone.0022999

Aleksandrov, A. A., Memetova, K. S., Stankevich, L. N., & Uplisova, K. O. (2017). Effects of Russian-Language Word Frequency on Mismatch Negativity in Auditory Event-Related Potentials. *Neuroscience and Behavioral Physiology*, *47*(9), 1043–1050. https://doi.org/10.1007/s11055-017-0510-3

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, *5*(2), 183–207. https://doi.org/10.1017/S0952675700002268

Baldeweg, T., Klugman, A., Gruzelier, J., & Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia Research*, *69*(2–3), 203–217. https://doi.org/10.1016/j.schres.2003.09.009

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonetics by computer* (6.0.18) [Computer software]. Phonetic Science.

Bohn, O.-S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *The Journal of the Acoustical Society of America*, *110*(1), 504–515. https://doi.org/10.1121/1.1380415

Bomba, M. D., Choly, D., & Pang, E. W. (2011). Phoneme discrimination and mismatch negativity in English and Japanese speakers: *NeuroReport*, *22*(10), 479–483. https://doi.org/10.1097/WNR.0b013e328347dada

Bonte, M., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*(12), 2765–2774. https://doi.org/10.1016/j.clinph.2005.08.012

Bouwer, F. L., & Honing, H. (2015). Temporal attending and prediction influence the perception of metrical rhythm: Evidence from reaction times and ERPs. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01094

Brunellière, A., Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2009). Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition*, *111*(3), 390–396. https://doi.org/10.1016/j.cognition.2009.02.013

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The Word Frequency Effect: A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Chládková, K., Urbanec, J., Skálová, S., & Kremláček, J. (2021). Newborns' neural processing of native vowels reveals directional asymmetries. *Developmental Cognitive Neuroscience*, *52*, 101023. https://doi.org/10.1016/j.dcn.2021.101023

Corcoran, C. M., Stoops, A., Lee, M., Martinez, A., Sehatpour, P., Dias, E. C., & Javitt, D. C. (2018). Developmental trajectory of mismatch negativity and visual event-related potentials in healthy controls: Implications for neurodevelopmental vs. neurodegenerative models of schizophrenia. *Schizophrenia Research*, *191*, 101–108. https://doi.org/10.1016/j.schres.2017.09.047

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, *1394*, 79–89. https://doi.org/10.1016/j.brainres.2011.04.001

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 757–772. https://doi.org/10.1037/a0030862

Cummings, A., Madden, J., & Hefta, K. (2017). Converging evidence for [coronal] underspecification in English-speaking adults. *Journal of Neurolinguistics*, *44*, 147–162. https://doi.org/10.1016/j.jneuroling.2017.05.003

Cummings, A., Ogiela, D. A., & Wu, Y. C. (2020). Evidence for [Coronal] Underspecification in Typical and Atypical Phonological Development. *Frontiers in Human Neuroscience*, *14*, 580697. https://doi.org/10.3389/fnhum.2020.580697

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*(2), 141–177. https://doi.org/10.1016/0010-0285(87)90010-7

Dalebout, S. D., & Fox, L. G. (2000). Identification of the Mismatch Negativity in the Responses of Individual Listeners. *Journal of the American Academy of Audiology*, *11*(1), 11.

De Boeck, P., & Jeon, M. (2019). An Overview of Models for Response Times and Processes in Cognitive Tests. *Frontiers in Psychology*, *10*, 102. https://doi.org/10.3389/fpsyg.2019.00102

de Jonge, M. J. I., & Boersma, P. (2015). French high-mid vowels are underspecified for heigh. *Proceedings of the 18th International Congress of Phonetic Sciences*, 5.

de Rue, N. P. W. D. (2014). *Directional asymmetries in vowel perception*. Radboud University.

de Rue, N. P. W. D., Snijders, T. M., & Fikkert, P. (2021). Contrast and Conflict in Dutch Vowels. *Frontiers in Human Neuroscience*, *15*, 629648. https://doi.org/10.3389/fnhum.2021.629648

Deacon, D., Gomes, H., Nousak, J. M., Ritter, W., & Javitt, D. (2000). Effect of frequency separation and stimulus rate on the mismatch negativity: An examination of the issue of refractoriness in humans. *Neuroscience Letters*, *287*(3), 167–170. https://doi.org/10.1016/S0304-3940(00)01175-7

Deguchi, C., Chobert, J., Brunellière, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. *Brain Research*, *1366*, 149–161. https://doi.org/10.1016/j.brainres.2010.09.104

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, *8*(4), 919–924.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, *9*(8), 1885–1888. https://doi.org/10.1097/00001756-199806010-00040

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience*, *12*(4), 635–647. https://doi.org/10.1162/089892900562390

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400.

Clinical Neurophysiology, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Emmendorfer, A. K., Bonte, M., Jansma, B. M., & Kotz, S. A. (2021). *Phonological and temporal regularities lead to differential ERP effects in self- and externally generated speech* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.05.04.442414

Emmendorfer, A. K., Correia, J. M., Jansma, B. M., Kotz, S. A., & Bonte, M. (2020). ERP mismatch response to phonological and temporal regularities in speech. *Scientific Reports*, *10*(1), 9917. https://doi.org/10.1038/s41598-020-66824-x

Endrass, T., Mohr, B., & Pulvermuller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, *19*(6), 1653–1660. https://doi.org/10.1111/j.1460-9568.2004.03247.x

Eriksson, A., & Traunmüller, H. (1999). PERCEPTION OF VOCAL EFFORT AND SPEAKER DISTANCE on the basis of vowel utterances. *Proceedings of the 14th ICPhS*, 2469–2472.

Eriksson, A., & Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception & Psychophysics*, *64*(1), 131–139. https://doi.org/10.3758/BF03194562

Eulitz, C., & Lahiri, A. (2004). Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583. https://doi.org/10.1162/089892904323057308

Eulitz, C., Lahiri, A., Obleser, J., & Reetz, H. (2003). *Brain Electric Activity Reflects the Underspecification of Phonological Features in the Mental Lexicon*. 1631–1633.

Fitzgerald, K., & Todd, J. (2020). Making Sense of Mismatch Negativity. *Frontiers in Psychiatry*, *11*, 468. https://doi.org/10.3389/fpsyt.2020.00468

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, *95*(6), 3623–3641. https://doi.org/10.1121/1.409931

Fletcher, H., & Munson, W. A. (1933). Loudness, Its Definition, Measurement and Calculation. *The Journal of the Acoustical Society of America*, *5*(82), 82–108. https://doi.org/10.1121/1.1915637

Florentine, M., Popper, A. N., & Fay, R. R. (Eds.). (2011). *Loudness* (Vol. 37). Springer New York. https://doi.org/10.1007/978-1-4419-6712-1

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving Nonnative Vowels: The Effect of Context on Perception as Evidenced by Event-Related Brain Potentials. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1496–1510. https://doi.org/10.1044/1092-4388(2005/104)

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1545–1559. https://doi.org/10.1037/a0012481

Fu, Z., & Monahan, P. J. (2021). Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants. *Frontiers in Human Neuroscience*, *15*, 609898. https://doi.org/10.3389/fnhum.2021.609898

Gallucci, M. (2019). *GAMLj: General analyses for linear models. [Jamovi module].* [Computer software]. https://gamlj.github.io/

García, P. B., & Froud, K. (2018). Perception of American English vowels by sequential Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *21*(1), 80–103. https://doi.org/10.1017/S1366728916000808

Garrapa, L., Bottari, D., Grimaldi, M., Pavani, F., Calabrese, A., Benedetto, M. D., & Vitale, S. (2013). *Processing of /i/ and /u/ in Italian Cochlear-Implant Children: A Behavioral and Neurophysiologic Study*. 2272–2276.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Glave, R. D., & Rietveld, A. C. M. (1975). Is the effort dependence of speech loudness explicable on the basis of acoustical cues? *The Journal of the Acoustical Society of America*, *58*(4), 875–879. https://doi.org/10.1121/1.380737

Glave, R. D., & Rietveld, A. C. M. (1979). Bimodal cues for speech loudness. *The Journal of the Acoustical Society of America*, *66*(4), 1018–1022. https://doi.org/10.1121/1.383320

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*, 14. https://doi.org/10.3389/fnhum.2014.00279

Hamamura, M., & Iwamiya, S. (2016). Relationship of the differences in perceived loudness of sound and the optimum listening level between males and females. *Acoustical Science and Technology*, *37*(1), 40–41. https://doi.org/10.1250/ast.37.40

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135–178. https://doi.org/10.1016/j.cogpsych.2016.05.001

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70. https://doi.org/10.1017/S0022226705003683

Henrich, K., & Scharinger, M. (2022). Predictive Processing in Poetic Language: Event-Related Potentials Data on Rhythmic Omissions in Metered Speech. *Frontiers in Psychology*, *12*, 782765. https://doi.org/10.3389/fpsyg.2021.782765

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, *109*(49), 20095–20100. https://doi.org/10.1073/pnas.1213390109

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007

Hestvik, A., Shinohara, Y., Durvasula, K., Verdonschot, R. G., & Sakai, H. (2020). Abstractness of human speech sound representations. *Brain Research*, *1732*, 146664. https://doi.org/10.1016/j.brainres.2020.146664

Højlund, A., Gebauer, L., McGregor, W. B., & Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: An MEG study. *Language, Cognition and Neuroscience*, *34*(5), 545–560. https://doi.org/10.1080/23273798.2019.1572204

Hsu, C.-H., Evans, J. P., & Lee, C.-Y. (2015). Brain responses to spoken F 0 changes: Is H special? *Journal of Phonetics*, *51*, 82–92. https://doi.org/10.1016/j.wocn.2015.02.003

Hume, E. (2004). Deconstructing Markedness: A Predictability-Based Approach. *Annual Meeting of the Berkeley Linguistics Society*, *30*(1), 182–198.

Hume, E. (2008). Markedness and the Language User. *Phonological Studies*, *11*.

Hume, E. (2006). *Language-Specific and Universal Markedness: An information-theoretic apporach* [Vortrag]. Colloquium on Information Theory & Phonology, Linguistic Society of America Annual Meeting.

Hwang, S.-O. K., Monahan, P. J., & Idsardi, W. J. (2010). Underspecification and asymmetries in voicing perception. *Phonology*, *27*(2), 205–224. https://doi.org/10.1017/S0952675710000102

Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., & Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters*, *321*(3), 133–136. https://doi.org/10.1016/S0304-3940(01)02408-9

Ilvonen, T., Kujala, T., Kozou, H., Kiesiläinen, A., Salonen, O., Alku, P., & Näätänen, R. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neuroscience Letters*, *366*(3), 235–240. https://doi.org/10.1016/j.neulet.2004.05.024

Jacobsen, T., Horenkamp, T., & Schröger, E. (2003). Preattentive Memory-Based Comparison of Sound Intensity. *Audiology and Neurotology*, *8*(6), 338–346. https://doi.org/10.1159/000073518

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., & Winkler, I. (2004). Pre-attentive auditory processing of lexicality. *Brain and Language*, *88*(1), 54–67. https://doi.org/10.1016/S0093-934X(03)00156-1

Jacobsen, T., Schroger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, *41*(4), 654–659. https://doi.org/10.1111/1469-8986.2004.00175.x

Johnson, K. (2015). Vowel perception asymmetry in auditory and phonemic listening. *UC Berkeley PhonLab Annual Report*, *11*(11), 26.

Karypidis, C. (2007). *Order effects and vowel decay in short-term memory: The neutralization hypothesis*. 657–660. https://halshs.archives-ouvertes.fr/halshs-00167116

Kazanina, N., Bowers, J. S., & Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*, *25*(2), 560–585. https://doi.org/10.3758/s13423-017-1362-0

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schroger, E., & Jacobsen, T. (2007). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology*, *67*, 131–143. https://doi.org/10.1016/j.ijpsycho.2007.10.012

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and Late Mismatch Negativity Elicited by Words and Speech-Like Stimuli in Children. *Brain and Language*, *76*(3), 332–339. https://doi.org/10.1006/brln.2000.2426

Kotzor, S., Wetterlin, A., & Lahiri, A. (2017). Symmetry or asymmetry: Evidence for underspecification in the mental lexicon. In A. Lahiri & S. Kotzor (Eds.), *The Speech Processing Lexicon* (pp. 85–106). De Gruyter. https://doi.org/10.1515/9783110422658-005

Kotzor, S., Zhou, B., & Lahiri, A. (2020). (A)Symmetry in vowel features in verbs and pseudoverbs: ERP evidence. *Neuropsychologia*, *143*, 107474. https://doi.org/10.1016/j.neuropsychologia.2020.107474

Kriengwatana, B. P., & Escudero, P. (2017). Directional Asymmetries in Vowel Perception of Adult Nonnative Listeners Do Not Change Over Time With Language Experience. *Journal of Speech, Language, and Hearing Research*, *60*(4), 1088–1093. https://doi.org/10.1044/2016_JSLHR-H-16-0050

Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, *112*(9), 1712–1719. https://doi.org/10.1016/S1388-2457(01)00625-3

Ladefoged, P., & McKinney, N. P. (1963). Loudness, Sound Pressure, and Subglottal Pressure in Speech. *The Journal of the Acoustical Society of America*, *35*(4), 454–460. https://doi.org/10.1121/1.1918503

Lahiri, A. (2018). Predicting universal phonological contrasts. In L. M. Hyman & F. Plank (Eds.), *Phonological Typology* (pp. 229–272). De Gruyter. https://doi.org/10.1515/9783110451931-007

Lahiri, A., & Reetz, H. (2002a). Underspecified Recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (p. 40). Mouton de Gruyter.

Lahiri, A., & Reetz, H. (2002b). Underspecified Recognition. In *Laboratory Phonology 7* (pp. 637–675). Mouton de Gruyter.

Lahiri, A., & Reetz, H. (2010a). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Lahiri, A., & Reetz, H. (2010b). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Langers, D. R. M., van Dijk, P., Schoenmaker, E. S., & Backes, W. H. (2007). fMRI activation in relation to sound intensity and loudness. *NeuroImage*, *35*(2), 709–718. https://doi.org/10.1016/j.neuroimage.2006.12.013

Law, S., Fung, R., & Kung, C. (2013). Asymmetric Mismatch Negativity and Underlying Representations of Lexical Tones in Cantonese. *Procedia - Social and Behavioral Sciences*, *94*, 82–83. https://doi.org/10.1016/j.sbspro.2013.09.039

Lawyer, L. A., & Corina, D. P. (2018). Putting underspecification in context: ERP evidence for sparse representations in morphophonological alternations. *Language, Cognition and Neuroscience*, *33*(1), 50–64. https://doi.org/10.1080/23273798.2017.1359635

Li, X., & Chen, Y. (2015). Representation and Processing of Lexical Tone and Tonal Variants: Evidence from the Mismatch Negativity. *PLOS ONE*, *10*(12), e0143097. https://doi.org/10.1371/journal.pone.0143097

Lidji, P., Jolicœur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology*, *121*(4), 533–541. https://doi.org/10.1016/j.clinph.2009.12.018

Lindblom, B., & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics*, *17*(1–2), 107–121. https://doi.org/10.1016/S0095-4470(19)31516-5

Lipski, S. C., Lahiri, A., & Eulitz, C. (2007). *Differential hight specification in front vowels for german speakers and turkish-german bilinguals: An electroencephalographic study*. 809–812.

Litovsky, R. (2015). Development of the auditory system. In *Handbook of Clinical Neurology* (Vol. 129, pp. 55–72). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00003-2

Liu, L., Ong, J. H., Tuninetti, A., & Escudero, P. (2018). One Way or Another: Evidence for Perceptual Asymmetry in Pre-attentive Learning of Non-native Contrasts. *Frontiers in Psychology*, *9*, 162. https://doi.org/10.3389/fpsyg.2018.00162

Liu, Y., Polka, L., Masapollo, M., & Ménard, L. (2021). Disentangling the roles of formant proximity and stimulus prototypicality in adult vowel perception. *JASA Express Letters*, *1*, 9. https://doi.org/10.1121/10.0003041

Llompart, M. (2021). Lexical and Phonetic Influences on the Phonolexical Encoding of Difficult Second-Language Contrasts: Insights From Nonword Rejection. *Frontiers in Psychology*, *12*, 659852. https://doi.org/10.3389/fpsyg.2021.659852

Mager, R., Falkenstein, M., Störmer, R., Brand, S., Müller-Spahn, F., & Bullinger, A. H. (2005). Auditory distraction in young and middle-aged adults: A behavioural and event-related potential study. *Journal of Neural Transmission*, *112*(9), 1165–1176. https://doi.org/10.1007/s00702-004-0258-0

Marks, L. E., & Florentine, M. (2011). Measurement of Loudness, Part I: Methods, Problems, and Pitfalls. In M. Florentine, A. N. Popper, & R. R. Fay (Eds.), *Loudness* (Vol. 37, pp. 17–56). Springer New York. https://doi.org/10.1007/978-1-4419-6712-1_2

Masapollo, M., Polka, L., & Ménard, L. (2015). Asymmetries in vowel perception: Effects of formant convergence and category "goodness". *The Journal of the Acoustical Society of America*, *137*(4), 2385–2385. https://doi.org/10.1121/1.4920678

Masapollo, M., Polka, L., & Ménard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition*, *166*, 358–370. https://doi.org/10.1016/j.cognition.2017.06.001

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., & Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 1103–1118. https://doi.org/10.1037/xhp0000518

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America*, *141*(4), 2857–2869. https://doi.org/10.1121/1.4981006

Masapollo, M., Zhao, T. C., Franklin, L., & Morgan, J. L. (2019). Asymmetric discrimination of nonspeech tonal analogues of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(2), 285–300. https://doi.org/10.1037/xhp0000603

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Bahavior Research Methods*, *44*(2), 314–324. http://dx.doi.org/10.3758/s13428-011-0168-7

Meng, Y., Kotzor, S., Xu, C., Wynne, H., & Lahiri, A. (2021a). Asymmetric Influence of Vocalic Context on Mandarin Sibilants: Evidence From ERP Studies. *Frontiers in Human Neuroscience*, *15*, 617318. https://doi.org/10.3389/fnhum.2021.617318

Meng, Y., Kotzor, S., Xu, C., Wynne, H. S. Z., & Lahiri, A. (2021b). Mismatch negativity (MMN) as an index of asymmetric processing of consonant duration in fake Mandarin geminates. *Neuropsychologia*, *163*, 108063. https://doi.org/10.1016/j.neuropsychologia.2021.108063

Mohanan, K. P. (1991). On the bases of radical underspecification. *Natural Language and Linguistic Theory*, *9*(2), 285–325. https://doi.org/10.1007/BF00134678

Monahan, P. J. (2018). Phonological Knowledge and Speech Comprehension. *Annual Review of Linguistics*, *4*(1), 21–47. https://doi.org/10.1146/annurev-linguistics-011817-045537

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*. Academic Press.

Moore, B. C. J. (2005). Loudness, Pitch and Timbre. In B. E. Goldstein (Ed.), *Blackwell Handbook of Sensation & Perception* (pp. 408–436). Blackwell Publishing.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21. https://doi.org/10.1111/1469-8986.3810001

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*(1–3), 347–352. https://doi.org/10.1016/0304-3940(89)90844-6

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

Nam, Y., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition*, *155*, 57–66. https://doi.org/10.1016/j.cognition.2016.06.005

Novak, G. P., Ritter, W., Vaughan, H. G., & Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory. Discrimination task. *Electroeneephalography and Clinical Neurophvsiology*, *75*, 255–275.

Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch Detection and the Latency of Temporal Judgments. *Psychophysiology*, *29*(4), 398–411. https://doi.org/10.1111/j.1469-8986.1992.tb01713.x

Novitski, N., Tervaniemi, M., Huotilainen, M., & Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research*, *20*(1), 26–36. https://doi.org/10.1016/j.cogbrainres.2003.12.011

Nudga, N., Urbanec, J., Oceláková, Z., Kremláček, J., & Chládková, K. (2021). Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults. *Frontiers in Human Neuroscience*, *15*, 643655. https://doi.org/10.3389/fnhum.2021.643655

Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Vowels. *Journal of Cognitive Neuroscience*, *16*(1), 31–39. https://doi.org/10.1162/089892904322755539

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 466–479. https://doi.org/10.1016/0013-4694(91)90064-B

Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., Näätänen, R., & Huotilainen, M. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *International Journal of Psychophysiology*, *87*(1), 103–110. https://doi.org/10.1016/j.ijpsycho.2012.11.010

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peltola, M. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*, 25–28. https://doi.org/10.1016/S0304-3940(03)00997-2

Peltola, M. S. (2007). *Speech sound perception and neural representations*. 123–126.

Peltola, M. S., & Aaltonen, O. (2005). Long-Term Memory Trace Activation for Vowels Depends on the Mother Tongue and the Linguistic Context. *Journal of Psychophysiology*, *19*(3), 159–164. https://doi.org/10.1027/0269-8803.19.3.159

Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, *482*(1), 71–75. https://doi.org/10.1016/j.neulet.2010.07.010

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., Sockalingam, R., & Chenery, H. J. (2004). Automatic Auditory Processing of English Words as Indexed by the Mismatch Negativity, Using a Multiple Deviant Paradigm: *Ear and Hearing*, *25*(3), 284–301. https://doi.org/10.1097/01.AUD.0000130800.88987.03

Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., Finnigan, S., & Alku, P. (2004). Processing of English Words with Fine Acoustic Contrasts and Simple Tones: A Mismatch Negativity Study. *Journal of the American Academy of Audiology*, *15*(1), 47–66. https://doi.org/10.3766/jaaa.15.1.6

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study. *Journal of Cognitive Neuroscience*, *12*(6), 1038–1055. https://doi.org/10.1162/08989290051137567

Pisoni, D. B. (1980). Variability of Vowel Formant Frequencies and the Quantal Theory of Speech: A First Report. *Phonetica*, *37*(5–6), 285–305. https://doi.org/10.1159/000259999

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290. https://doi.org/10.3758/BF03213946

Poeppel, D., & Monahan, P. J. (2008). Speech Perception: Cognitive Foundations and Cortical Implementation. *Current Directions in Psychological Science*, *17*(2), 80–85. https://doi.org/10.1111/j.1467-8721.2008.00553.x

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., Eischen, S. E., & Collins, G. E. (1994). P300 from a single auditory stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 253–261. https://doi.org/10.1016/0168-5597(94)90068-X

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the Perception of Mandarin Tones: Evidence From Mismatch Negativity. *Journal of Experimental Psychology*, *42*(10), 1547–1570.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961–2977. https://doi.org/10.1121/1.400734

Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, *100*(1), 577–592. https://doi.org/10.1121/1.415884

Polka, L., & Bohn, O.-S. (2003a). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2003b). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2011a). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., & Bohn, O.-S. (2011b). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., Molnar, M., Zhao, T. C., & Masapollo, M. (2021). Neurophysiological Correlates of Asymmetries in Vowel Perception: An English-French Cross-Linguistic Event-Related Potential Study. *Frontiers in Human Neuroscience*, *15*, 607148. https://doi.org/10.3389/fnhum.2021.607148

Polka, L., Ruan, Y., & Masapollo, M. (2019). Understanding Vowel Perception Biases – It's Time to Take a Meta-analytic Approach. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 561–582). Dept. of English, School of Communication & Culture.

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The Interplay Between Input and Initial Biases: Asymmetries in Vowel Perception During the First Year of Life: Interplay Between Input and Initial Biases. *Child Development*, *83*(3), 965–976. https://doi.org/10.1111/j.1467-8624.2012.01740.x

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory Traces for Words as Revealed by the Mismatch Negativity. *NeuroImage*, *14*(3), 607–616. https://doi.org/10.1006/nimg.2001.0864

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R. J., & Marslen-Wilson, W. D. (2006). Tracking speech comprehension in space and time. *NeuroImage*, *31*(3), 1297–1305. https://doi.org/10.1016/j.neuroimage.2006.01.030

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, *41*(1), 106–112. https://doi.org/10.1111/j.1469-8986.2003.00135.x

Ratcliff, R. (1993). Methods for Dealing With Reaction Times Outliers. *Psychological Bulletin*, *114*(3), 510–532.

Remijn, G. B., & Kojima, H. (2010). Active versus passive listening to auditory streaming stimuli: A near-infrared spectroscopy study. *Journal of Biomedical Optics*, *15*(3), 037006. https://doi.org/10.1117/1.3431104

Rennies, J., Holube, I., & Verhey, J. L. (2013). Loudness of Speech and Speech-Like Signals. *Acta Acustica United with Acustica*, *99*(2), 268–282. https://doi.org/10.3813/AAA.918609

Rice, K. (2007). Markedness in phonology. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 79–97). University Press.

Riedinger, M., Nagels, A., Werth, A., & Scharinger, M. (2021). Asymmetries in Accessing Vowel Representations Are Driven by Phonological and Acoustic Properties: Neural and Behavioral Evidence From Natural German Minimal Pairs. *Frontiers in Human Neuroscience*, *15*, 20. https://doi.org/10.3389/fnhum.2021.612345

Rinker, T., Alku, P., Brosch, S., & Kiefer, M. (2010). Discrimination of native and non-native vowel contrasts in bilingual Turkish–German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain and Language*, *113*(2), 90–95. https://doi.org/10.1016/j.bandl.2010.01.007

Ritter, W., De Sanctis, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology*, *43*(5), 423–430. https://doi.org/10.1111/j.1469-8986.2006.00423.x

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, *111*(1–2), 13–23. https://doi.org/10.1016/S0166-4328(00)00139-X

RStudio Team. (2022). *RStudio: Integrated Development Environment for R.* (2022.2.0.443) [Computer software]. PBC. http://www.rstudio.com

Sams, M., Paavilainen, P., & Alho, K. (1985). Auditory Freqeuncy Discrimination and Event-related Potentials. *Electroencephalography and Clinical Neurophysiology*, 437–448.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Cognitive Brain Research*, *16*(2), 250–256. https://doi.org/10.1016/S0926-6410(02)00280-X

Scharinger, M. (2006). Featural representations of the New Zealand English short front vowels. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (p. 5). Australian Speech Science & Technology Association Inc.

Scharinger, M. (2009). Minimal representations of alternating vowels. *Lingua*, *119*(10), 1414–1425. https://doi.org/10.1016/j.lingua.2007.12.009

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., & Obleser, J. (2012). A Sparse Neural Code for Some Speech Sounds but Not for Others. *PLoS ONE*, *7*(7), 1-11 (e40953). https://doi.org/10.1371/journal.pone.0040953

Scharinger, M., Domahs, U., Klein, E., & Domahs, F. (2016). Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus. *Brain and Language*, *163*, 42–49. https://doi.org/10.1016/j.bandl.2016.09.002

Scharinger, M., & Idsardi, W. J. (2014). Sparseness of vowel category structure: Evidence from English dialect comparison. *Lingua*, *140*, 35–51. https://doi.org/10.1016/j.lingua.2013.11.007

Scharinger, M., Idsardi, W. J., & Poe, S. (2011a). A Comprehensive Three-dimensional Cortical Map of Vowel Space. *Journal of Cognitive Neuroscience*, *23*(12), 3972–3982. https://doi.org/10.1162/jocn_a_00056

Scharinger, M., Idsardi, W. J., & Poe, S. (2011b). Neuromagnetic reflections of harmony and constraint violations in Turkish. *Laboratory Phonology*, *2*(1), 99–123. https://doi.org/10.1515/labphon.2011.003

Scharinger, M., & Lahiri, A. (2010). Height Differences in English Dialects: Consequences for Processing and Representation. *Language and Speech*, *53*(2), 245–272. https://doi.org/10.1177/0023830909357154

Scharinger, M., Lahiri, A., & Eulitz, C. (2010). Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics*, *23*(4), 383–399. https://doi.org/10.1016/j.jneuroling.2010.02.005

Scharinger, M., Merickel, J., Riley, J., & Idsardi, W. J. (2011). Neuromagnetic evidence for a featural distinction of English consonants: Sensor- and source-space data. *Brain and Language*, *116*(2), 71–82. https://doi.org/10.1016/j.bandl.2010.11.002

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the Processing of Vowel Height. *Journal of Speech, Language, and Hearing Research*, *55*(3), 903–918. https://doi.org/10.1044/1092-4388(2011/11-0065)

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *NeuroImage*, *128*, 293–301. https://doi.org/10.1016/j.neuroimage.2016.01.003

Scharinger, M., Steinberg, J., & Tavano, A. (2017). Integrating speech in time depends on temporal expectancies and attention. *Cortex*, *93*, 28–40. https://doi.org/10.1016/j.cortex.2017.05.001

Schild, U., Röder, B., & Friedrich, C. K. (2012). Neuronal spoken word recognition: The time course of processing variation in the speech signal. *Language and Cognitive Processes*, *27*(2), 159–183. https://doi.org/10.1080/01690965.2010.503532

Schluter, K., Politzer-Ahles, S., Al Kaabi, M., & Almeida, D. (2017). Laryngeal Features Are Phonetically Abstract: Mismatch Negativity Evidence from Arabic, English, and Russian. *Frontiers in Psychology*, *8*, 746. https://doi.org/10.3389/fpsyg.2017.00746

Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, *32*(1), 55–65. https://doi.org/10.1111/j.1469-8986.1995.tb03406.x

Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology*, *100*, 517–526.

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication*, *45*(4), 425–434. https://doi.org/10.1016/j.specom.2004.12.001

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997a). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286. https://doi.org/10.1006/jpho.1997.0043

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997b). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286. https://doi.org/10.1006/jpho.1997.0043

Sendlmeier, W. F., & Seebode, J. (2006). *Formantkarten des deutschen Vokalsystems*. 4.

Seshadri, G., & Yegnanarayana, B. (2009). Perceived loudness of speech based on the characteristics of glottal excitation source. *The Journal of the Acoustical Society of America*, *126*(4), 2061. https://doi.org/10.1121/1.3203668

Shtyrov, Y., Kimppa, L., Pulvermüller, F., & Kujala, T. (2011). Event-related potentials reflecting the frequency of unattended spoken words: A neuronal index of connection strength in lexical memory circuits? *NeuroImage*, *55*(2), 658–668. https://doi.org/10.1016/j.neuroimage.2010.12.002

Shtyrov, Y., Osswald, K., & Pulvermuller, F. (2008). Memory Traces for Spoken Words in the Brain as Revealed by the Hemodynamic Correlate of the Mismatch Negativity. *Cerebral Cortex*, *18*(1), 29–37. https://doi.org/10.1093/cercor/bhm028

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport*, *13*(4), 521–525. https://doi.org/10.1097/00001756-200203250-00033

Shtyrov, Y., & Pulvermüller, F. (2007). Language in the Mismatch Negativity Design: Motivations, Benefits, and Prospects. *Journal of Psychophysiology*, *21*(3–4), 176–187. https://doi.org/10.1027/0269-8803.21.34.176

Simon, E., Sjerps, M. J., & Fikkert, P. (2014). Phonological representations in children's native and non-native lexicon. *Bilingualism: Language and Cognition*, *17*(1), 3–21. https://doi.org/10.1017/S1366728912000764

Sittiprapaporn, W., Tervaniemi, M., Chindaduangratn, C., & Kotchabhakdi, N. (2005). Preattentive discrimination of across-category and within-category change in consonant???vowel syllable: *NeuroReport*, *16*(13), 1513–1518. https://doi.org/10.1097/01.wnr.0000175618.46677.07

Steriade, D. (1995). Underspecification and Markedness. In J. A. Goldsmith (Ed.), *The Handbook of phonological theory* (pp. 114–174). Blackwell Publishing.

Strauß, A., Kotz, S. A., Scharinger, M., & Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *NeuroImage*, *97*, 387–395. https://doi.org/10.1016/j.neuroimage.2014.04.005

Sundara, M., & Polka, L. (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition*, *106*(1), 234–258. https://doi.org/10.1016/j.cognition.2007.01.011

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*(1), 232–242. https://doi.org/10.1016/j.cognition.2007.12.013

Szymanski, M. D., Yund, E. W., & Woods, D. L. (1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *The Journal of the Acoustical Society of America*, *106*(6), 3492–3505. https://doi.org/10.1121/1.428202

Takegata, R., & Morotomi, T. (1999). Integrated neural representation of sound and temporal features in human auditory sensory memory: An event-related potential study. *Neuroscience Letters*, *274*(3), 207–210. https://doi.org/10.1016/S0304-3940(99)00711-9

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, *266*(2), 109–112. https://doi.org/10.1016/S0304-3940(99)00267-0

Takegata, R., Paavilainen, P., Naatanen, R., & Winkler, I. (2001). Preattentive processing of spectral, temporal, and structural characteristics of acoustic regularities: A mismatch negativity study. *Psychophysiology*, *38*(1), 92–98. https://doi.org/10.1111/1469-8986.3810092

Takegata, R., Syssoeva, O., Winkler, I., Paavilainen, P., & Näätänen, R. (2001). Common neural mechanism for processing onset-to-onset intervals and silent gaps in sound sequences: *Neuroreport*, *12*(8), 1783–1787. https://doi.org/10.1097/00001756-200106130-00053

Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of Place of Articulation Changes on Auditory Neural Activity: A Magnetoencephalography Study. *PLoS ONE*, *4*(2), e4452. https://doi.org/10.1371/journal.pone.0004452

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*(2–3), 157–167. https://doi.org/10.1016/0301-0511(94)90036-1

Tervaniemi, M., Schröger, E., Saher, M., & Näätänen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans – a mismatch negativity study. *Neuroscience Letters*, *290*(1), 66–70. https://doi.org/10.1016/S0304-3940(00)01290-8

The jamovi project. (2021). *Jamovi* (2.2.5.0) [Computer software]. https://www.jamovi.org.

Tiitinen, H., May, P., & Reinikainen, K. (1994). *Attentive novelty detection in humans is governed by pre-aHentive sensory memory*. *372*.

Todd, J., & Michie, P. T. (2000). Do perceived loudness cues contribute to duration mismatch negativity (MMN)?: *NeuroReport*, *11*(17), 3771–3774. https://doi.org/10.1097/00001756-200011270-00035

Tsuji, S., & Cristia, A. (2017). Which Acoustic and Phonological Factors Shape Infants' Vowel Discrimination? Exploiting Natural Variation in InPhonDB. *Interspeech 2017*, 2108–2112. https://doi.org/10.21437/Interspeech.2017-1468

Tuomainen, J., Savela, J., Obleser, J., & Aaltonen, O. (2013). Attention modulates the use of spectral attributes in vowel discrimination: Behavioral and event-related potential evidence. *Brain Research*, *1490*, 170–183. https://doi.org/10.1016/j.brainres.2012.10.067

Tyler, M. D., Nil, M. D., Nil, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica*, *71*(1), 4–21. https://doi.org/10.1159/000356237

Vannest, J. J., Karunanayaka, P. R., Altaye, M., Schmithorst, V. J., Plante, E. M., Eaton, K. J., Rasmussen, J. M., & Holland, S. K. (2009). Comparison of fMRI data from passive listening and active-response story processing tasks in children. *Journal of Magnetic Resonance Imaging*, *29*(4), 971–976. https://doi.org/10.1002/jmri.21694

Verá-Constán, F., & Sebastián-Gallés, N. (2008). *Asimetrías en la percepión de vocales: Se mantienen las anclas perspectivas?* Paper presented at the fifth SEPEX, San Sebastian, Spain.

Wang, X.-D., Gu, F., He, K., Chen, L.-H., & Chen, L. (2012). Preattentive Extraction of Abstract Auditory Rules in Speech Sound Stream: A Mismatch Negativity Study Using Lexical Tones. *PLoS ONE*, *7*(1), e30027. https://doi.org/10.1371/journal.pone.0030027

Wayland, R., Chen, S., Zhou, F., & Hong, Y. (2019). *Directional asymmetry in lexical tone perception*. 060005. https://doi.org/10.1121/2.0001300

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, 475–482.

Wiese, R. (2000). *The Phonology of German*. Oxford University Press.

Wolff, C., & Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters*, *311*(1), 37–40. https://doi.org/10.1016/S0304-3940(01)02135-8

Wottawa, J., Adda-Decker, M., & Isel, F. (2022). Neurophysiology of non-native sound discrimination: Evidence from German vowels and consonants in successive French–German bilinguals using an MMN oddball paradigm. *Bilingualism: Language and Cognition*, *25*(1), 137–147. https://doi.org/10.1017/S1366728921000468

Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2013). Voice quality in affect cueing: Does loudness matter? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00335

Yasin, I. (2007). Hemispheric differences in processing dichotic meaningful and non-meaningful words. *Neuropsychologia*, *45*(12), 2718–2729. https://doi.org/10.1016/j.neuropsychologia.2007.04.009

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, *1072*(1), 175–185. https://doi.org/10.1016/j.brainres.2005.12.004

Yu, Y. H., & Shafer, V. L. (2021). Neural Representation of the English Vowel Feature [High]: Evidence From /ɛ/ vs. /ɪ/. *Frontiers in Human Neuroscience*, *15*, 629517. https://doi.org/10.3389/fnhum.2021.629517

Zhang, H., Eppes, A., & Diaz, M. T. (2019). Task difficulty modulates age-related differences in the behavioral and neural bases of language production. *Neuropsychologia*, *124*, 254–273. https://doi.org/10.1016/j.neuropsychologia.2018.11.017

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., & Kuhl, P. K. (2019). Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. *Brain and Language*, *194*, 77–83. https://doi.org/10.1016/j.bandl.2019.05.002