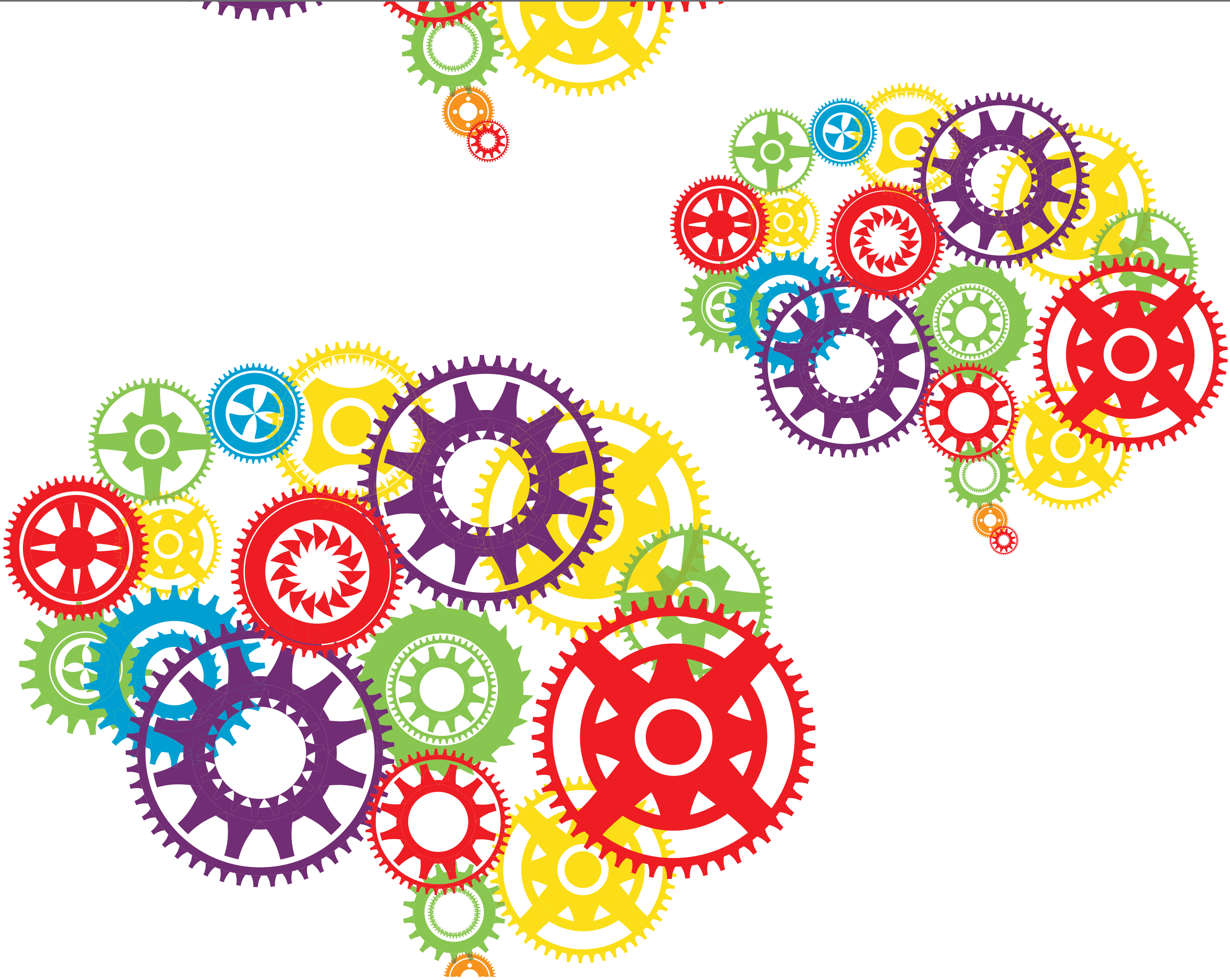
PHONOLOGICAL REPRESENTATIONS AND MISMATCH NEGATIVITY ASYMMETRIES

EDITED BY: Arild Hestvik, Valerie L. Shafer, Aditi Lahiri and Mathias Scharinger

PUBLISHED IN: Frontiers in Human Neuroscience



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-733-7
DOI 10.3389/978-2-88974-733-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PHONOLOGICAL REPRESENTATIONS AND MISMATCH NEGATIVITY ASYMMETRIES

Topic Editors:
Arild Hestvik, University of Delaware, United States
Valerie L. Shafer, The City University of New York, United States
Aditi Lahiri, University of Oxford
Mathias Scharinger, University of Marburg, Germany

Citation: Hestvik, A., Shafer, V. L., Lahiri, A., Scharinger, M., eds. (2022). Phonological Representations and Mismatch Negativity Asymmetries. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-733-7

Table of Contents

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 22 December 2020
doi: 10.3389/fnhum.2020.580697

Evidence for [Coronal] Underspecification in Typical and Atypical Phonological Development

Alycia E. Cummings[1]*, *Diane A. Ogiela*[1] *and Ying C. Wu*[2]

[1] *Department of Communication Sciences and Disorders, Idaho State University, Meridian, ID, United States,* [2] *Swartz Center for Computational Neuroscience, University of California, San Diego, San Diego, CA, United States*

OPEN ACCESS

Edited by:
Mathias Scharinger,
University of Marburg, Germany

Reviewed by:
Norbert Maïonchi-Pino,
UMR6024 Laboratoire de Psychologie Sociale et Cognitive (LAPSCO), France
Yang Zhang,
University of Minnesota Health Twin Cities, United States

***Correspondence:**
Alycia E. Cummings
cummalyc@isu.edu

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *06 July 2020*
Accepted: *30 November 2020*
Published: *22 December 2020*

Citation:
Cummings AE, Ogiela DA and Wu YC (2020) Evidence for [Coronal] Underspecification in Typical and Atypical Phonological Development. Front. Hum. Neurosci. 14:580697. doi: 10.3389/fnhum.2020.580697

The Featurally Underspecified Lexicon (FUL) theory predicts that [coronal] is the language universal default place of articulation for phonemes. This assumption has been consistently supported with adult behavioral and event-related potential (ERP) data; however, this underspecification claim has not been tested in developmental populations. The purpose of this study was to determine whether children demonstrate [coronal] underspecification patterns similar to those of adults. Two English consonants differing in place of articulation, [labial] /b/ and [coronal] /d/, were presented to 24 children (ages 4–6 years) characterized by either a typically developing phonological system (TD) or a phonological disorder (PD). Two syllables, /bɑ/ and /dɑ/, were presented in an ERP oddball paradigm where both syllables served as the standard and deviant stimulus in opposite stimulus sets. Underspecification was examined with three analyses: traditional mean amplitude measurements, cluster-based permutation tests, and single-trial general linear model (GLM) analyses of single-subject data. Contrary to previous adult findings, children with PD demonstrated a large positive mismatch response (PMR) to /bɑ/ while the children with TD exhibited a negative mismatch response (MMN); significant group differences were not observed in the /dɑ/ responses. Moreover, the /bɑ/ deviant ERP response was significantly larger in the TD children than in the children with PD. At the single-subject level, more children demonstrated mismatch responses to /dɑ/ than to /bɑ/, though some children had a /bɑ/ mismatch response and no /dɑ/ mismatch response. While both groups of children demonstrated similar responses to the underspecified /dɑ/, their neural responses to the more specified /bɑ/ varied. These findings are interpreted within a proposed developmental model of phonological underspecification, wherein children with PD are functioning at a developmentally less mature stage of phonological acquisition than their same-aged TD peers. Thus, phonological underspecification is a phenomenon that likely develops over time with experience and exposure to language.

Keywords: ERP, underspecification, MMN, phonology, children, phonological disorder

INTRODUCTION

Accurate speech perception is a complex process (Aslin and Smith, 1988). For example, auditory sensory information must first be detected, and then transformed into a neural representation of the event, with meaning eventually attributed to the auditory input. More specifically, phonological representations are formed by decoding of the speech signal, which requires, in part, the extraction and sequencing of phonetic features from the auditory signal (Scott and Wise, 2004). Being able to accurately perceive speech sounds allows for the accurate formation of phonemic categories, and more importantly, the accurate identification of words. Thus, it is important to form detailed phonological representations so that accurate speech production can occur.

Children appear to be born with the capability of differentiating [nearly] all the sounds of human speech (Eimas et al., 1971; Eimas, 1975). However, within the first year of life, infants' phonetic sensitivity decreases as they attend to the statistical distributions of sounds in the input (Kuhl, 2000, 2010). For example, infants can discriminate native vowel phonemes by 6 months of age and native consonants by 10 months of age (Werker and Tees, 1984; Kuhl et al., 1992; Werker and Hensch, 2015). Thus, while young children have the ability perceive subtle differences in sounds, it is only with time and language experience that they assign phonological meaning to the sounds. This suggests that over time, children learn which features are necessary for phonemic categorization in their native language(s) (Cheour et al., 1998; Kuhl et al., 2008).

Together, these findings suggest that auditory speech perception changes during phonological development (Nittrouer and Miller, 1997). Initially, sounds can be perceived and differentiated, but are not necessarily assigned to a phonological representation. Sound is given phonological meaning only after features are identified and categorized into distinct phonological representations. These phonological representations subsequently change how sound is perceived, as the auditory system goes from identifying all sound distinctions to making distinctions that are relevant to a given language. It is possible that children initially perceive differences between a wide variety of phonetic features in speech sounds, but as they develop the phonological representations for their language, they become less sensitive to those features that are irrelevant or redundant and may assign default status to those that are most frequent. Thus, as children develop their phonological representations, they may not yet be refined to adult-like levels. Children could initially store redundant features in their phonological representations, as they have not yet learned those features are unnecessary for phonemic categorization. As a result, some children's phonological representations might contain more features than those of adults.

Being able to accurately perceive speech sounds allows for the accurate formation of phonemic categories and the accurate identification of words. That being said, speech perception and production must also be efficient (Chomsky and Halle, 1968; Eulitz and Lahiri, 2004). Having to access and process extremely detailed representations of each phoneme would likely be inefficient for rapid speech processing. Moreover, it is questionable whether all features need to be stored in a phonological representation. One proposed solution to this problem is phonological underspecification, during which only the contrastive or not otherwise predictable phonological information (i.e., distinctive features) is stored for each phoneme (Kiparsky, 1985; Archangeli, 1988; Mohanan, 1991; Steriade, 1995; Hestvik and Durvasula, 2016).

While underspecification is proposed to be a language-universal phenomenon, the course of its development is presently unknown. As a language universal, is it present from birth? Or, is it something that develops over time, similar to the acquisition of phonemic categories? One aspect of phonological development could involve the establishment of specified and underspecified phonemes. Bernhardt (1992) proposed that children first develop and define the phonological role of underspecified features and then gradually define the phonological role of more specified features. For example, the [coronal] consonantal place of articulation feature (i.e., produced with tongue tip or blade, such as /t/ or /d/) is typically assumed to be less specified than other places of articulation (e.g., [labial] (i.e., produced with lips, such as /p/ or /b/) or [dorsal] (i.e., produced with dorsum of tongue, such as /k/ or /ɡ/) (e.g., Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013; Cummings et al., 2017). Given that [coronal] is the proposed default place of articulation, children should acquire this underspecified place of articulation early in development. If children slowly acquire specified features, more marked place of articulation features such as [labial] would become established at a later age. As a result, children are predicted to first acquire the least specified phonemes, and more specified phonemes are added over time as features are defined and categorized. Thus, American English-speaking children are expected to produce phonemes with 90% accuracy by the following ages: 2;11 (years; months)—/p b d m n h w/; 3;11—/t k ɡŋ f j/ 4;11—/v s z ʃ ʧ ʤ l/; 5;11—/ʒ ð ɹ/; and 6;11—/θ/ (Crowe and McLeod, 2020).

While behavioral studies have successfully identified children's categorical speech perception abilities, behavioral tasks offer little if any insight into their underlying phonological representations. However, neuroimaging tools have proven useful in examining phonological underspecification. Neural markers of phonological underspecification have primarily been examined using the framework established by the Featurally Underspecified Lexicon (FUL) model (Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002, 2010).

FUL predicts asymmetries in speech processing when an underspecified phoneme is contrasted with a more fully specified phoneme. A sound can directly *match* when the features extracted from the acoustic signal are the same as those in the phonological representation. A sound would be a *mismatch* when the features extracted from the acoustic signal are distinct from those in the phonological representation. A sound is a *no-mismatch* when the features extracted from the acoustic signal are consistent with the phonological representation, but because the phonological representation is not specified for a certain feature present in the speech signal, the input and the representation cannot exactly match (Schluter et al., 2016).

FUL's predictions have been tested using electrophysiological methodologies, such as event-related potentials (ERPs) that often measure mismatch negativity (MMN) responses (Näätänen et al., 2007). The MMN is an attention-independent neurophysiological response elicited by an acoustically different (deviant) stimulus when presented in a series of homogenous (standard) stimuli. Thus, the MMN is an automatic auditory change detection response in the brain and is thought to reflect stimulus discrimination and sensory memory (Sams et al., 1985); it is elicited by any discriminable acoustic contrast. The MMN is sensitive to language-specific speech sound representations (Näätänen et al., 1997; Kraus et al., 1998; Winkler et al., 1999; Phillips et al., 2000; Näätänen, 2001). Indeed, there is evidence to suggest that the MMN response is deviant in children with known language disabilities in response to tones (Korpilahti and Lang, 1994; Rinker et al., 2007; Ahmmed et al., 2008) and to speech syllables (Kraus et al., 1996; Uwer et al., 2002; Shafer et al., 2005; Volkmer and Schulte-Körne, 2018). Importantly, the MMN has been shown to index a person's ability to *behaviorally* discriminate between standard and deviant stimuli (Sams et al., 1985; Kraus et al., 1996). Children with better phoneme processing abilities have demonstrated larger MMN responses than children scoring lower on a phoneme processing test (Linnavalli et al., 2017).

In terms of underspecification, the MMN varies depending on whether the specified or underspecified phoneme is the standard (and deviant), as the standard stimulus sets up the feature expectations that the deviant stimulus will match, mismatch, or no-mismatch. In the match condition, the same feature is present in the deviant and standard stimuli, resulting in no MMN response. In the no-mismatch condition, the underspecified phoneme is the standard stimulus, which does not set up a feature expectation for the more specified deviant stimulus. As a result, little or no MMN response is expected in the no-mismatch condition. The true mismatch condition occurs when the more specified phoneme is the standard stimulus and sets up a specific feature expectation for the less specified deviant stimulus. The feature extracted from the deviant stimulus signal directly conflicts with that of the standard, resulting in a large MMN response.

FUL predicts that [coronal] phonemes have the default place of articulation because they contain less distinctive feature information in their phonological representations than phonemes with other places of articulation, such as [labial] or [dorsal]. As such, most of the previous ERP studies examining FUL have focused on place of articulation contrasts in German consonants and vowels (Eulitz and Lahiri, 2004; Scharinger and Lahiri, 2010; Cornell et al., 2011, 2013; Scharinger et al., 2012). While these studies provided support for [coronal] underspecification, few electrophysiological studies have tested [coronal] underspecification in English. Cummings et al. (2017) examined underspecification of /d/ and /b/, classified as [coronal] and [labial] respectively, in English-speaking adults. Each consonant was presented in a consonant-vowel (CV) combination (e.g., /bɑ/). Consistent with the predictions of FUL, the less specified /dɑ/ elicited a large MMN while no MMN was elicited by the more specified /bɑ/. Interestingly, not all participants demonstrated reliable mismatch responses. This suggested that [coronal] underspecification might not be a language universal phenomenon, at least as measured by the MMN (Scharinger et al., 2011).

Another way to test the language universal prediction of [coronal] underspecification is to examine the speech processing patterns of children. Thus, the primary goal of this study was to determine whether [coronal] underspecification occurs in young children. Using the same stimuli and stimulus presentation paradigm as Cummings et al. (2017), two early-acquired English consonants differing in place of articulation, [labial] /b/ and [coronal] /d/, were presented to 24 children (ages 4–6 years). If [coronal] underspecification is a language universal, it was predicted that the children would demonstrate asymmetrical response patterns similar to those of the adults in Cummings et al. (2017). That is, due to a place of articulation feature mismatch, the /dɑ/ deviant should elicit a large response when presented within the /bɑ/ standards, resulting in a large MMN. Alternatively, the /bɑ/ deviant would be a no-mismatch to the /dɑ/ standards, resulting in a much smaller MMN response. Such a result would indicate that young children have adult-like phonological representations, as suggested by the FUL.

An alternative, but not necessarily opposing, possibility is that /bɑ/ and /dɑ/ would elicit small and symmetrical MMN responses due to little response differences between the standards and deviants. Such a result would suggest that the two phonemes no-mismatch one another because there is no place of articulation feature contrast. This would be consistent with the predicted development of underspecification (Bernhardt, 1992). Thus, in this situation, [coronal] would still be the underspecified feature; however, the more marked [labial] place of articulation would not yet be correctly defined in the phonological representations. As a result, it could just be a matter of time before adult-like [coronal] underspecification patterns are present in children's responses.

In order to capture potential developmental trends and variations in phonological underspecification, children with typically developing *and* disordered phonological systems were included in the study. No previous study has used electrophysiological methods to examine speech processing in children with disordered phonological systems. Thus, a secondary goal of this study was to determine whether children with typically developing (TD) phonological systems and children with phonological disorders (PD) have distinctive speech processing neural signatures.

Phonological disorders are one subtype of speech sound disorders (McLeod and Baker, 2017). Children with speech sound disorders can demonstrate 'any combination of difficulties with perception, articulation/motor production, and/or phonological representation of speech segments (consonants and vowels), phonotactics (syllable and word shapes), and prosody (lexical and grammatical tones, rhythm, stress, and intonation) that may impact speech intelligibility and acceptability' (International Expert Panel on Multilingual Children's Speech, 2012, p. 1). It has been claimed that the speech errors children with PD make may be due to knowledge deficits at the level of phonemic detail and/or at the level of phonemic contrasts (Rvachew and Jamieson, 1995). Specifically, speech sound perception problems may arise, at least in some cases, from faulty representation of the speech signal in the central auditory processing centers (Kraus, 2001). Children with PD may have difficulty creating phonological representations due to their inaccurate perception of speech sounds (Macken, 1980; Chaney, 1988; McGregor and Schwartz, 1992).

While stable perceptual representations of speech sounds would allow children to form detailed phonological representations that can be generalized across experiences, unstable neural encoding of speech could affect children's ability to process rapidly changing acoustic information that differentiates phonemes (Carr et al., 2016). This unstable neural encoding could lead to the creation of fuzzy (i.e., less detailed) phonological representations that do not allow children with PD to accurately discriminate between sounds that share similar articulatory features (e.g., voicing, articulatory placement, and/or articulatory manner). More specifically, if children with PD have fuzzy phonological representations, their ability to accurately discriminate one sound from another may be impaired. This suggests that if children do not have an appropriately detailed underlying phonological representation to access during speech production, speech production errors would likely occur. Thus, ultimately, children's speech production abilities might be highly dependent on their speech perception abilities (Scott, 2012).

While no published research has examined underlying perceptual mechanisms implicated in PD, studies involving children with developmental language disorders, reading disorders, and/or childhood apraxia of speech have identified atypical electrophysiological discriminatory responses to speech sounds (Kraus et al., 1996; Uwer et al., 2002; Sharma et al., 2006; Froud and Khamis-Dakwar, 2012; Volkmer and Schulte-Körne, 2018). With 18–25% of children with PD going on to develop reading difficulties or receive a dyslexia diagnosis (Cabbage et al., 2018), one potential deficit that children with PD might share with children with developmental language disorders and reading disabilities is an impairment in phonological processing or phonological representation (Elbro and Jensen, 2005; Boada and Pennington, 2006; Pennington and Bishop, 2009; Cabbage et al., 2018). It is plausible that the underlying neural responses representing the discrimination of speech sounds in children with PD could also be atypical in nature. Thus, it is likely that children with PD might have phonological representations that are different from those of TD children.

One possibility is that children with PD have developmentally immature phonological systems, possibly due to exceptionally sparse, or fuzzy, phonological representations. Evidence for this hypothesis could be provided by distinct mismatch response patterns found in the children with PD, as compared to the TD children. For example, in young children, stimulus mismatch responses are often either not a clear negativity or are fully positive in polarity (Cheour et al., 2002; Maurer et al., 2003). Positive mismatch responses (PMR) have been associated with neural development (Mueller et al., 2012). Specifically, different neural networks could be involved in the PMR and MMN, with superficial neural networks being recruited to produce the MMN while deep cortical neurons might generate the PMR (Ponton et al., 1999, 2002; Kral and Eggermont, 2007). PMR responses have often been reported in children who have or are at-risk-for dyslexia (Volkmer and Schulte-Körne, 2018). Given that dyslexia is phonologically based, these results suggest that children with poorer or less detailed phonological representations, such as children with PD, might demonstrate developmentally immature PMR to speech sounds. Thus, if there is a developmental trajectory of stimulus mismatch responses, children would first demonstrate a PMR that gradually shifts in polarity to eventually result in a large MMN. The transition from PMR to MMN might occur much earlier in TD children, as compared to children with impaired phonological processing.

METHODS

Participants

Twenty-four children who were native speakers of (American) English participated in the study. Twelve of the children had typically developing (TD) phonological systems (9 male, 3 female; mean age: 5.8 years, range: 4.58–6.92 years). Twelve of the children had been previously diagnosed with a phonological disorder (PD) by a certified speech-language pathologist (6 male, 6 female; mean age 5.5 years, range: 4.00–6.92 years). All participants had normal vision and hearing within normal limits as determined by a standard audiometric screening and resided in a monolingual English-speaking household. The children with PD met the additional following criteria (**Table 1**):

TABLE 1 | Characteristics of the typically developing (TD) children and children with phonological disorders (PD).

	TD (n = 12)	PD (n = 12)
Assessment age in years	5.81 (0.96)	5.51 (1.04)
	Range = 4.33–6.92	Range = 4.00–6.92
	$t_{(22)}$ =0.732, $p > 0.47$	
GFTA–3: Standard score	107.17 (7.47)	65.75 (13.69)
	Range = 90–114	Range = 40–80
	$t_{(22)}$ =9.202, $p < 0.0001$	
GFTA-3: Percent Consonants Correct (PCC)	96.34 (3.85)	70.84 (14.75)
	Range = 86–100	Range = 35–85
	$t_{(22)}$ =5.795, $p < 0.0001$	
KLPA-3: Standard score	105.08 (7.22)	67.67 (9.28)
	Range = 90–112	Range = 50–79
	$t_{(22)}$ =11.028, $p < 0.0001$	
Hearing	Within normal limits	Within normal limits

Mean group scores and score range are presented. Standard deviations are reported within parentheses.
GFTA-3 refers to the standard score on the Goldman-Fristoe Test of Articulation 3rd Edition (Goldman and Fristoe, 2015). Standard scores between 85 and 115 are considered to be in the normal range.
KLPA-3 refers to the standard score on the Khan-Lewis Phonological Analysis 3rd Edition (Khan and Lewis, 2015). Standard scores between 85 and 115 are considered to be in the normal range.
Based on the Lifespan Reference Data for PCC (Austin and Shriberg, 1997), 5-year-old males should have an average PCC score of 86.9% while 5-year-old females should have an average PCC score of 87.3%.

an oral-peripheral mechanism exam completed within normal limits (Robbins and Klee, 1987); speech articulation scores on the *Goldman-Fristoe Test of Articulation—3rd edition* (GFTA-3; Goldman and Fristoe, 2015) at least 1.25 standard deviations below the mean (standard scores of 80 or below), phonological process scores (i.e., speech error patterns) on the *Khan-Lewis Phonological Analysis—3rd edition* (KLPA-3; Khan and Lewis, 2015) at least 1.25 standard deviations below the mean (standard scores below 80), and a percent consonants correct (PCC) score (Shriberg and Kwiatkowski, 1982) for all the consonants on the GFTA-3 of 85% or less. Children with TD speech had standard scores of 90 or higher on the GFTA-3 and KLPA-3 and GFTA-3 PCC scores of 85% or higher.

Regardless of their group assignment, all children correctly produced /b/ and /d/ in all the words on the GFTA-3 (i.e., four occurrences of /b/ and five occurrences of /d/). Thus, a behavioral speech perception task was not employed because all children were capable of accurately producing the study's phonemes of interest. This study was approved by the university institutional review board and a parent of each participant signed informed consent in accordance with the university human research protection program.

Stimuli

The stimuli were the same as those used in Cummings et al. (2017) with adult participants; please refer there for more detailed information. Briefly, syllables (consonant + /ɑ/) were pronounced by a male North American English speaker. The average intensity of all the syllable stimuli was normalized to 65 dB SPL. Syllable duration was minimally modified (by shortening the vowel duration) so that all syllables were 375 ms in length. The vowel editing process began by identifying the most consistent, steady state portion (i.e., the middle) of the vowel. Then, a single sinusoid cycle of the vowel production was measured at 8 ms for /bɑ/ and 9 ms for /dɑ/. To achieve the target 375 ms for the entire syllable, 2 sinusoid cycles (i.e., 16 ms) were deleted in the /bɑ/ syllable and four cycles (i.e., 36 ms) were deleted in the /dɑ/ syllable. Each syllable token used in the study was correctly identified by at least 15 adult listeners.

Children heard two oddball stimulus sets, each containing the same four English speech consonant-vowel (CV) syllables: "ba" (/bɑ/), "da" (/dɑ/), "pa" (/pɑ/), and "ga" (/ɡɑ/). In one stimulus set, /bɑ/ served as the standard syllable, with the other three CV syllables serving as deviants. In the second stimulus set, /dɑ/ served as the standard syllable, with the other three syllables being deviants. Only responses to the /bɑ/ and /dɑ/ syllables will be addressed further since they served as both standard and deviant stimuli, which allowed for the creation of same-stimulus identity difference waves.

Stimulus Presentation

The stimuli were presented in blocks containing 237 standard stimuli and 63 deviant stimuli (21 per deviant), with five blocks of each stimulus set being presented to each participant (i.e., 10 blocks total). Each block lasted ∼6 min and the participants were given a break between blocks when necessary. Within the block, the four stimuli were presented using an oddball paradigm in which the three deviant stimuli (probability = 0.07 for each) were presented in a series of standard stimuli (probability = 0.79). Stimuli were presented in a pseudorandom sequence and the onset-to-onset inter-stimulus interval varied randomly between 600 and 800 ms. The syllables were delivered by stimulus presentation software (Presentation software, www.neurobs.com). The syllable sounds were played via two loudspeakers situated 30 degrees to the right and left from the midline 120 cm in front of a participant, which allowed the sounds to be perceived as emanating from the midline space. The participants sat in a sound-treated room and watched a silent cartoon video of their choice. The recording of the ERPs took ∼1 h.

EEG Recording and Averaging

Sixty-six channels of continuous EEG (DC-128 Hz) were recorded using an ActiveTwo data acquisition system (Biosemi, Inc., Amsterdam, Netherlands) at a sampling rate of 256 Hz. This system provides "active" EEG amplification at the scalp that substantially minimizes movement artifacts. The amplifier gain on this system is fixed, allowing ample input range (−264 to 264 mV) on a wide dynamic range (110 dB) Delta- Sigma ($\Delta\Sigma$) 24-bit AD converter. Sixty-four channel scalp data were recorded using electrodes mounted in a stretchy cap according to the International 10–20 system. Two additional electrodes were placed on the right and left mastoids. Eye movements were monitored using FP1/FP2 (blinks) and F7/F8 channels (lateral movements, saccades). During data acquisition, all channels were referenced to the system's internal loop (CMS/DRL sensors located in the centro-parietal region), which drives the average potential of a subject (the Common Mode voltage) as close as possible to the Analog-Digital Converter reference voltage (the amplifier "zero"). The DC offsets were kept below 25 microvolts at all channels. Off-line, data were re-referenced to the common average of the 64 scalp electrode tracings.

Data processing followed an EEGLAB (Delorme and Makeig, 2004) processing pipeline. Briefly, data were high-pass filtered at 0.5 Hz using a pass-band filter. Line noise was removed using the CleanLine EEGLAB plugin. Bad channels were rejected using the trimOutlier EEGLAB plugin and the removed channels were interpolated. Source level contributions to channel EEG were decomposed using Adaptive Mixed Model Independent Component Analysis (AMICA) (Palmer et al., 2008) in EEGLAB (http://www.sccn.ucsd.edu/eeglab). Non-artifact, independent component (IC) scalp topographies were modeled as projections of single equivalent dipoles and clustered on the basis of dipole locations (Jung et al., 2000; Delorme and Makeig, 2004).

Epochs containing time points within a window encompassing 100 ms pre-auditory stimulus to 800 ms post-stimulus were baseline-corrected with respect to the pre-stimulus interval and averaged by stimulus type. On average, the individual data of TD children contained 763 (SD = 72) /bɑ/ standard syllable trials, 706 (SD = 162) /dɑ/ standard syllable trials, 85 (SD = 19) /bɑ/ deviant syllable trials, and 91 (SD = 9) /dɑ/ deviant syllable trials. On average, the individual data of children with PD contained 799 (SD = 172) /bɑ/ standard syllable trials, 895 (SD = 185) /dɑ/ standard syllable trials, 95

(SD = 17) /bɑ/ deviant syllable trials, and 96 (SD = 21) /dɑ/ deviant syllable trials.

ERP and EEG Measurements

Three different data analysis strategies were used in the present study: (1) traditional mean amplitude repeated measure ANOVA analyses using averaged data, (2) cluster-based permutation analyses of averaged data (Bullmore et al., 1999; Groppe et al., 2011), and (3) general linear modeling of epoched (i.e., unaveraged) data (Pernet et al., 2011).

Mean Amplitude Measurements of Averaged Data

In an oddball paradigm, the MMN is typically examined by subtracting the standard ERP response from the deviant response in difference waves. The dual stimulus set nature of the present study allowed for the creation of "same-stimulus," or identity, difference waveforms. These difference waves were created by subtracting the ERP response of a stimulus serving as the standard from that of the same stimulus serving as the deviant, across stimulus sets. For example, the ERP response for /bɑ/ as the standard was subtracted from the ERP response for /bɑ/ as the deviant (of the reversed stimulus set) (Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013). The creation of identity difference waveforms eliminates the potential confound that variations in ERP morphology may result from acoustic stimulus differences, since the same stimulus is used to elicit both the standard and deviant responses.

Peak measurement of MMN was a multi-step process. While the MMN is typically maximal over fronto-central midline electrode sites (Näätänen et al., 1992), evidence of MMN-type activity was present in many electrodes. That observation, along with the fact that underspecification has not been examined in children before, nor has any prior ERP study examined phonological processing in children with PD, led to the inclusion of 28 electrodes in the analyses (F5/F6, F3/F4, F1/F2, Fz, FC5/FC6, FC3/FC4, FC1/FC2, FCz, C5/C6, C3/C4, C1/C2, Cz, CP5/CP6, CP3/CP4, CP1/CP2, and CPz)[1]. The most prominent mismatch response was observed in both groups 200–300 ms post-syllable onset. As this was consistent with the timing of the adult MMN response (Cummings et al., 2017), this time window was selected for data analysis. Phonological underspecification in the identity difference waves was analyzed using a Group (TD, PD) x Phoneme Type (/bɑ/, /dɑ/) x Anterior-Posterior (four levels) × Left-Right (seven levels) repeated measure ANOVA.

Given that the difference waves were generated from the standard and deviant syllable ERPs, the mean amplitude measurements of the standard and deviant waveforms were taken from the same 100 ms time as that of the MMN: 200–300 ms post-syllable onset. In terms of ERP waveform morphology, this measurement approximately captured the auditory N2. Phonological underspecification in these standard and deviant ERP waveforms was analyzed using a Group (TD, PD) x Phoneme Type (/bɑ/, /dɑ/) × Trial Type (Standard, Deviant) x Anterior-Posterior (4 levels) x Left-Right (seven levels) repeated measure ANOVA. Partial eta squared (η^2) effect sizes are also reported for all significant effects and interactions. When applicable, Geiser-Greenhouse corrected *p*-values are reported.

Cluster Mass Permutation Tests of Averaged Data

The ERPs were submitted to repeated measures two-tailed cluster-based permutation tests (Bullmore et al., 1999; Groppe et al., 2011). These permutation test analyses provide better spatial and temporal resolution than conventional ANOVAs while maintaining weak control of the family-wise alpha level (i.e., it corrects for the large number of comparisons). To estimate the distribution of the null hypothesis, 2,500 permutations were used, which was more than twice the number recommended for a family-wise alpha level of 0.05 (Manly, 2006). These analyses enabled identification of differences between the underspecified /dɑ/ and the more specified /bɑ/. Thus, the high temporal resolution of this analysis could be used to identify a specific time period during which indices of underspecification were present.

Five different tests were conducted: (1) /bɑ/ vs. /dɑ/ identity MMN difference waveforms, (2) /bɑ/ vs. /dɑ/ standard ERPs, (3) /bɑ/ vs. /dɑ/ deviant ERPs, (4) /bɑ/ standard vs. /bɑ/ deviant ERPs, and (5) /dɑ/ standard vs. /dɑ/ deviant ERPs. These tests were conducted to identify group level (TD vs. PD) differences. The tests were also conducted separately for each group to examine phoneme type and/or trial type differences that were specific to TD children and/or children with PD. Each test included 28 different electrodes that encompassed four different anterior-posterior levels and seven different laterality measures: F5/F6, F3/F4, F1/F2, Fz, FC5/FC6, FC3/FC4, FC1/FC2, FCz, C5/C6, C3/C4, C1/C2, Cz, CP5/CP6, CP3/CP4, CP1/CP2, CPz. All of the time points (measured every 4 ms; 91 total time points) between 0 and 350 ms at the 28 scalp electrodes were included in the test (i.e., 2,548 total comparisons).

T-tests were performed for each comparison using the original data and 2,500 random within-participant permutations of the data. For each permutation, all t-scores corresponding to uncorrected *p*-values of 0.05 of less were formed into clusters. Electrodes within about 5.44 cm of one another were considered spatial neighbors, and adjacent time points were considered temporal neighbors. The sum of the t-scores in each cluster was the "mass" of that cluster. The most extreme cluster mass in each of the 2,501 sets of tests was recorded and used to estimate the distribution of the null hypothesis (i.e., no difference between conditions). The permutation cluster mass percentile ranking of each cluster from the observed data was used to derive *p*-values assigned to each member of the cluster. T-scores that were not included in a cluster were given a *p*-value of 1.

General Linear Modeling (GLM) of Epoched Data

GLM analyses were used to help account for the correlation in time and space dimensions found in EEG data, and to provide an alternate analysis technique to the repeated measure ANOVAs commonly used in ERP data analysis. They were also more robust to potential noise introduced by the trial number imbalance found in the standard and deviant syllable data generated by the oddball paradigm.

[1]These electrodes encompass four anterior-posterior levels (Frontal, Frontal-Central, Central, Central-Parietal) and seven left-right laterality levels (Far Left-5, Mid Left-3, Close Left-1, Midline-z, Close Right-2, Mid Right-4, Far Right-6).

Following the protocol described in previous studies (Rousselet et al., 2011; Cummings et al., 2017), subjects' epoched data were modeled using LIMO EEG, an open source Matlab toolbox for hierarchical GLM, compatible with EEGLAB: https://gforge.dcn.ed.ac.uk/gf/project/limo_eeg/ (Pernet et al., 2011). The general linear model was used to examine single-trial ERP amplitudes, in microvolts, independently at each time point and each electrode. Parameters (β-values) were estimated at each electrode and time point independently, yielding a matrix of 64 (electrodes) x 103 (time points, from 0 to 400 ms post-stimulus in 4 ms steps) for each regressor. Similar electrode x time point matrices were computed for R^2, F, and *p*-values for both the overall models and for each regressor (partial *F*-values). Probability values were determined using a permutation approach for which trial labels were permuted 1,000 times using a bootstrap-t technique (Wilcox, 2005). To examine underspecification of /dɑ/ as compared to /bɑ/ in standard and deviant trials, four GLM analyses of epoched data were conducted at the single-subject level: (1) /bɑ/ vs. /dɑ/ standards, (2) /bɑ/ vs. /dɑ/ deviants, (3) /bɑ/ standards vs. /bɑ/ deviants, and (4) /dɑ/ standards vs. /dɑ/ deviants.

Single subject GLM analyses

For each analysis, bootstrap paired *t*-tests were computed between the contrasts of interest at all time points across the entire scalp. Due to the small number of deviant trials and hence a low signal-to-noise ratio, most individual participants' analyses were not significant when controlled for multiple comparisons. Thus, for the purpose of examining phonological underspecification at a single-subject level, uncorrected data are reported.

To quantify and compare the individual results, two analyses were conducted. First, using the full-scalp uncorrected comparison analyses, the data of each participant were examined for a significant phoneme type or trial type difference (positive or negative) of at least 20 continuous milliseconds at electrode FCz during the 200–400 ms time window, which was the time window of the mismatch response observed in the grand averaged waveforms. The second analysis involved identifying a significant continuous 20 ms phoneme type or trial type difference in at least five separate electrode sites during the 200–400 ms time window; this analysis did not have to include electrode FCz, though this electrode was included in some cases.

RESULTS

In the ERP waveforms of both the TD children and children with PD, the standard and deviant /bɑ/ and /dɑ/ syllables elicited auditory P1/P2 at ca. 115 ms and auditory N2 at ca. 300 ms (**Figure 1**). In the same-stimulus identity difference waves of the TD children, an MMN response was observed at ca. 250 ms. Conversely, in the same-stimulus identity difference waves of the children with PD, a positive mismatch response (PMR) (Mueller et al., 2012) was evident at ca. 275 ms (**Figure 1**).

ERP Mean Amplitude Results

Identity Difference Waves: MMN

The MMN mean amplitude did not differ between the TD children and the children with PD ($p > 0.28$). For the combined groups, the mean amplitude of the /dɑ/ identity difference waveform was not significantly different from the /bɑ/ identity waveform ($p > 0.94$). A main effect of electrode location anteriority was found [$F_{(3,66)} = 4.894$, $p < 0.03$, $\eta^2 = 0.182$]. Mismatch responses tended to be more negative across posterior electrodes as compared to anterior electrodes, though pairwise comparisons revealed no significant differences. No main effect of left-right electrode location was observed ($p > 0.91$) (**Figures 2**, **3**).

However, a significant group x phoneme type interaction was observed [$F_{(1,22)} = 9.817$, $p < 0.006$, $\eta^2 = 0.309$] (**Figure 4**). When examining just the responses elicited by /bɑ/, the TD children had a negative response (i.e., MMN) while the children with PD demonstrated a large positive mismatch response (i.e., PMR) [$F_{(1,22)} = 6.126$, $p < 0.03$, $\eta^2 = 0.218$] (**Figures 4**, **8B**); significant group differences were not observed in the /dɑ/ mismatch responses ($p > 0.32$). The TD children's mismatch responses to /bɑ/ and /dɑ/ did not significantly differ ($p > 0.08$) (**Figures 4**, **9A**). The /bɑ/ mismatch response of the children with PD was significantly more positive than was the mismatch response to /dɑ/ [$F_{(1,11)} = 7.744$, $p < 0.02$, $\eta^2 = 0.413$] (**Figures 4**, **9D**). Thus, this interaction was primarily driven by the opposing directions of the group mismatch responses elicited by /bɑ/.

Standard and Deviant Waveforms

No group main effect was found ($p > 0.30$); overall, the ERP mean amplitudes of the TD children and the children with PD did not differ. A main effect of phoneme type was found, as across both groups, the ERP mean amplitude of the /bɑ/ waveform was significantly more negative (i.e., larger) than that of /dɑ/ [$F_{(1,22)} = 25.049$, $p < 0.0001$, $\eta^2 = 0.532$]. No main effect of trial type was observed ($p > 0.37$); the standard and deviant waveforms did not differ from each other. Across both groups, a main effect of left-right [$F_{(6,132)} = 4.341$, $p < 0.02$, $\eta^2 = 0.165$] electrode location was found. Significantly smaller responses were recorded over the most lateral electrodes (Far Left/Far Right), as compared with their nearest neighbors (Mid Left/Mid Right; all $p < 0.02$); no other electrode left-right effects were significant. No effect of anterior-posterior electrode location was observed ($p > 0.17$) (**Figures 5**, **6**).

On the other hand, a group x phoneme type interaction was observed [$F_{(1,22)} = 7.347$, $p < 0.02$, $\eta^2 = 0.250$] (**Figure 7**). While the TD children's ERP responses to /bɑ/ were significantly more negative than their responses to /dɑ/ [$F_{(1,11)} = 59.771$, $p < 0.0001$, $\eta^2 = 0.845$], no phoneme type differences were observed in the children with PD ($p > 0.21$). A group x phoneme type x trial type interaction was also found [$F_{(1,22)} = 9.817$, $p < 0.006$, $\eta^2 = 0.309$]. Significant group differences were not observed when responses to the /bɑ/ standards ($p > 0.46$), /dɑ/ standards ($p > 0.47$), and /dɑ/ deviants ($p > 0.89$) were examined separately. However, the ERP responses elicited by the /bɑ/

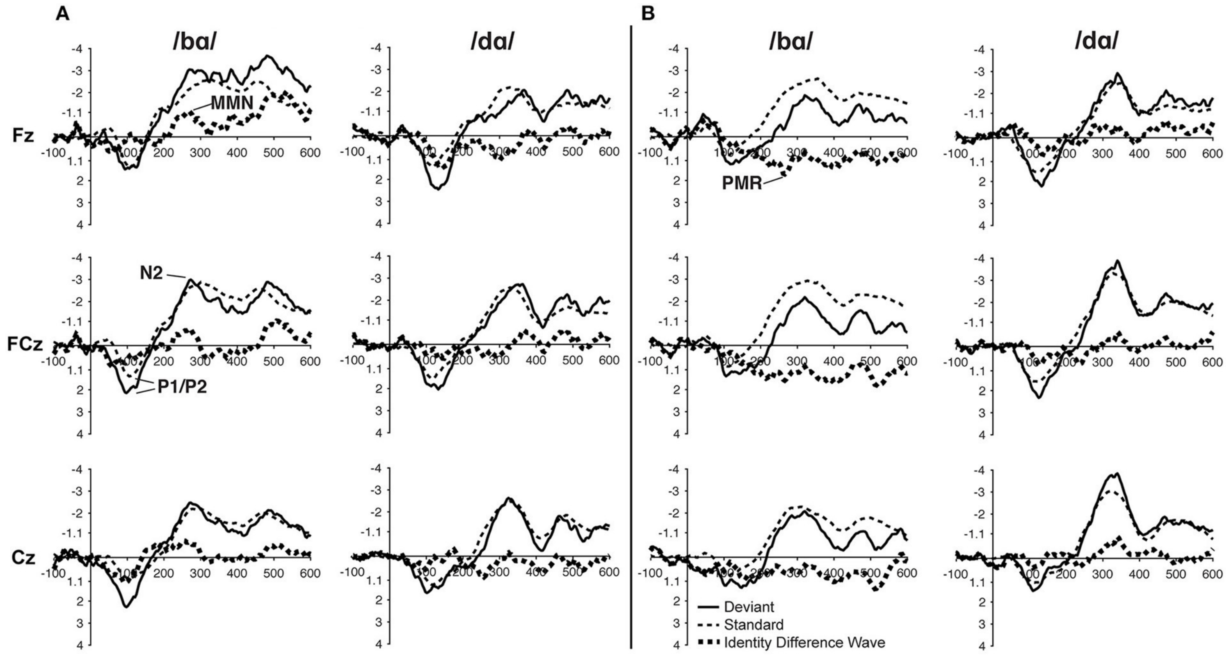


FIGURE 1 | ERP waveforms elicited in the **(A)** typically developing (TD) children and **(B)** children with phonological disorders (PD). In each panel, the /bɑ/ syllable response is presented on the left and the /dɑ/ syllable response is on the right. The deviant waveforms represent the neural responses when the deviant syllable was presented within a stream of the opposite syllable standards. Subtracting the standard syllable response from the deviant syllable response resulted in the identity difference waves. Note that negative is plotted *up* in all waveforms.

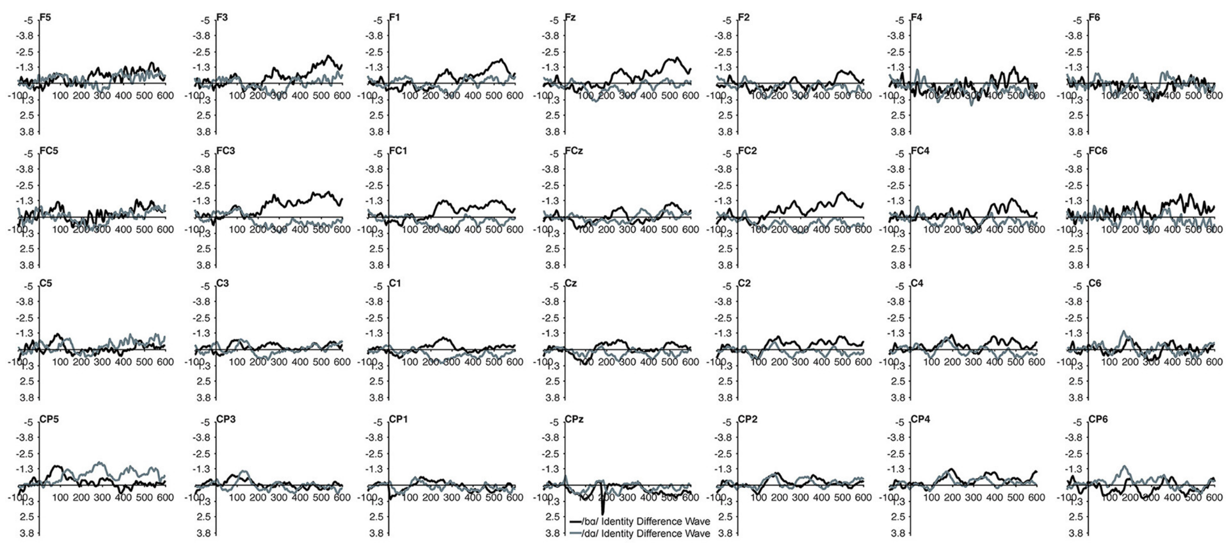


FIGURE 2 | Identity difference waveforms of the TD children elicited by /bɑ/ and /dɑ/ across all 28 electrodes included in all analyses. The /bɑ/ responses are in black while the /dɑ/ responses are in gray.

deviants in the TD children were significantly larger than those of the children with PD [$F_{(1,22)} = 4.442$, $p < 0.05$, $\eta^2 = 0.168$] (**Figures 7**, **8A**). Thus, phonological underspecification group differences were most prevalent in the responses elicited by the /bɑ/ deviants.

Cluster Permutation Analysis Results

Identity Difference Waveforms

TD vs. PD identity difference waves

Cluster-level mass permutation procedures encompassing the timeline of the P1/P2, N2, and MMN (0–350 ms) were applied

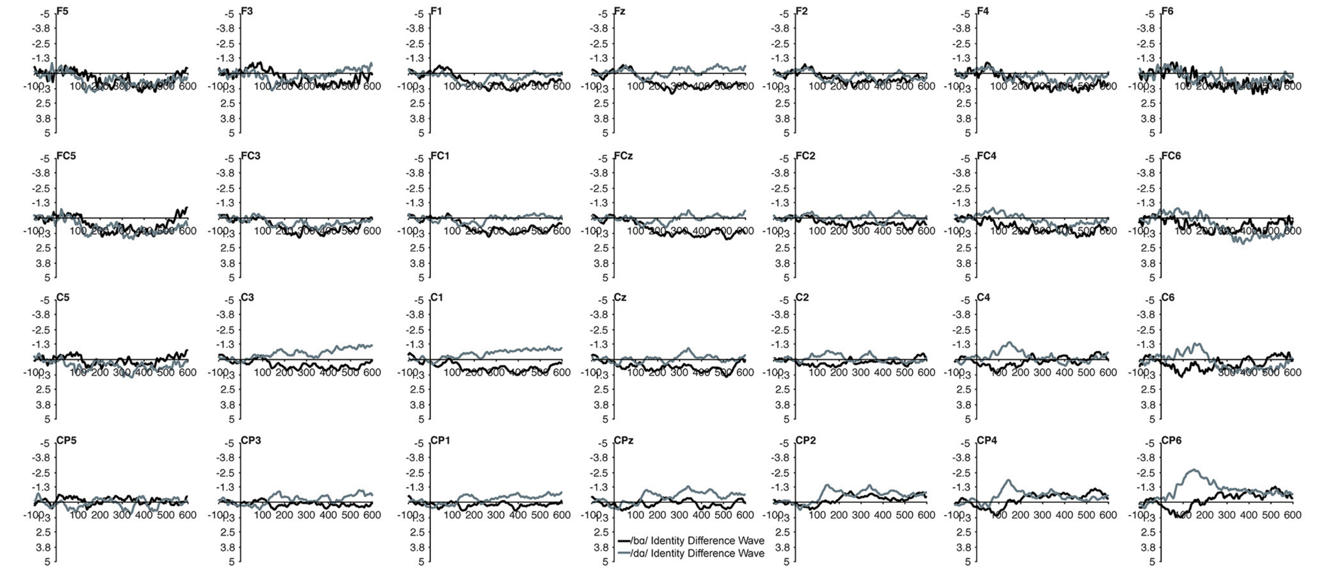


FIGURE 3 | Identity difference waveforms of the children with PD elicited by /bɑ/ and /dɑ/ across all 28 electrodes included in all analyses. The /bɑ/ responses are in black while the /dɑ/ responses are in gray. Error bars represent SEM.

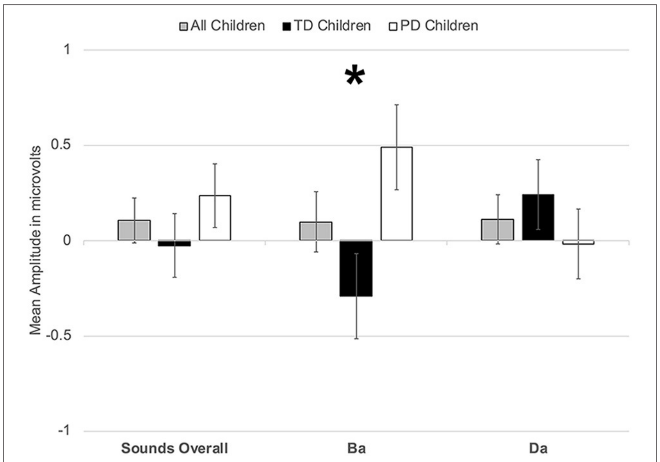


FIGURE 4 | Average mean amplitudes for mismatch responses measured in identity difference waves from 200 to 300 ms post-syllable onset. Responses from the typically developing children are in black, responses from children with phonological disorders are in white, and combined responses across groups are in gray. Error bars represent SEM. The TD children demonstrated a negative mismatch response (MMN) to /bɑ/ while the children with PD demonstrated a positive mismatch response (PMR). The The * symbol represents a significant difference.

to the data. **Figure 8B** shows one cluster extending from 152 to 316 ms that signified the time period during which the /bɑ/ difference waves of the TD children differed from those of the children with PD; the smallest significant t-score was: $t_{(23)} = 2.07$, $p < 0.05$. This effect was driven by essentially opposite mismatch responses, with an MMN response present in the difference waves of the TD children, while a PMR was present in the difference waves of the children with PD. No significant clusters were identified when examining group differences in the /dɑ/ identity difference wave responses.

TD identity difference waves: /bɑ/ vs /dɑ/

No significant clusters were identified when examining the difference between the /bɑ/ and /dɑ/ identity difference waves.

PD identity difference waves: /bɑ/ vs /dɑ/

A predominantly right hemisphere cluster extending from 62 to 214 ms signified the time period during which the /dɑ/ difference wave differed from the /bɑ/ difference wave; the smallest significant t-score was: $t_{(11)} = 2.208$, $p < 0.007$ (**Figure 9D**). During this time period, a PMR response was seen in the /bɑ/ difference wave while no significant mismatch response was evident in the /dɑ/ difference wave.

Standard and Deviant Waveforms

TD vs. PD standard and deviant waveforms

Consistent with the mean amplitude measurements, the group comparison of /bɑ/ deviant responses yielded a significant cluster. A broadly distributed effect from 125 to 296 ms signified the time period during which the /bɑ/ deviants of the TD children differed from those of the children with PD; the smallest significant t-score was: $t_{(22)} = 2.075$, $p < 0.02$ (**Figure 8A**). Since this time window primarily encompassed the time period between the peak of the auditory P1/P2 and the peak of the auditory N2 ERP responses, this difference implies that the /bɑ/ deviant response elicited in the TD children was significantly more negative (i.e., larger N2) than that of the children with PD. No other significant clusters were identified in the group comparisons.

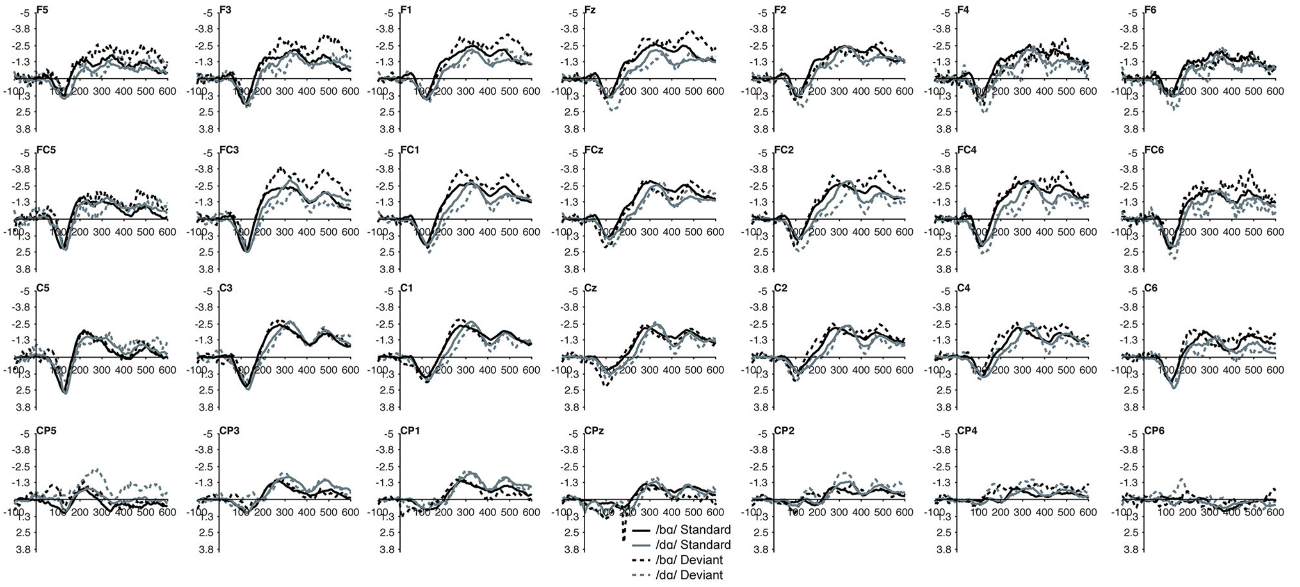


FIGURE 5 | Standard and deviant waveforms of the TD children elicited by /bɑ/ and /dɑ/ across all 28 electrodes included in all analyses. The /bɑ/ responses are in black while the /dɑ/ responses are in gray. Standard responses are solid lines while deviant response are dashed lines.

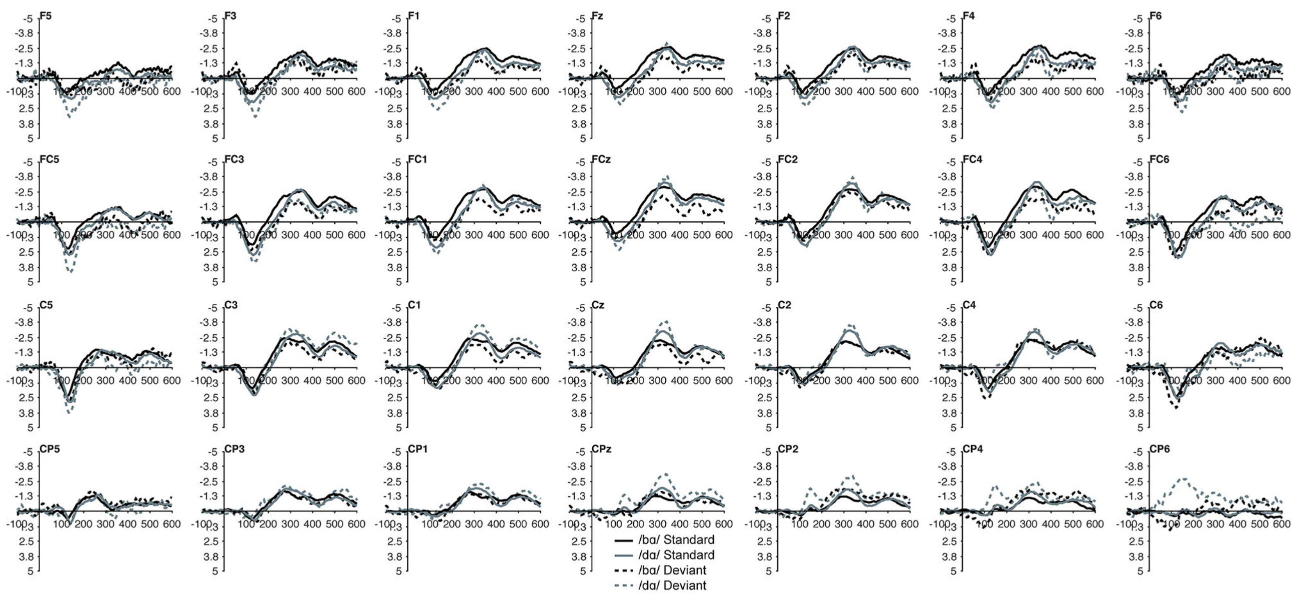


FIGURE 6 | Standard and deviant waveforms of the children with PD elicited by /bɑ/ and /dɑ/ across all 28 electrodes included in all analyses. The /bɑ/ responses are in black while the /dɑ/ responses are in gray. Standard responses are solid lines while deviant response are dashed lines.

TD standard and deviant waveforms

No significant clusters were identified when examining the difference between the /bɑ/ standards and /bɑ/ deviants, or the difference between /dɑ/ standards and /dɑ/ deviants. Thus, neither phoneme type elicited a mismatch response wherein the deviant stimulus was reliably larger (or smaller) than the corresponding standard stimulus.

When contrasting the phoneme differences, a broadly distributed effect from 128 to 296 ms signified the time period during which the /dɑ/ standards differed from the /bɑ/ standards; the smallest significant t-score was: $t_{(11)} = -2.02$, $p < 0.001$ (**Figure 9B**). This difference implies that the /bɑ/ standards elicited more negative (i.e., larger N2) responses than did the /dɑ/ standards.

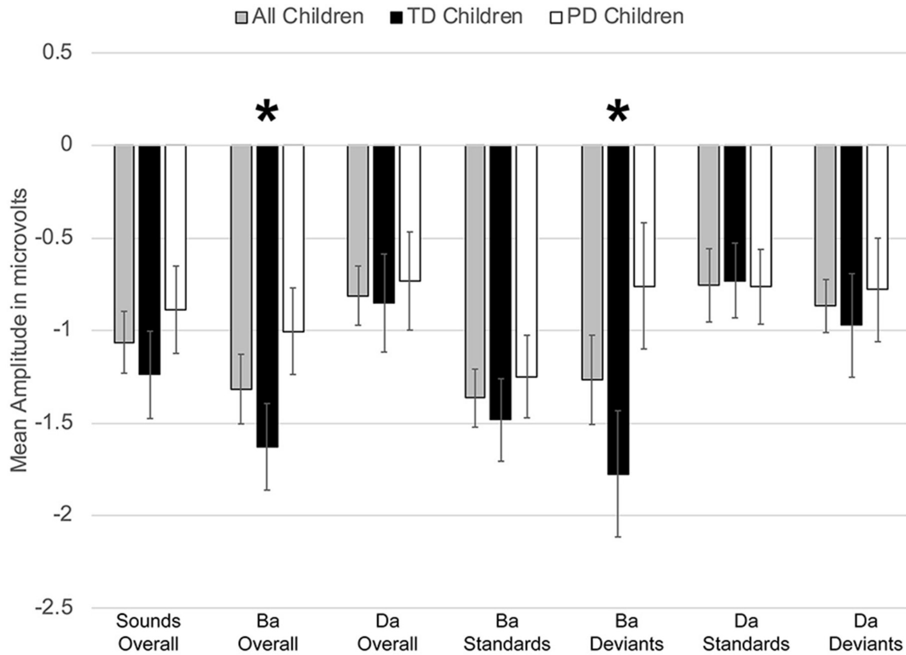


FIGURE 7 | Average mean amplitudes for standard and deviant ERPs from 200 to 300 ms post-syllable onset. Responses from the typically developing children are in black, responses from children with phonological disorders are in white, and combined responses across groups are in gray. Error bars represent SEM. The ERP responses elicited by the /bɑ/ deviants in the TD children were significantly larger (i.e., more negative) than those of the children with PD. The * symbol represents a significant difference.

Similarly, when the /bɑ/ and /dɑ/ deviant syllables were contrasted, a significant effect from 132 to 316 ms signified the time period during which the /bɑ/ deviants elicited more negative (i.e., larger N2) responses than the /dɑ/ deviants; the smallest significant t-score was: $t_{(11)} = -2.211$, $p < 0.002$ (**Figure 9C**). This effect was more localized to the fronto-central and central electrode locations.

PD standard and deviant waveforms

No significant clusters were identified when examining the difference between the /bɑ/ standards and /bɑ/ deviants, or the difference between /dɑ/ standards and /dɑ/ deviants. A broadly distributed effect from 50 to 296 ms signified the time period during which the /bɑ/ standards differed from the /dɑ/ standards; the smallest significant t-score was: $t_{(11)} = -2.204$, $p < 0.002$ (**Figure 9E**). Similarly, when the /bɑ/ and /dɑ/ deviant syllables were contrasted, a significant right hemisphere localized cluster was present from 62 to 183 ms; the smallest significant t-score was: $t_{(11)} = 2.20$, $p < 0.04$ (**Figure 9F**).

Single-Trial GLM Analyses

The single-subject data show the prevalence of potential [coronal] underspecification in English-speaking children. Individual participants' full scalp analyses of the four comparisons are presented in the online **Supplementary Material**. While the vast majority of the participants' individual significant effects occurred during the time period of the mismatch response (∼200–400 ms post-syllable onset), some effects were observed both before and after this time window. **Table 2** provides an overview of the single-subject findings.

TD Children

In the /bɑ/ standard vs. deviant analysis, 3/12 participants demonstrated a significant difference between the trial types at FCz, with 7/12 participants demonstrating a significant difference elsewhere across at least five electrodes. In the /dɑ/ standard vs. deviant analysis, 2/12 participants had a significant difference between trial types at FCz, though 10/12 participants demonstrated a significant trial type difference elsewhere. Thus, more TD children demonstrated some evidence of a /dɑ/ mismatch response than a /bɑ/ mismatch response. Indeed, five children only demonstrated distributed mismatch responses to /dɑ/, while two children only demonstrated responses to /bɑ/.

When the standards of the two syllables were compared, ERP activities in 6/12 participants differentiated /bɑ/ and /dɑ/ at FCz, and all 12 participants demonstrating a sensitivity to phoneme type across other scalp locations. The differences between the /bɑ/ and /dɑ/ deviants were not quite as prevalent, as 3/12 participants demonstrated a phoneme type difference at FCz, and 10/12 participants showed a phoneme type difference at other electrode sites.

Children With PD

In the /bɑ/ standard vs. deviant analysis, 4/12 participants demonstrated a significant difference between the trial types at FCz, with 9/12 participants demonstrating a significant trial

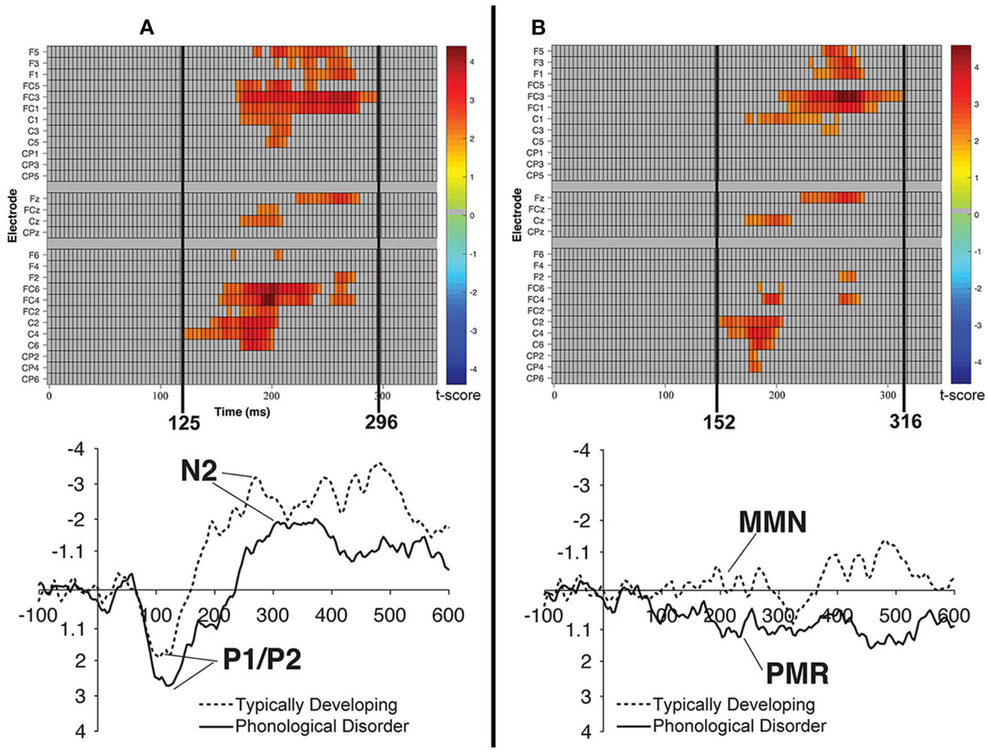


FIGURE 8 | Raster diagrams and waveforms illustrating group (TD vs. PD) differences in the processing of the /bɑ/ stimuli. On the left **(A)** are group differences in response to the /bɑ/ deviant. TD children demonstrated a more negative response than did children with PD. On the right **(B)** are group differences in response to the /bɑ/ identity difference wave. TD children demonstrated a negative mismatch response while children with PD demonstrated a positive mismatch response. For the raster diagrams, colored rectangles indicate electrodes/time points in which the ERPs to one stimulus are significantly different from those to another. The color scale dictates the size of the *t*-test result, with dark red and blue colors being more significant. Gray areas indicate electrodes/time points at which no significant differences were found. Note that the electrodes are organized along the y-axis somewhat topographically. Electrodes on the left and right sides of the head are grouped on the figure's top and bottom, respectively; midline electrodes are shown in the middle. Within those three groupings, y-axis top-to-bottom corresponds to scalp anterior-to-posterior.

type difference elsewhere across at least five electrodes. In the /dɑ/ standard vs. deviant analysis, 3/12 participants showed a significant trial type difference at FCz and 10/12 participants demonstrated a significant trial type difference elsewhere. Thus, evidence of mismatch responses to both /bɑ/ and /dɑ/ was present in the children with PD. Two children only demonstrated mismatch responses to /dɑ/, one child only demonstrated a mismatch response to /bɑ/, and one child did not demonstrate a response to either syllable.

When the standards of the two syllables were compared, 1/12 participants demonstrated a significant difference between /bɑ/ and /dɑ/ at FCz, while 11/12 participants demonstrated a phoneme type difference across other scalp locations. The differences between the /bɑ/ and /dɑ/ deviants were less clear, as 2/12 participants demonstrated a significant phoneme type difference at FCz, and 9/12 participants showed a phoneme type difference at other electrode sites.

DISCUSSION

This study examined phonological underspecification in 4- to 6-year-old children with typically developing (TD) and disordered (PD) phonological systems. Two phonemes, [labial] /b/ and [coronal] /d/, were presented to children within consonant-vowel syllables. In the TD children, no asymmetrical MMN responses were elicited by the underspecified /dɑ/ and specified /bɑ/; children's mismatch responses were equivocal. On the other hand, in the children with PD, the /bɑ/ mismatch response was significantly more positive than was the mismatch response to /dɑ/. Both of these findings were contrary to previous to adult [coronal] underspecification findings (Cummings et al., 2017) wherein the underspecified /dɑ/ elicited a larger MMN than did the specified /bɑ/.

Analysis of single-subject responses via GLM did not reveal any predictable underspecification patterns. Overall, more participants demonstrated a measurable /dɑ/ mismatch response as compared to /bɑ/. However, there was variability, as some children had a /bɑ/ mismatch response, but no /dɑ/ mismatch response. That so many children did not demonstrate mismatch asymmetries in their processing of /bɑ/ and /dɑ/ provides strong evidence against the language universality of [coronal] underspecification. Or, at the very least, the predictions should be modified to consider how [coronal] underspecification might develop in children.

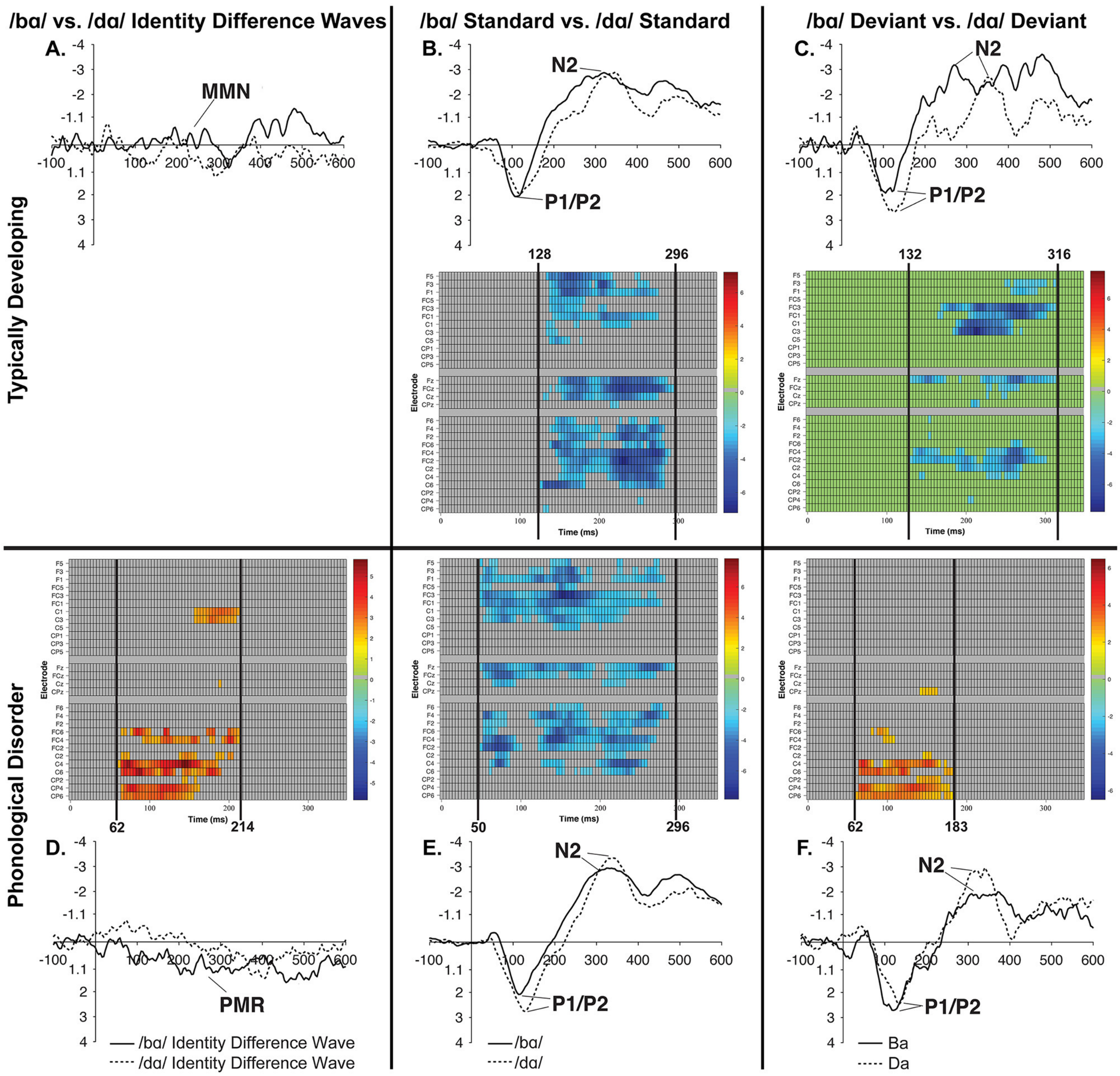


FIGURE 9 | Raster diagrams and waveforms illustrating /bɑ/ and /dɑ/ processing differences in the TD children and children with PD. Responses from TD children are on the top **(A–C)**, while responses from children with PD are on the bottom **(D–F)**. In the left column are /bɑ/ and /dɑ/ identity difference wave comparisons **(A,D)**, the middle column contrasts /bɑ/ and /dɑ/ standards **(B,E)**, and the right column contrasts /bɑ/ and /dɑ/ deviants **(C,F)**. Both groups of children demonstrated similar processing patterns in the standard and deviant analyses, as the /bɑ/ elicited more negative responses than did the /dɑ/. The /bɑ/ mismatch response of the children with PD was significantly more positive than the mismatch response to /dɑ/, especially over the right hemisphere electrodes; no significant differences were observed in the identity difference waves of the TD children.

A Developmental Model of Phonological Underspecification

Since children did not demonstrate the FUL-predicted adult MMN asymmetry pattern, it is likely that underspecification develops over time. It is proposed that there is a developmental trajectory for phonological underspecification, and the TD children and children with PD were functioning at different points on this continuum. Schluter et al. (2016) predicted multiple types of potential underspecification MMN responses depending on the type of feature specification and the manner in which the phonological system characterized sounds. This type of framework has been adopted here to represent three potential stages in the development of phonological underspecification based on the current data from children with PD (possibly representing an earlier stage of development), children with TD, and adults. That is, the data suggest a developmental model of phonological underspecification (**Figure 10**).

Within the proposed developmental model, the first stage of phonological underspecification occurs when the child is developing and defining the phonological representation of the underspecified phoneme, /d/ (**Figure 10—I**). Since underspecified phonemes are considered to be less complex than specified phonemes, it is proposed that the features of the underspecified phonemes are developed prior to those of the specified phonemes. Thus, in this stage, the features of the specified phoneme may not yet be defined. In the case of /b/ and /d/, [coronal] is defined and temporarily specified within the /d/ representation, while the [labial] place of articulation feature is not yet a part of the /b/ phonological representation. This means that when the underspecified phoneme, /d/, is the standard, it sets up a [coronal] feature expectation. When /b/ is the deviant, its representation does not contain this feature; as such, this feature contrast would be considered a mismatch. Alternatively, when the /b/ is the standard, it does not specify a place of articulation feature. As a result, even though /d/ had a contrasting feature, this contrast would be considered a no-mismatch, resulting in a small or no mismatch response. In the present study, the children with PD demonstrated this response pattern, as a positive mismatch response was evident in the /bɑ/ identity difference waves, while no mismatch response was present in the /dɑ/ difference waves.

The model proposes that the second stage of phonological underspecification occurs when the child develops and defines the phonological representation of the more specified phoneme (**Figure 10—II**). The features of the more complex specified phoneme, /b/, are now phonologically defined in the phonological representation, while the features of the underspecified phoneme continue to be temporarily specified. Thus, in the case of /b/ and /d/, the [labial] and [coronal] places of articulation have now been phonologically assigned to their respective phonemes. With both phonemes having fully specified phonological representations, specific features are expected in both oddball conditions. When the /b/ is the standard, [labial] is expected; when /d/ is the standard, [coronal] is expected. Thus, when either phoneme is the deviant stimulus, the feature contrast would be a mismatch. As a result, both the specified and underspecified phonemes are predicted to elicit mismatch responses—no asymmetrical response pattern should occur. The TD children in the present study demonstrated this response pattern, as mismatch responses were present in both the /bɑ/ and /dɑ/ identity difference waves.

The final stage of the proposed developmental model of phonological underspecification occurs when only the features of the specified phoneme are defined in the phonological representation, while the underspecified phoneme's features are no longer specified, but considered to be the default (**Figure 10—III**). This stage assumes that over time children learn which phoneme features are phonologically defined within the representation, and which ones are default features. In this stage, [coronal] is assumed to be the default place of articulation unless other evidence is provided. This means that [labial] is specified for /b/, while /d/ does not have a specified place of articulation. In terms of the oddball paradigm, the specified phoneme then sets up a specific feature expectation, which in this case is [labial]. When the underspecified phoneme, /d/, is the deviant, the feature contrast results in a mismatch. Alternatively, when the underspecified phoneme is the standard, its lack of a place of articulation creates no feature expectation. As a result, even though the specified phoneme has a contrasting feature, a no-mismatch occurs when it is the deviant—resulting in a small or no mismatch response. The adults in Cummings et al. (2017) followed this response pattern.

This proposed developmental model is consistent with previous proposals of children's speech processing (Bernhardt, 1992; Kuhl, 2000). Recall that Bernhardt (1992) proposed that children first develop and define underspecified features and then gradually add more specified features. This claim is supported by the data of the children with PD. The children with PD appeared to develop the underspecified [coronal] feature prior to that of the more specified [labial]. That is, while children had phonologically assigned the [coronal] feature to /d/, [labial] was not yet defined in the phonological representation of /b/. This suggests that children have access to more phonological information about /d/ than /b/ early in development. This acquisition pattern could be the explanation for why the children with PD demonstrated exactly opposite MMN patterns as those of adults. That is, the children had the [coronal] feature defined in their phonological representations while adults' representations contained [labial].

Moreover, data from the TD children suggest that children first develop and phonologically define all the features of /b/ and /d/ prior to refining their phonological representations through the elimination of the redundant default [coronal] feature. This phonological organization of the /b/ and /d/ representations likely resulted in mismatch responses for both phonemes. These findings suggest that the children had not yet learned that certain default underspecified features, such as [coronal], do not need to be assigned to phonological representations. Thus, the composition of phonological representations appears to develop and change over time.

In sum, the proposed developmental model of phonological underspecification is meant to provide a framework within which to examine the universality of the theory in children. The particular predictions of underspecification will need to be extensively tested in developmental populations to provide converging, or diverging, evidence for the proposal put forth here. Longitudinal and/or cross-sectional studies could clarify how phonological representations are established in development.

A Developmental Trajectory for Mismatch Responses

Two types of mismatch responses, negative (MMN) and positive (PMR), were observed in the present study. In the TD children, an MMN was present in the /bɑ/ identity difference waves while a PMR was present in the /dɑ/ difference waves. Moreover, a PMR was observed in the /bɑ/ difference waves of the children with PD. While the present study considered both the MMN and PMR as characterizing mismatch responses within the FUL framework,

TABLE 2 | Single-subject results from the four LIMO GLM analyses.

Child	Age	Group	/bɑ/ Standards vs. /bɑ/ Deviants		/dɑ/ Standards vs. /dɑ/ Deviants		/bɑ/ Standards vs. /dɑ/ Standards		/bɑ/ Deviants vs. /dɑ/ Deviants	
			FCz	5 elect.	FCz	5 elect.	FCz	5 elect.	FCz	5 elect.
1	5.00	TD			X	X	X	X	X	X
2	6.33	TD	X	X				X		X
3	6.33	TD		X	X	X	X	X		X
4	6.92	TD		X				X		X
5	6.67	TD				X	X	X		
6	5.00	TD		X		X	X	X		X
7	4.33	TD		X		X	X	X		X
8	6.17	TD	X	X		X		X	X	X
9	4.92	TD				X		X		
10	6.67	TD				X	X	X	X	X
11	6.75	TD	X			X		X		X
12	4.58	TD		X		X		X		X
13	6.92	PD		X		X		X		X
14	5.33	PD		X		X				
15	5.50	PD				X		X		
16	4.00	PD			X	X		X		X
17	5.08	PD	X	X	X	X		X		X
18	5.83	PD	X	X				X		X
19	6.33	PD		X		X		X		
20	4.83	PD	X	X		X	X	X	X	X
21	6.83	PD		X		X		X		X
22	5.17	PD		X		X		X		X
23	4.00	PD	X	X	X	X		X	X	X
24	6.58	PD						X		X

Each participant's data were examined for a significant (though uncorrected for multiple comparisons) phoneme type or trial type difference of at least 20 continuous milliseconds at electrode FCz, and across at least five separate electrode sites (that did not have to include FCz), between 200 and 400 ms post-syllable onset—the time window of the mismatch response. An "X" identifies participants who showed a significant effect for each measurement.

the polarity differences of the mismatch responses suggest a developmental trajectory, with children first demonstrating a PMR that gradually shifts in polarity to an MMN.

The polarity differences in the mismatch response provide additional evidence of a developmental trajectory of phonological underspecification in children. That is, feature contrasts that are acquired first elicit the PMR. Once the features are more established, the feature contrasts elicit the MMN. For example, the children with PD only demonstrated a PMR, suggesting early-developing phonological knowledge pertaining to just the less complex, underspecified phoneme. Conversely, the TD children demonstrated both positive and negative mismatch responses. The MMN was associated with the underspecified phoneme and the PMR was associated with the specified phoneme. In this situation, the TD children's knowledge of the underspecified phoneme was more extensive, leading to the MMN. Thus, they may have still been in the process of developing and defining the features of the more specified phoneme, resulting in the PMR. Finally, the more specified phoneme elicited the MMN in adults (Cummings et al., 2017). It is assumed that the adults had adequate and extensive knowledge of the phoneme, which elicited the negative mismatch response. Thus, the longer a feature had been assigned phonological meaning and defined within a phonological representation, the more negative its mismatch response. Mismatch polarity could characterize the depth and detail of phonological knowledge.

Typically Developing vs. Disordered Phonological Systems

As discussed above, the TD children and children with PD were likely following the same developmental trajectory of phonological underspecification, but were functioning at different stages. That is, the TD children demonstrated developmentally more mature responses than the children with PD, based on the pattern and polarity of their mismatch responses. Thus, children with PD do not appear to have the same extent of phonological knowledge as their same-aged, TD peers. This finding is consistent with previous evidence that language disordered populations demonstrate different mismatch responses as compared to their TD peers. For example, as compared to their TD peers, PMR responses have often been reported in children who have or are at-risk-for dyslexia (Volkmer and Schulte-Körne, 2018).

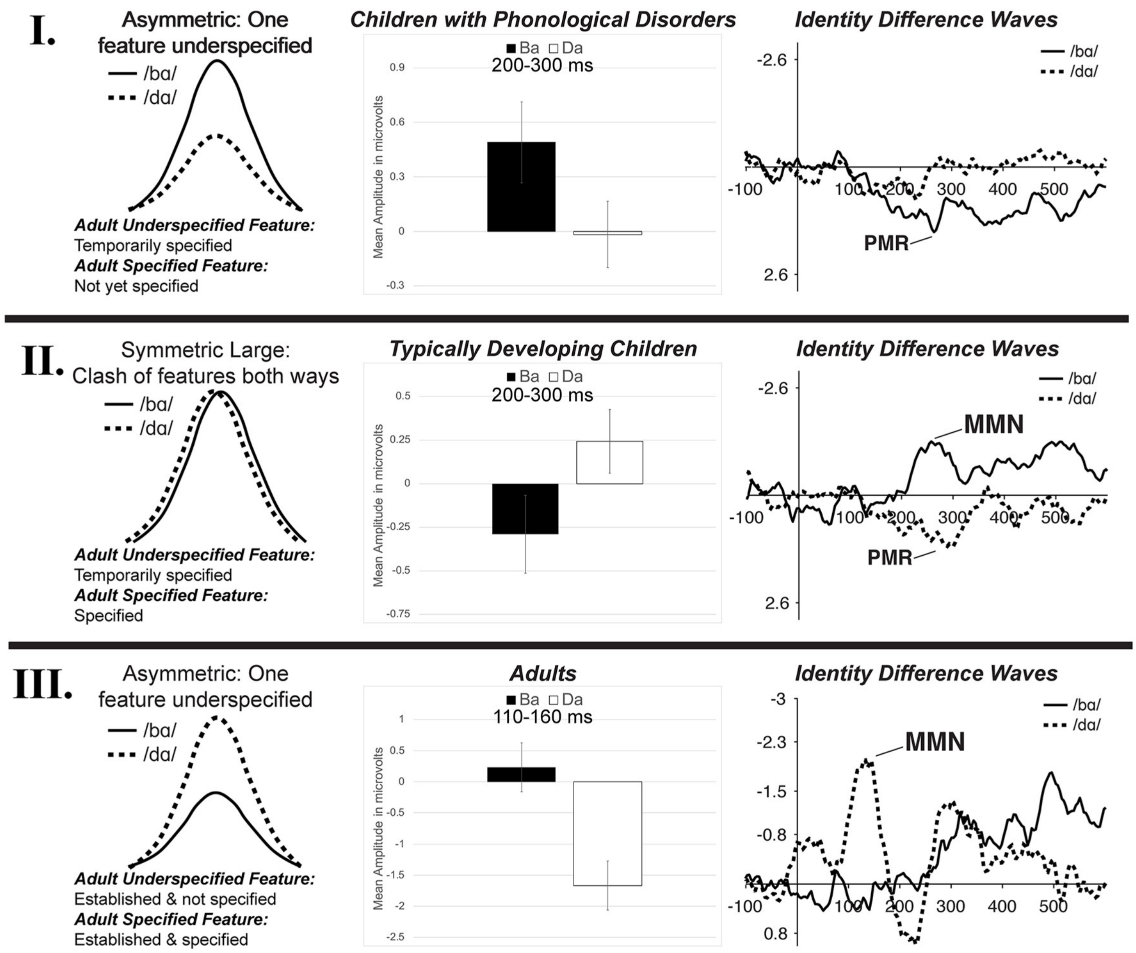


FIGURE 10 | The three stages of the proposed developmental model of phonological underspecification. Children with PD are hypothesized to be in **Stage I**, TD children are hypothesized to be in **Stage II**, while adults are presumed to be in **Stage III**. The left column represents a schematic representation of possible mismatch responses in each stage (based on Schluter et al., 2016). The proposed development and specification of features is listed, with /d/ presumed to be the underspecified phoneme and /b/ the specified phoneme. The middle column displays the mean amplitude of the mismatch responses found in the /bɑ/ and /dɑ/ identity difference waves of each population during the mismatch analysis window. The right column presents the identity differences waves of each population from electrode FC1. See text for more detail.

Overall, the processing differences in the two groups primarily centered around the more specified sound, /b/. Based on their mismatch response pattern, it would appear that the children with PD did not have the specified place of articulation feature [labial] sufficiently developed and defined within their phonological representations while the TD children did. Alternatively, the children with PD did not differ from the TD children in their standard and deviant ERP responses to /dɑ/. This finding suggests that there is an early processing advantage for less specified phonemes. Even if default features are initially stored in the phonological representations, those features may be easier to acquire than specified features.

This study provides evidence that the entire phonological system is not impaired in children with PD, as the underspecified /d/ phoneme elicited neural responses similar to those of TD children. This may be surprising, given their significant differences on the standardized speech articulation and phonological assessments. If children produce speech by accessing their underlying phonological representations, it would be assumed that children with large numbers of speech production errors have incorrect representations. However, it is important to remember that the children in both groups could accurately produce /d/, as evidenced by their test scores. Thus, it is possible that the phonological representations of /d/ were similar in both groups of children. Conversely, the data suggest that the representations of /b/ were not the same in the two groups of children, even though the children could also accurately produce that phoneme. As no previous study has examined the neural indices of speech processing in children with PD, it is presently unknown how children with PD would respond to phonemes they could not produce correctly. It is possible that an incorrect or extremely sparse phonological representation of affected phonemes is an underlying mechanism of the speech production errors observed in children with PD.

This work also has important potential clinical implications. It is nearly impossible for speech-language pathologists (SLPs) to predict how well a child might perform in treatment, as outcomes are varied. While behavioral predictors of speech treatment outcomes have not yet been identified, it is possible that the neural patterns children demonstrate prior to beginning treatment might indicate how well they will learn to produce a treated sound. That is, ERP responses could be indices of children's speech production ability and/or speech treatment effectiveness. Specifically, the present evidence suggests that children with PD might be recruiting developmentally immature neural networks for the processing of speech sounds, which might not allow for full and accurate processing and discrimination of phonological information. SLPs could use such information to design intervention programs that target not only speech production, but general phonological knowledge and/or speech perception skills, which could lead to better overall intervention outcomes.

Limitations

The present study was the first to examine neural indices of phonological underspecification in children, both with typically developing and disordered phonological systems. Specifically, the present study focused on preschool-aged children between 4- and 6-years of age, as that is the age in which the highest percentage of children are diagnosed with PD (Shriberg et al., 1999; Law et al., 2000). As such, it is likely we missed the earliest stages of phonological development whose precursors are present in the infant speech perception work (Kuhl, 2000; Werker and Hensch, 2015). It would be useful to examine underspecification in younger children to see if TD children demonstrate Stage I at an earlier age than children with PD, and to see if there could be an even earlier stage that we could not identify with our present population and age groups. Moreover, examining underspecification in older children who have theoretically acquired all of their speech sounds (McLeod and Crowe, 2018; Crowe and McLeod, 2020) could provide information about how and when adult-like phonological knowledge is acquired.

Individual differences within and across groups are an inherent confounds when working with children in general, and disordered populations in particular. For example, while the TD children's mismatch responses to /bɑ/ and /dɑ/ did not reliably differ, a strong trend was observed. That this trend did not reach significance suggests that there might have been some variability in the TD sample. Moreover, while all children met the basic criterion to be included in one group or the other, there was still a range of severity of speech production difficulty in the children with PD. The single-trial analyses (**Supplementary Figures 1–6**) show that participants demonstrated a wide range of responses to the stimuli. It is possible that there are subtypes of PD, and different neural response patterns could be used to identify them. However, much larger groups of children would be necessary to address this issue. Future studies with new and/or larger groups of participants can provide converging evidence for the underspecification evidence provided here.

Identity difference waves were included to control for basic differences in acoustic detail present in the /bɑ/ and /dɑ/ stimuli. Still, it is still possible that the physical acoustic differences of /bɑ/ and /dɑ/ alone were responsible for the observed MMN response differences (Näätänen et al., 2007). However, from a sonority standpoint (Clements, 1990), the sonority difference between /b/ and /ɑ/ is acoustically the same as that of /d/ and /ɑ/. Thus, neither consonant establishes a stronger syllable onset than the other; they are acoustically functioning at a similar level. The frequency of occurrence, or phonotactic probability, of phoneme combinations could have also affected the MMN responses (Bonte et al., 2005; Näätänen et al., 2007). However, the phonotactic probability (Vitevitch and Luce, 2004) of the single phonemes /b/ and /d/ were nearly identical (0.0512 and 0.0518, respectively) and the phonotactic probability of the syllables were quite similar (0.0039 and 0.0023, respectively). Thus, it does not seem that the frequency with which children encounter the phonemes and syllable combinations in the ambient language were driving the response differences.

It is possible that the general acoustic perceptual ability was different in the two groups of children. For example, while the left inferior frontal gyrus (IFG) has been associated with articulatory-based speech codes (Poeppel et al., 2008), it has been suggested that atypical right hemisphere (i.e., IFG) processing may impact phonological processing (Goswami, 2011). When looking at the PD children's responses in **Figure 9**, many of the phoneme response differences in Panels A and C are only found in the right hemisphere electrodes, which is not the case for the TD children in Panel C. Moreover, the timing of the stimulus differences in Panels B and C are different for the two groups of children, with the children with PD demonstrating much earlier stimulus differences. These earlier differences could be due more to acoustic-level processing, rather than phonological-level processing. This could indicate that these children are attending more to the acoustic differences of the stimuli, rather than the more relevant phonological feature information of the phonemes. Thus, there is some evidence that children with PD recruit atypical neural networks during speech processing tasks.

As this is the first study to address neural indices of phonological underspecification in children, caution should be taken to avoid over-interpretation. Although it is possible that the results could be due to the acoustic differences between the stimuli, the findings are consistent with an underspecification model. Future studies could address how the effects of acoustic differences can be distinguished from the effects of phonological underspecification in typical and atypical phonological development. This could be accomplished by comparing ERP responses elicited by phonemes to pure tones, or other non-linguistic stimuli, in TD children and children with PD.

Conclusion

FUL predicts that [coronal] phonemes have the default place of articulation because they contain less distinctive feature information in their phonological representations than phonemes with other places of articulation. However, these language universal underspecification claims had not been tested in developmental populations until now. Neither the

TD children, nor the children with PD, demonstrated the FUL-predicted mismatch asymmetry patterns seen in adult data. In fact, the children with PD demonstrated the exact opposite mismatch response pattern from that of adults, while TD children demonstrated mismatch responses to both specified and unspecified phonemes. Moreover, while both groups of children demonstrated similar responses to the underspecified /dɑ/, their neural responses to the more specified /bɑ/ varied. Thus, the children with PD did not appear to have the same level of phonetic information, or specification, in their phonological representations as TD children. These findings were interpreted within a proposed developmental model of phonological underspecification, wherein children with PD are functioning at a developmentally less mature stage of phonological acquisition than their same-aged TD peers. Thus, phonological specification, and underspecification, are phenomena that that likely develop over time with experience and exposure to language.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Idaho State University Human Subjects Committee and the University of North Dakota Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AC created the stimuli, tested participants, prepared and analyzed the data, and helped write the manuscript. DO and YW helped analyze the data and write the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by NIH grant R15DC013359 (from the National Institute on Deafness and Other Communication Disorders) awarded to the first author. The third author was supported, in part, by NSF grant 1540943. Open access publication funds were provided by Idaho State University as part of the first author's start-up package.

ACKNOWLEDGMENTS

We would like to thank Janet Hallgrimson for her assistance during the testing of participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.580697/full#supplementary-material

REFERENCES

Ahmmed, A. U., Clarke, E. M., and Adams, C. (2008). Mismatch negativity and frequency representational width in children with specific language impairment. *Dev. Med. Child Neurol.* 50, 938–944. doi: 10.1111/j.1469-8749.2008.03093.x

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

Aslin, R. N., and Smith, L. B. (1988). Perceptual development. *Annu. Rev. Psychol.* 39, 435–473. doi: 10.1146/annurev.ps.39.020188.002251

Austin, D., and Shriberg, L. D. (1997). *Lifespan Reference Data for Ten Measures of Articulation Competence Using the Speech Disorders Classification System (SDSC)*. Madison, WI: Phonology Project Waisman Center on Mental Retardation and Human Development.

Bernhardt, B. (1992). The application of nonlinear phonological theory to intervention with one phonologically disordered child. *Clin. Linguist. Phon.* 6, 283–316. doi: 10.3109/02699209208985537

Boada, R., and Pennington, B. F. (2006). Deficient implicit phonological representations in children with dyslexia. *J. Exp. Child Psychol.* 95, 153–193. doi: 10.1016/j.jecp.2006.04.003

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., and Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clin. Neurophysiol.* 116, 2765–2774. doi: 10.1016/j.clinph.2005.08.012

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253

Cabbage, K. L., Farquharson, K., Iuzzini-Seigel, J., Zuk, J., and Hogan, T. P. (2018). Exploring the overlap between dyslexia and speech sound production deficits. *Lang. Speech Hear. Serv. Sch.* 49, 774–786. doi: 10.1044/2018_LSHSS-DYSLC-18-0008

Carr, K. W., Tierney, A., White-Schwoch, T., and Kraus, N. (2016). Intertrial auditory neural stability supports beat synchronization in preschoolers. *Dev. Cogn. Neurosci.* 17, 76–82. doi: 10.1016/j.dcn.2015.12.003

Chaney, C. (1988). Identification of correct and misarticulated semivowels. *J. Speech Hear. Disord.* 53, 252–261. doi: 10.1044/jshd.5303.252

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* 1, 351–353. doi: 10.1038/1561

Cheour, M., Ceponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., et al. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scand. J. Psychol.* 43, 33–39. doi: 10.1111/1467-9450.00266

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper & Row.

Clements, G. (1990). "The role of the sonority cycle in core syllabification," in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech,* eds. J. Kingston and M. Beckman (Cambridge: Cambridge University Press), 283–333. doi: 10.1017/CBO9780511627736.017

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

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol.* 39, 757–772. doi: 10.1037/a0030862

Crowe, K., and McLeod, S. (2020). Children's english consonant acquisition in the United States: a review. *Am. J. Speech Lang. Pathol.* 29, 2155–2169. doi: 10.1044/2020_AJSLP-19-00168

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

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: discrimination of the /r–l/ distinction by young infants. *Percept. Psychophys.* 18, 341–347. doi: 10.3758/BF03211210

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. doi: 10.1126/science.171.3968.303

Elbro, C., and Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scand. J. Psychol.* 46, 375–384. doi: 10.1111/j.1467-9450.2005.00468.x

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Froud, K., and Khamis-Dakwar, R. (2012). Mismatch negativity responses in children with a diagnosis of childhood apraxia of speech (CAS). *Am. J. Speech Lang. Pathol.* 21, 302–312. doi: 10.1044/1058-0360(2012/11-0003)

Goldman, R., and Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation (3rd ed.)*. Minneapolis, MN: Pearson.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10. doi: 10.1016/j.tics.2010.10.001

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

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

International Expert Panel on Multilingual Children's Speech (2012). *Multilingual Children with Speech Sound Disorders: Position paper. Research Institute for Professional Practice, Learning and Education (RIPPLE).* Bathurst, NSW: Charles Sturt University.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111, 1745–1758. doi: 10.1016/S1388-2457(00)00386-2

Khan, L., and Lewis, N. (2015). *Khan-Lewis Phonological Analysis (3rd ed.)*. Minneapolis, MN: Pearson.

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonol. Yearbook* 2, 85–138. doi: 10.1017/S0952675700000397

Korpilahti, P., and Lang, H. A. (1994). Auditory ERP components and mismatch negativity in dysphasic children. *Electroencephalogr. Clin. Neurophysiol.* 91, 256–264. doi: 10.1016/0013-4694(94)90189-9

Kral, A., and Eggermont, J. J. (2007). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res. Rev.* 56, 259–269. doi: 10.1016/j.brainresrev.2007.07.021

Kraus, N. (2001). Auditory pathway encoding and neural plasticity in children with learning problems. *Audiol. Neuro Otol.* 6, 221–227. doi: 10.1159/000046837

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., and Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273, 971–973. doi: 10.1126/science.273.5277.971

Kraus, N., McGee, T. J., and Koch, D. B. (1998). Speech sound representation, perception, and plasticity: a neurophysiologic perceptive. *Audiol. Neuro Otol.* 3, 168–182. doi: 10.1159/000013788

Kuhl, P. K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11850–11857. doi: 10.1073/pnas.97.22.11850

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron* 67, 713–727. doi: 10.1016/j.neuron.2010.08.038

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364

Lahiri, A., and Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245–294. doi: 10.1016/0010-0277(91)90008-R

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phon.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Labphon 7,* eds. C. Gussenhoven and N. Warner (Berlin: Mouton), 637–676. doi: 10.1515/9783110197105.637

Law, J., Boyle, J., Harris, F., Harkness, A., and Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *Int. J. Lang. Commun. Disord.* 35, 165–188. doi: 10.1080/136828200247133

Linnavalli, T., Putkinen, V., Huotilainen, M., and Tervaniemi, M. (2017). Phoneme processing skills are reflected in children's MMN responses. *Neuropsychologia* 101, 76–84. doi: 10.1016/j.neuropsychologia.2017.05.013

Macken, M. A. (1980). The child's lexical representation: the ' puzzle-puddle-pickle' evidence. *J. Linguist.* 16, 1–17. doi: 10.1017/S0022226700006307

Manly, B. F. J. (2006). *Randomization, Bootstrap and Monte Carlo Methods in Biology, Third Edition*. Boca Raton, FL: CRC Press.

Maurer, U., Bucher, K., Brem, S., and Brandeis, D. (2003). Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. *Clin. Neurophysiol.* 114, 808–817. doi: 10.1016/S1388-2457(03)00032-4

McGregor, K. K., and Schwartz, R. G. (1992). Converging evidence for underlying phonological representation in a child who misarticulates. *J. Speech Hear. Res.* 35, 596–603. doi: 10.1044/jshr.3503.596

McLeod, S., and Crowe, K. (2018). Children's consonant acquisition in 27 languages: a cross-linguistic review. *Am. J. Speech Lang. Pathol.* 27, 1546–1571. doi: 10.1044/2018_AJSLP-17-0100

McLeod, S., and Baker, E. (2017). *Children's Speech: An Evidence-Based Approach to Assessment and Intervention*. New York, NY: Pearson.

Mohanan, K. P. (1991). On the bases of radical underspecification. *Nat. Lang. Linguistic Theory* 9, 285–325. doi: 10.1007/BF00134678

Mueller, J. L., Friederici, A. D., and Männel, C. (2012). Auditory perception at the root of language learning. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15953–15958. doi: 10.1073/pnas.1204319109

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, 1–21. doi: 10.1111/1469-8986.3810001

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Näätänen, R., Teder, W., Alho, K., and Lavikainen, J. (1992). Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport* 3, 493–496. doi: 10.1097/00001756-199206000-00009

Nittrouer, S., and Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *J. Acoust. Soc. Am.* 101, 2253–2266. doi: 10.1121/1.418207

Palmer, J. A., Makeig, S., Kreutz-Delgado, K., and Rao, B. D. (2008). "Newton method for the ICA mixture model," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing,* (IEEE: Las Vegas), 1805–1808. doi: 10.1109/ICASSP.2008.4517982

Pennington, B. F., and Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annu. Rev. Psychol.* 60, 283–306. doi: 10.1146/annurev.psych.60.110707.163548

Pernet, C. R., Chauveau, N., Gaspar, C., and Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Comput. Intell. Neurosci.* 2011:831409. doi: 10.1155/2011/831409

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1071–1086. doi: 10.1098/rstb.2007.2160

Ponton, C. W., Eggermont, J. J., Khosla, D., Kwong, B., and Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin. Neurophysiol.* 113, 407–420. doi: 10.1016/S1388-2457(01)00733-7

Ponton, C. W., Moore, J. K., and Eggermont, J. J. (1999). Prolonged deafness limits auditory system developmental plasticity: evidence from an evoked potentials study in children with cochlear implants. *Scand. Audiol.* 51, 13–22.

Rinker, T., Kohls, G., Richter, C., Maas, V., Schulz, E., and Schecker, M. (2007). Abnormal frequency discrimination in children with SLI as indexed by mismatch negativity (MMN). *Neurosci. Lett.* 413, 99–104. doi: 10.1016/j.neulet.2006.11.033

Robbins, J., and Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *J. Speech Hear. Disord.* 52, 271–277. doi: 10.1044/jshd.5203.271

Rousselet, G. A., Gaspar, C. M., Wieczorek, K. P., and Pernet, C. R. (2011). Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Front. Psychol.* 2:137. doi: 10.3389/fpsyg.2011.00137

Rvachew, S., and Jamieson, D. (1995). "Learning new speech contrasts: evidence from adults learning a second language and children with speech disorders," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* ed. W. Strange (Timonium, MD: York Press), 411–432.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., and Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PLoS ONE* 7:e40953. doi: 10.1371/journal.pone.0040953

Scharinger, M., and Lahiri, A. (2010). Height differences in english dialects: consequences for processing and representation. *Lang. Speech 53*, 245–272. doi: 10.1177/0023830909357154

Scharinger, M., Merickel, J., Riley, J., and Idsardi, W. J. (2011). Neuromagnetic evidence for a featural distinction of english consonants: sensor- and source-space data. *Brain Lang.* 116, 71–82. doi: 10.1016/j.bandl.2010.11.002

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of english fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Scott, S. K. (2012). The neurobiology of speech perception and production—can functional imaging tell us anything we did not already know? *J. Commun. Disord.* 45, 419–425. doi: 10.1016/j.jcomdis.2012.06.007

Scott, S. K., and Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45. doi: 10.1016/j.cognition.2002.12.002

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., and Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *J. Cogn. Neurosci.* 17, 1168–1180. doi: 10.1162/0898929054475217

Sharma, M., Purdy, S. C., Newall, P., Wheldall, K., Beaman, R., and Dillon, H. (2006). Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder. *Clin. Neurophysiol.* 117, 1130–1144. doi: 10.1016/j.clinph.2006.02.001

Shriberg, L. D., and Kwiatkowski, J. (1982). Phonological disorders III: a procedure for assessing severity of involvement. *J. Speech Hear. Disord.* 47:256. doi: 10.1044/jshd.4703.256

Shriberg, L. D., Tomblin, J. B., and McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *J. Speech Lang. Hear. Res.* 42, 1461–1481. doi: 10.1044/jslhr.4206.1461

Steriade, D. (1995). "Underspecification and markedness," in *The Handbook of Phonological Theory,* ed. J. A. Goldsmith (New York, NY: Blackwell Publishing), 114–174. doi: 10.1111/b.9780631201267.1996.00006.x

Uwer, R., Albrecht, R., and von Suchodoletz, W. (2002). Automatic processing of tones and speech stimuli in children with specific language impairment. *Dev. Med. Child Neurol.* 44, 527–532. doi: 10.1111/j.1469-8749.2002.tb00324.x

Vitevitch, M. S., and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behav. Res. Methods Instr. Comput.* 36, 481–487. doi: 10.3758/BF03195594

Volkmer, S., and Schulte-Körne, G. (2018). Cortical responses to tone and phoneme mismatch as a predictor of dyslexia? A systematic review. *Schizophr. Res.* 191, 148–160. doi: 10.1016/j.schres.2017.07.010

Werker, J. F., and Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annu. Rev. Psychol.* 66, 173–196. doi: 10.1146/annurev-psych-010814-015104

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing, 2nd ed*. New York, NY: Elsevier Academic Press.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn. Brain Res.* 7, 357–369. doi: 10.1016/S0926-6410(98)00039-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 17 February 2021
doi: 10.3389/fnhum.2021.585817

Phonological Underspecification: An Explanation for How a Rake Can Become Awake

Alycia E. Cummings[1]*, *Ying C. Wu*[2] *and Diane A. Ogiela*[1]

[1]*Department of Communication Sciences and Disorders, Idaho State University, Meridian, ID, United States,* [2]*Swartz Center for Computational Neuroscience, University of California, San Diego, San Diego, CA, United States*

Neural markers, such as the mismatch negativity (MMN), have been used to examine the phonological underspecification of English feature contrasts using the Featurally Underspecified Lexicon (FUL) model. However, neural indices have not been examined within the approximant phoneme class, even though there is evidence suggesting processing asymmetries between liquid (e.g., /ɹ/) and glide (e.g., /w/) phonemes. The goal of this study was to determine whether glide phonemes elicit electrophysiological asymmetries related to [consonantal] underspecification when contrasted with liquid phonemes in adult English speakers. Specifically, /ɹɑ/ is categorized as [+consonantal] while /wɑ/ is not specified [i.e., (–consonantal)]. Following the FUL framework, if /w/ is less specified than /ɹ/, the former phoneme should elicit a larger MMN response than the latter phoneme. Fifteen English-speaking adults were presented with two syllables, /ɹɑ/ and /wɑ/, in an event-related potential (ERP) oddball paradigm in which both syllables served as the standard and deviant stimulus in opposite stimulus sets. Three types of analyses were used: (1) traditional mean amplitude measurements; (2) cluster-based permutation analyses; and (3) event-related spectral perturbation (ERSP) analyses. The less specified /wɑ/ elicited a large MMN, while a much smaller MMN was elicited by the more specified /ɹɑ/. In the standard and deviant ERP waveforms, /wɑ/ elicited a significantly larger negative response than did /ɹɑ/. Theta activity elicited by /ɹɑ/ was significantly greater than that elicited by /wɑ/ in the 100–300 ms time window. Also, low gamma activation was significantly lower for /ɹɑ/ vs. /wɑ/ deviants over the left hemisphere, as compared to the right, in the 100–150 ms window. These outcomes suggest that the [consonantal] feature follows the underspecification predictions of FUL previously tested with the place of articulation and voicing features. Thus, this study provides new evidence for phonological underspecification. Moreover, as neural oscillation patterns have not previously been discussed in the underspecification literature, the ERSP analyses identified potential new indices of phonological underspecification.

Keywords: ERP, EEG, underspecification, MMN, ERSP, theta, gamma, phonology

Edited by:
Arild Hestvik,
University of Delaware, United States

Reviewed by:
Wolfram Ziegler,
Ludwig Maximilian University of Munich, Germany
Karthik Durvasula,
Michigan State University, United States

Correspondence:
Alycia E. Cummings
cummalyc@isu.edu

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *21 July 2020*
Accepted: *25 January 2021*
Published: *17 February 2021*

Citation:
Cummings AE, Wu YC and Ogiela DA (2021) Phonological Underspecification: An Explanation for How a Rake Can Become Awake. Front. Hum. Neurosci. 15:585817. doi: 10.3389/fnhum.2021.585817

INTRODUCTION

Distinctive features are often described as the functional units of phonological systems (Chomsky and Halle, 1968). Phonemes are composed of combinations of features, with each phoneme being distinguished from all other phonemes by at least one feature. Phonological underspecification theories propose that only the distinctive features that differentiate a phoneme are present in the adult phonological representation (Kiparsky, 1985; Archangeli, 1988; Mohanan, 1991; Steriade, 1995). Specifically, underspecification identifies some features as "default" and others as "marked." Default features are not stored within the phonological representation because they are assumed to be predictable by phonological rule. Conversely, marked features are the contrastive, or not otherwise predictable, phonological information that must be specified and stored. A marked phoneme is presumed to require the storage of more distinctive features in its phonological representation as compared to an unmarked phoneme. Thus, marked phonemes are considered to be more phonologically specified than unmarked phonemes.

By only storing specified features within the phonological representation, underspecification can improve speech processing efficiency when encountering the wide variability present in natural speech (Chomsky and Halle, 1968; Eulitz and Lahiri, 2004). Indeed, evidence for the effectiveness of phonological underspecification can be found in speech production. For example, phonological code retrieval in adults is slower when naming words beginning with marked phonemes, such as /ɹ/, as compared to unmarked phonemes, such as /b/ (Cummings et al., 2016).

The application of underspecification is also often observed in speech production errors. Specifically, speech errors typically affect specified features and phonemes rather than underspecified features and phonemes (Fromkin, 1973; Levelt et al., 1999; Brown, 2004). For example, approximants are involved in a common phonological process, called liquid gliding, found in the productions of both typically developing children and those with speech sound disorders (Shriberg, 1980; Broen et al., 1983). That is, many young English-speaking children incorrectly produce pre-vocalic /ɹ/ as [w] (e.g., 'rake' is pronounced as 'wake'); however, children rarely, if ever, produce /w/ as [ɹ][1]. Thus, during typical and atypical development, children tend to incorrectly produce phonemes with specified features (Stoel-Gammon and Dunn, 1985; Grunwell, 1987). Such evidence suggests that the underlying phonological representations can affect speech production. A better understanding of how specified and underspecified features are stored within phonological representations has important clinical implications for speech-language pathologists working with clients who have speech production errors. Due to the high frequency of the liquid gliding phonological process in pediatric American English-speaking populations, the examination of the /ɹ/-/w/ contrast is of particular interest.

As underlying phonological representations cannot be easily, if at all, accessed behaviorally, neuroimaging tools have proven useful in examining phonological underspecification. Neural markers of phonological underspecification have primarily been examined using the framework established by the Featurally Underspecified Lexicon (FUL) model (Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002, 2010). Phonological underspecification has been found in vowels (Diesch and Luce, 1997; Eulitz and Lahiri, 2004; Cornell et al., 2011; Scharinger et al., 2012), as well as in consonants such as stops (Cummings et al., 2017), nasals (Cornell et al., 2013), and fricatives (Schluter et al., 2016). Many of these studies have indexed underspecification using the mismatch negativity (MMN), which is a well-studied event-related potential (ERP) peak that is elicited by auditory oddballs elicited within a stream of standard stimuli (Näätänen and Winkler, 1999; Picton et al., 2000; Näätänen et al., 2007). The MMN is a neurophysiological index of auditory change detection. As the deviant oddball becomes more different from the standard, MMN amplitude increases and latency decreases. Thus, the timing and size of the MMN may reflect the amount of perceived difference between the standard and the deviant stimuli (Tiitinen et al., 1994; Näätänen et al., 1997).

Within the FUL framework, the size of the MMN depends on the degree of specification of the features extracted from the stimuli (Winkler et al., 1999; Eulitz and Lahiri, 2004; Scharinger et al., 2011). For example, a true mismatch occurs when the more specified sound is the standard and the less specified sound is the deviant in the MMN oddball paradigm. In this situation, large MMN responses are elicited by the less specified deviant sound because it violates the feature expectations established by the standard. Conversely, a no-mismatch occurs when the less specified sound serves as the standard and the more specified sound is the deviant. In this context, no conflict between the phonetic features is identified because the feature was not specified by the standard. Thus, a very small, or no, MMN is elicited. Because of the predicted size differences of the MMN responses, the true mismatch contrast could be considered an easier feature comparison to make than the no-mismatch feature comparison.

Neural indices of underspecification have not been examined within the English approximant phoneme class, even though there is evidence suggesting processing asymmetries exist between liquid (e.g., /ɹ/) and glide (e.g., /w/) phonemes (Greenberg, 1975; Shriberg, 1980; Edwards, 1983; Clements, 1990). These asymmetries suggest that /ɹ/ and /w/ might differ in how they are stored within a phonological representation. While /ɹ/ and /w/ share several distinctive features (Chomsky and Halle, 1968), liquid phonemes are specified as [+consonantal] while glide phonemes are considered semi-vowels and are not specified for that feature (i.e., [−consonantal]. The basic definition of [consonantal] is: "... sounds [are] produced with a radical obstruction in the midsagittal region of the vocal tract; nonconsonantal sounds are produced without such an obstruction." (Chomsky and Halle, 1968, p. 302). That is, [consonantal] phonemes are produced with varying amounts of constriction created by the labial, coronal, and/or dorsal articulators in the oral cavity. This feature

[1]The /ɹ/ is generally considered to be more complex and specified than /w/ (Greenberg, 1975) as it is acquired later in development (McLeod and Crowe, 2018) and occurs in fewer world languages than /w/ (Maddieson, 1984).

classification essentially places vowels, glides, and laryngeal consonants in one natural sound class: [−consonantal], while the other consonant phonemes, including /ɹ/, are in a separate sound class: [+consonantal]. Thus, glide phonemes can be considered underspecified for [consonantal] in comparison to liquid phonemes.

While [consonantal] never functions as the sole feature responsible for distinguishing phonemes (Hume and Odden, 1996), it is hypothesized that constriction is the primary distinguishing feature of /ɹ/ and /w/, at least in American English. There are many ways that the American pre-vocalic /ɹ/ can be produced (Preston et al., 2020), with the /ɹ/ productions broadly described as either being "retroflex" or "bunched" in nature. Regardless of the type of production used, two separate constrictions are necessary for /ɹ/ to be produced: palatal constriction and pharyngeal constriction (Delattre and Freeman, 1968; Gick, 1999; Secord et al., 2007). The palatal constriction is made with the dorsum of the tongue being brought near the soft palate while the pharyngeal constriction is achieved with tongue root retraction. Indeed, problems with vocal tract constriction, and arguably the application of [consonantal], are often observed in children with speech sound disorders as they often have difficulty achieving adequate palatal and pharyngeal constriction necessary for an "accurate" /ɹ/ production. That is, they produce /ɹ/ with /w/-level constriction, which is not enough, either in terms of the amount and/or place of constriction.

Previous studies examining underspecification within the FUL-MMN paradigm relied on strict superset-subset relationships between the specified and underspecified features [e.g., contrasting voiced stops differing only by place of articulation: (coronal) vs. (labial); Cummings et al., 2017]. Such a contrast is not available for /ɹ/ and /w/ because they vary both in terms of manner and place of articulation. Thus, a strict application of FUL cannot be applied to identify the underspecification differences in /ɹ/ and /w/. Nevertheless, there are other potential ways to identify phonological underspecification in this contrast, which could then be tested using FUL-based predictions in an MMN paradigm.

Feature Geometry is an alternative way of organizing features in a hierarchical relationship that reflects the configuration of the vocal tract and articulators in a tree diagram. It allows for broad feature groupings (e.g., manner and place of articulation) to be associated with individual features. This means that some features, such as those at the root node (e.g., consonantal, sonorant) dominate place (e.g., coronal, labial) nodes[2] (Bernhardt and Stoel-Gammon, 1994; Clements and Hume, 1995; Halle et al., 2000; Lahiri and Reetz, 2002). It is assumed that the determination of features in the higher nodes of the tree will impact the features available at lower nodes in the tree. Thus, the idea of markedness is present in both feature geometry and underspecification theory.

Given the hypothesis that constriction is the distinguishing articulatory property of /w/ and /ɹ/, the feature geometry theory of Clements and Hume (1995) was used (**Figure 1**). The Clements and Hume (1995) model is a constriction-based approach that defines most phonemes in terms of their constriction location and degree. This means that the place features (i.e., the articulators and dependents) define the constriction location while the articulator-free features define constriction degree (i.e., consonantal/vocoid, sonorant, approximant, and continuant). Three major class features are located at the root node: [sonorant], [approximant], and [vocoid]. As [vocoid] is the terminal opposite of [consonantal], we will refer to /w/ as [−consonantal] ([+vocoid]) and /ɹ/ as [+consonantal] ([-vocoid]). These distinct [+consonantal] and [−consonantal] designations place /ɹ/ on the C-place tier and /w/ on the V-place tier, respectively. Both phonemes are [+sonorant], [+approximant], and [+continuant].

In this model, the place nodes for vowels and consonants are on separate tiers, designated *V*-place and *C*-place, respectively, with the vocalic node linking under the *C*-place node. The actual constriction location (i.e., place of articulation) is largely the same for both vowels and consonants: [labial], [coronal], and [dorsal]. As a result, consonant and vowel articulators are placed on the same tier. In addition, [coronal] has two dependents: [anterior] and [distributed]. This means that coronal itself is not the terminal place of articulation—[anterior] or [distributed] is; conversely, [labial] and [dorsal] are terminal. This feature tree organization leads to /w/ being characterized as [−consonantal, labial] while /ɹ/ contains the features [+consonantal, coronal: +distributed]. With [coronal: +distributed] being located lower on the feature tree than [labial], /ɹ/ more specified for the place of articulation than /w/. Thus, following Clements and Hume (1995), as compared to /w/, /ɹ/ is more specified both in terms of the manner of articulation [+consonantal] and place of articulation [coronal: +distributed].

In regards to /ɹ/ and /w/, the feature [consonantal] ([vocoid]) is located on the highest node of the Clements and Hume (1995) tree (**Figure 1**). As such, processing this feature should dominate the processing of features at lower nodes, including the place of articulation nodes (i.e., *C*-Place and *V*-Place). That is, feature geometry theory predicts that the presence or absence of the [consonantal] feature will be the relevant contrasting feature of /w/ and /ɹ/. While the Clements and Hume (1995) model does not have the same organization as FUL, both can identify features and phonemes that are less specified than others. It is assumed that if underspecification is a language universal phenomenon, the underspecification-specific MMN predictions of FUL would hold regardless of whether FUL was strictly adhered to, or if another theoretical interpretation of underspecification was used. Thus, the Clements and Hume (1995) model was employed as the framework for the [consonantal] underspecification of /w/, as compared to /ɹ/. This prediction was then tested in the present study using the FUL-based predictions in an MMN paradigm.

Feature geometry can also provide a framework to explain children's acquisition of phonemes and speech production errors. That is, higher and dominant nodes in the hierarchy are proposed to be acquired before subordinate nodes (Bernhardt, 1992; Core, 1997). Moreover, default features would be acquired early in development, with minimal specification present in the phonological representations (Bernhardt and Gilbert, 1992).

[2]There is some debate as to the organization of the tree, but all theories agree upon a root node, and separate class nodes.

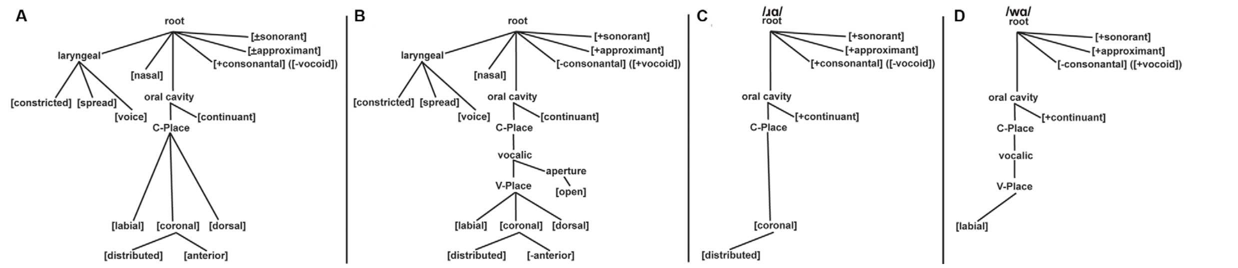


FIGURE 1 | Feature geometry trees based on Clements and Hume (1995). Panel **(A)** displays the full feature geometry tree for consonants. Panel **(B)** displays the full feature geometry tree for vocoids (i.e., vowels and glides). Panel **(C)** displays the features of /ɹ/: [+consonantal] ([-vocoid]), [+sonorant], [+approximant], [+continuant], and [coronal: +distributed]. Panel **(D)** displays the features of /w/: [–consonantal] ([+vocoid]), [+sonorant], [+approximant], [+continuant], and [labial].

The liquid gliding phonological process could then be explained by the early acquisition of /w/, which is not specified for [consonantal] and is the default feature. Only after the [consonantal] feature of /ɹ/ is fully established in the phonological representation is the gliding pattern suppressed in children's production. As the basic definition of [consonantal] suggests that articulatory precision (i.e., constriction control) is necessary, it seems logical that an underspecified [–consonantal] phoneme (i.e., /w/) would be acquired prior to a [+consonantal] phoneme (i.e., /ɹ/). Thus, there is speech production evidence for the underspecification of [consonantal] in typical and atypical development.

There is clear evidence from developmental and clinical (i.e., disordered speech) data that there is a relationship between /w/ and /ɹ/ in American English, with young children and children with speech disorders substituting [w] for /ɹ/. Moreover, when adults mimic the speech of young children, they almost always substitute [w] for /ɹ/. Thus, the liquid gliding phonological process is an arguably ingrained stereotype of young children's speech—even for adults who have essentially no explicit knowledge of the phonological system. Given these observations, it was hypothesized that /w/ contains one or more default features leading to its common usage in development while /ɹ/ contains one or more specified features that limits its production early in development. The purpose of the study was to address this potential underspecified/specified feature relationship in adults before examining the neural processing patterns in children. Thus, this study aims to determine whether glide phonemes elicit [consonantal] underspecification-related electrophysiological asymmetries when contrasted with liquid phonemes in adult English speakers.

Following FUL's predictions and framework, if /w/ is less specified than /ɹ/ in terms of the manner of articulation, the former phoneme should elicit a larger MMN response than the latter phoneme. That is, a standard stream of /w/ phonemes would not set expectations for [consonantal], so when a deviant /ɹ/ is presented, it would be a no-mismatch. Thus, a small, or no, MMN response is predicted to occur in the no-mismatch situation. Conversely, hearing /ɹ/ as the standard stimulus would set up the expectation for [+consonantal], which would be violated by a deviant /w/. Thus, a large MMN response is predicted to occur in the true mismatch situation.

While underspecification has primarily been addressed with ERPs, subtle processing differences between distinct phonemes may not be detected due to the averaging of brain signals in traditional ERP methods. In contrast, time-frequency analyses provide an alternative approach that involves decomposing the spectral power of the EEG signal over time (Davidson and Indefrey, 2007; Cohen, 2014). Unlike ERPs, which only reveal phase-locked changes in the time series data, this approach affords both a view of changes in EEG signals that are phase-locked to stimulus onset (evoked responses), as well as a view of changes that are not phase-locked (induced responses). The synchronization of neuronal cell assemblies proposed to underlie increases in induced power has been hypothesized to mediate the binding of perceptual information (Singer and Gray, 1995). Experimental results have also implicated induced responses in various cognitive functions such as working memory (Gurariy et al., 2016) and attentional processes (Ward, 2003).

In keeping with these findings, there is reason to believe that phonological underspecification could also be indicated by neural oscillation patterns. For example, cortical oscillations in the theta (∼4–7 Hz) and low gamma (∼25–35 Hz) bands have been implicated in decoding syllabic and phonemic segments, respectively, from continuous speech (Luo and Poeppel, 2007; Ghitza, 2011; Giraud and Poeppel, 2012; Doelling et al., 2014; Di Liberto et al., 2015). That is, theta band has been proposed to represent higher-order syllable-level processing while low gamma band activities have been linked to phoneme feature-level processing (e.g., formant transitions, voicing). Possibly, one, or both, of these bands could demonstrate underspecification response asymmetries. As neural oscillation patterns underlying phonological underspecification have not previously been examined, this work was exploratory in nature and no specific hypotheses were proposed regarding theta and low gamma response patterns.

MATERIALS AND METHODS

Participants

Fifteen native speakers of (American) English (three males, 12 female; mean age: 21.71 years, range: 19–26 years) who were undergraduate students participated in the study. All of them had

TABLE 1 | The phonotactic probability in English of the phonemes and syllables used in the study.

	Consonant	Consonant + /ɑ/
/ɹ/	0.0501	0.0011
/w/	0.0203	0.0008
/b/	0.0512	0.0039
/d/	0.0518	0.0023

a normal or corrected-to-normal vision, and none had a history of speech, language, and/or hearing impairment. This study was approved by the university institutional review board and each participant signed informed consent following the university human research protection program.

Stimuli

Syllables (consonant + /ɑ/) were pronounced by a male North American English speaker. The syllables were digitally recorded in a sound isolated room (Industrial Acoustics Company, Inc., Winchester, UK) using a Beyer Dynamic (Heilbronn, Germany) Soundstar MK II unidirectional dynamic microphone and Behringer (Willich, Germany) Eurorack MX602A mixer. All syllables were digitized with a 16-bit AD converter at a 44.1 kHz sampling rate. The average intensity of all the syllable stimuli was normalized to 65 dB SPL.

The adults heard two oddball stimulus sets, each containing the same four English speech consonant-vowel (CV) syllables: "ra" (/ɹɑ/), "wa" (/wɑ/), "ba" (/bɑ/), and "da" (/dɑ/). In one stimulus set, /ɹɑ/ served as the standard syllable, with the other three CV syllables serving as deviants. In the second stimulus set, /wɑ/ served as the standard syllable, with the other three syllables being deviants. Only responses to the /ɹɑ/ and /wɑ/ syllables will be addressed further since they served as both standard and deviant stimuli, which allowed for the creation of same-stimulus identity difference waves. Since /bɑ/ and /dɑ/ deviants were incorporated to prevent MMN habituation, they were not examined. As initially recorded, the syllables varied slightly in duration, due to the individual phonetic make-up of each consonant. Syllable duration was minimally modified in /wɑ/ (by shortening the steady-state vowel duration by 24 ms) so that all syllables were 375 ms in length. Each syllable token used in the study was correctly identified by at least 15 adult listeners.

The phonotactic probability[3] of each phoneme and syllable were calculated using the online phonotactic probability calculator[4] (Vitevitch and Luce, 2004). These probability values are presented in **Table 1**. The singleton /ɹ/ occurs 2.5 times more frequently than /w/ in English. Similarly, the ɹɑ/ syllable in English occurs 1.375 times more frequently than that of /wɑ/.

Stimulus Presentation

The stimuli were presented in blocks containing 237 standard stimuli and 63 deviant stimuli (21 per deviant), with five blocks of each stimulus set being presented to each participant (10 total blocks). The stimulus sets were presented sequentially in the session, with all five blocks of one stimulus set (e.g., /ɹɑ/ standard set) being presented before the other stimulus set (e.g., /wɑ/ standard set); the presentation of the stimulus sets was counterbalanced across participants. Each block lasted approximately 6 min and the participants were given a break between blocks when necessary. Within the block, the four stimuli were presented using an oddball paradigm in which the three deviant stimuli (probability = 0.07 for each) were presented in a series of standard stimuli (probability = 0.79). Stimuli were presented in a pseudorandom sequence and the onset-to-onset inter-stimulus interval varied randomly between 600 and 800 ms. The syllables were delivered by stimulus presentation software (Presentation software, www.neurobs.com). The syllable sounds were played *via* two loudspeakers situated 30 degrees to the right and left from the midline 120 cm in front of a participant, which allowed the sounds to be perceived as emanating from the midline space. The participants sat in a sound-treated room and watched a silent cartoon video of their choice. The recording of the ERPs took approximately 1 h.

EEG Recording and Averaging

Sixty-six channels of continuous EEG (DC-128 Hz) were recorded using an ActiveTwo data acquisition system (Biosemi, Inc, Amsterdam, Netherlands) at a sampling rate of 256 Hz. This system provides "active" EEG amplification at the scalp that substantially minimizes movement artifacts. The amplifier gain on this system is fixed, allowing ample input range (−264 to 264 mV) on a wide dynamic range (110 dB) Delta- Sigma ($\Delta\Sigma$) 24-bit AD converter. Sixty-four channel scalp data were recorded using electrodes mounted in a stretchy cap according to the International 10-20 system. Two additional electrodes were placed on the right and left mastoids. Eye movements were monitored using FP1/FP2 (blinks) and F7/F8 channels (lateral movements, saccades). During data acquisition, all channels were referenced to the system's internal loop (CMS/DRL sensors located in the centro-parietal region), which drives the average potential of a subject (the Common Mode voltage) as close as possible to the Analog-Digital Converter reference voltage (the amplifier "zero"). The DC offsets were kept below 25 microvolts at all channels. Off-line, data were re-referenced to the common average of the 64 scalp electrode tracings.

Data processing followed an EEGLAB (Delorme and Makeig, 2004) pipeline. Briefly, data were high-pass filtered at 0.5 Hz using a pass-band filter. Line noise was removed using the CleanLine EEGLAB plugin. Bad channels were rejected using the trimOutlier EEGLAB plugin and the removed channels were interpolated. Source level contributions to channel EEG were decomposed using Adaptive Mixed Model Independent Component Analysis (AMICA; Palmer et al., 2008) in EEGLAB[5]. Artifactual independent components (ICs) were identified by their activation patterns, scalp topographies, and power spectra, and the contribution of these components to the channel EEG was zeroed (Jung et al., 2000; Delorme and Makeig, 2004). Epochs containing 100 ms pre-auditory stimulus to 800 ms

[3]Phonotactic probability refers to the frequency with which a phonological segment, such as /ɹ/, and a sequence of phonological segments, such as /ɹɑ/, occur in a given position in a word (Jusczyk et al., 1994).

[4]https://calculator.ku.edu/phonotactic/about

[5]http://www.sccn.ucsd.edu/eeglab

post-auditory stimulus time were baseline-corrected for the pre-stimulus interval and averaged by stimulus type. On average, individual data contained 804 (SD = 84) /ɹɑ/ standard syllable epochs (i.e., trials), 794 (SD = 79) /wɑ/ standard syllable epochs, 96 (SD = 9) /ɹɑ/ deviant syllable epochs, and 97 (SD = 9) /wɑ/ deviant syllable epochs.

ERP and EEG Measurements

Three different data analysis strategies were used in the present study: (1) traditional mean amplitude repeated measure ANOVA analyses using averaged data; (2) cluster-based permutation analyses of averaged data (Bullmore et al., 1999; Groppe et al., 2011); and (3) event-related spectral perturbation (ERSP) analyses (Makeig, 1993).

Mean Amplitude Measurements of Averaged Data

The dual stimulus set nature of the present study allowed for the creation of "same-stimulus", or identity, difference waveforms. These difference waves were created by subtracting the ERP response of a stimulus serving as the standard from that of the same stimulus serving as the deviant, across stimulus sets. For example, the ERP response for /ɹɑ/ as the standard was subtracted from the ERP response for /ɹɑ/ as the deviant (of the reversed stimulus set; Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013). The creation of identity difference waveforms eliminates the potential confound that may result from acoustic stimulus differences since the same stimulus is used to elicit both the standard and deviant responses. The waveforms were visually inspected from 0 to 400 ms, with the MMN appearing between approximately 100 and 250 ms post-syllable onset.

Since the MMN was present in 12 electrodes centered around the scalp midline (Fz, F1/F2, FCz, FC1/FC2, Cz, C1/C2, CPz, and CP1/CP2)[6], these electrodes were selected for the mean amplitude analyses. The MMN elicited by /wɑ/ extended for approximately 150 ms. Given the extended duration of the MMN, the mean amplitude measurement of the MMN was split into three 50 ms windows: 100–150, 150–200, and 200–250 ms post-stimulus onset. Phonological underspecification in the identity difference waves was analyzed separately in each time window using a Phoneme Type (/ɹɑ/, /wɑ/) × Anterior-Posterior (4 Levels) × Left-Right (3 Levels) repeated measure ANOVA.

Since the difference waves were generated from the standard and deviant syllable ERPs, the mean amplitude measurements of the standard and deviant waveforms were taken from the same three time windows as that of the MMN: 100–150, 150–200, and 200–250 ms post-syllable onset. In terms of ERP waveform morphology, these measurements approximately captured the auditory N1 (100–200 ms) and auditory P2 (200–250 ms). Phonological underspecification in these ERPs was analyzed separately for each time window using a Phoneme Type (/ɹɑ/, /wɑ/) × Trial Type (Standard, Deviant) × Anterior-Posterior (4 Levels) × Left-Right (3 Levels) repeated measure ANOVA. Partial eta squared (η^2) effect sizes are reported for all significant effects and interactions. When applicable, Geiser–Greenhouse corrected p-values are reported.

Cluster Mass Permutation Tests of Averaged Data

The ERPs were submitted to repeated measures two-tailed cluster-based permutation tests (Bullmore et al., 1999; Groppe et al., 2011) using the Mass Univariate ERP Toolbox for EEGLAB[7]. Four tests were conducted: (1) /ɹɑ/ standard vs. /ɹɑ/ deviant ERPs; (2) /wɑ/ standard vs. /wɑ/ deviant ERPs; (3) /ɹɑ/ vs. /wɑ/ standard ERPs; and (4) /ɹɑ/ vs. /wɑ/ deviant ERPs. Each test included the same 12 electrodes from the mean amplitude ERP measurements: F1/F2, Fz, FC1/FC2, FCz, C1/C2, Cz, CP1/CP2, and CPz. All of the time points (measured every 4 ms; 155 total time points) between 0 and 600 ms at the 12 scalp electrodes were included in the test (i.e., 1,860 total comparisons).

T-tests were performed for each comparison using the original data and 2,500 random within-participant permutations of the data. For each permutation, all t-scores corresponding to uncorrected p-values of 0.05 or less were formed into clusters. Electrodes within about 5.44 cm of one another were considered spatial neighbors, and adjacent time points were considered temporal neighbors. The sum of the t-scores in each cluster was the "mass" of that cluster. The most extreme cluster mass in each of the 2,501 sets of tests was recorded and used to estimate the distribution of the null hypothesis (i.e., no difference between conditions). The permutation cluster mass percentile ranking of each cluster from the observed data was used to derive p-values assigned to each member of the cluster. t-scores that were not included in a cluster were given a p-value of 1.

Event-Related Spectral Perturbation (ERSP) Analyses

ERSP analyses were performed to examine theta (4–7 Hz) and low gamma (25–35 Hz) band activities elicited by the /ɹɑ/ and /wɑ/ standard and deviant syllable stimuli. This approach was informed by prior work on speech syllable decoding (Ghitza, 2011; Giraud and Poeppel, 2012). ERSPs were computed from time-series data from 16 electrodes: F3/F4, F1/F2, FC3/FC4, FC1/FC2, C3/C4, C1/C2, CP3/CP4, CP1/CP2[8] (**Supplementary Figure 1**). Data were epoched from −0.6 ms before stimulus onset to 1.6 ms after. Estimates of spectral power for each of these EEG epochs were computed across 200 equally spaced time points along 100 frequency steps spanning 3–50 Hz using Morlet wavelets with cycles gradually increasing with frequency (Delorme and Makeig, 2004). ERSPs were created by converting spectral density estimates to log power, averaging across single trials, and subtracting the mean log power derived from the pre-stimulus baseline period of the same trials. The final output for each channel was a matrix of 100 frequency values (3–50 Hz) by 200 time points (−0.5 to 1 s).

[6] These electrodes encompass four anterior-posterior levels (Frontal, Frontal-Central, Central, Central-Parietal) and three left-right laterality levels [Left (1), Midline (z), Right (2)].

[7] https://openwetware.org/wiki/Mass_UnivariateERPToolbox.

[8] These electrodes encompassed two laterality levels (Left: F3, F1, FC3, FC1, C3, C1, CP3, CP1; Right: F4, F2, FC4, FC2, C4, C2, CP4, CP2), four anterior-posterior levels (Frontal, Frontal-Central, Central, Central-Parietal) and two electrode laterality levels (Far laterality: F3/F4, FC3/FC4, C3/C4, CP3/CP4; Close laterality: F1/F2, FC1/FC2, C1/C2, CP1/CP2).

It has been proposed that the decoding of auditory information during speech perception occurs during two distinct time scales—one which relates to syllable-level processing (∼200 ms) and one related to phoneme-level processing (∼25 ms; Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012; Doelling et al., 2014). As such, theta (4–7 Hz) bandwidth responses were measured in one 200 ms window occurring 100–300 ms post-syllable onset. Low gamma (25–35 Hz) bandwidth responses were measured separately in five 50 ms windows occurring 50–300 ms post-syllable onset.

For each participant, the magnitude of synchronized theta and gamma activity at each electrode was derived by averaging estimates of spectral power computed across steps within each of these bandwidths and across time points within the selected time interval. Phoneme-related differences in theta and low gamma power were examined in separate Phoneme Type (/ɹɑ/, /wɑ/) × Trial Type (Standard, Deviant) × Laterality (Left, Right) × Anterior-Posterior (4) × Electrode Laterality (Far, Close) repeated measure ANOVAs.

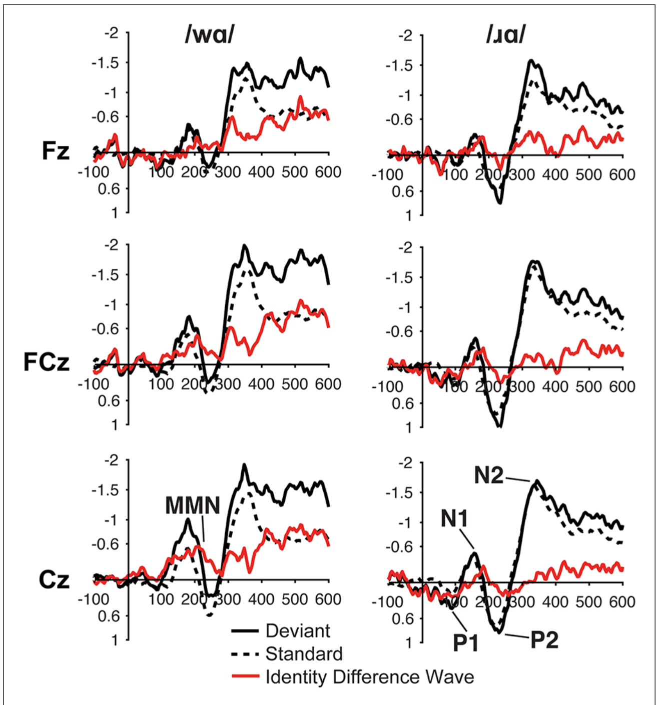


FIGURE 2 | Event-related potential (ERP) waveforms elicited by the /wɑ/ (left side) and /ɹɑ/ (right side) syllables in the ERP study. The deviant waveforms represent the neural responses when the deviant syllable was presented within a stream of the opposite syllable standards. Subtracting the standard syllable response from the deviant syllable response resulted in the identity difference waves. Note that negative is plotted *up* in all waveforms.

RESULTS

Only significant results for all analyses are reported.

ERP Mean Amplitude Results

For both the /ɹɑ/ and /wɑ/ syllables, the ERP waveforms elicited by the standard and deviant stimuli consisted of P1 at ca. 75 ms, N1 at ca. 150 ms, P2 at ca. 225 ms, and N2 at ca. 350 ms (**Figure 2**). In the same-stimulus identity difference waves, an MMN was visible in both the /ɹɑ/ and /wɑ/ identity waveforms at ca. 200 ms; the /wɑ/ MMN extended from ca. 100–250 while the /ɹɑ/ MMN extended from ca. 175–225 ms (**Figure 2**).

Identity Difference Waves: MMN

Individual participants' mean ERP responses to /ɹɑ/ and /wɑ/ stimuli are presented in **Supplementary Figure 2**. During the 200–250 time window, MMN responses elicited by /wɑ/ were significantly more negative than those elicited by /ɹɑ/ ($F_{(1,14)} = 5.479$, $p < 0.04$, $\eta^2 = 0.281$; **Figures 2, 3**). During the 150–200 ms time window, the overall magnitude of the MMN was larger over the left hemisphere, as compared to the right ($F_{(2,28)} = 5.343$, $p < 0.02$, $\eta^2 = 0.276$; **Supplementary Figure 3**).

Standard and Deviant Waveforms

The standard and deviant ERP responses elicited by /wɑ/ were significantly more negative than those elicited by /ɹɑ/ during both the 150–200 ms time window ($F_{(1,14)} = 12.448$, $p < 0.004$, $\eta^2 = 0.471$) and 200–250 ms time window ($F_{(1,14)} = 21.272$, $p < 0.0001$, $\eta^2 = 0.603$; **Figure 3** and **Supplementary Figure 4**). Deviant trials elicited significantly more negative responses than did standard responses during the 150–200 ms time window ($F_{(1,14)} = 10.029$, $p < 0.008$, $\eta^2 = 0.417$).

A phoneme × trial type interaction ($F_{(1,14)} = 5.481$, $p < 0.04$, $\eta^2 = 0.281$) was observed during the 200–250 ms time window (**Figure 3**). Whereas the /ɹɑ/ standard and /ɹɑ/ deviant responses did not reliably differ, ERPs elicited by /wɑ/ deviants were consistently more negative than the /wɑ/ standards ($F_{(1,14)} = 14.189$, $p < 0.003$, $\eta^2 = 0.503$).

ERP Summary

The FUL underspecification paradigm predicts that the underspecified phoneme deviant presented within a stream of the specified phoneme standards will elicit a large MMN response, as this situation creates a true feature mismatch context. The opposite stimulus presentation is predicted to elicit a small, or no, MMN response due to the feature no-match context. These hypotheses were supported. The underspecified /wɑ/ stimuli elicited significantly larger and more negative responses than did the specified /ɹɑ/.

Cluster Permutation Analysis Results

Four cluster-level mass permutation tests encompassing 0–600 ms were applied to the standard and deviant syllable data. The results of the tests are displayed in raster diagrams in **Figures 4A–C**.

No reliable clusters were identified when examining the difference between /ɹɑ/ standards and /ɹɑ/ deviants. On the other hand, one broadly distributed cluster extending from 132 to 600 ms signified a period during which the /wɑ/ deviants elicited more negative ERP responses than the /wɑ/ standards; the smallest significant *t*-score (in absolute values) was: $t_{(14)} = -2.159$, $p < 0.0001$ (**Figure 4A**).

When contrasting the standard syllables, two broadly distributed clusters extending from 175 to 230 ms and 242 to 343 ms signified two time periods during which the /ɹɑ/ standards differed from the /wɑ/ standards (**Figure 4B**); the

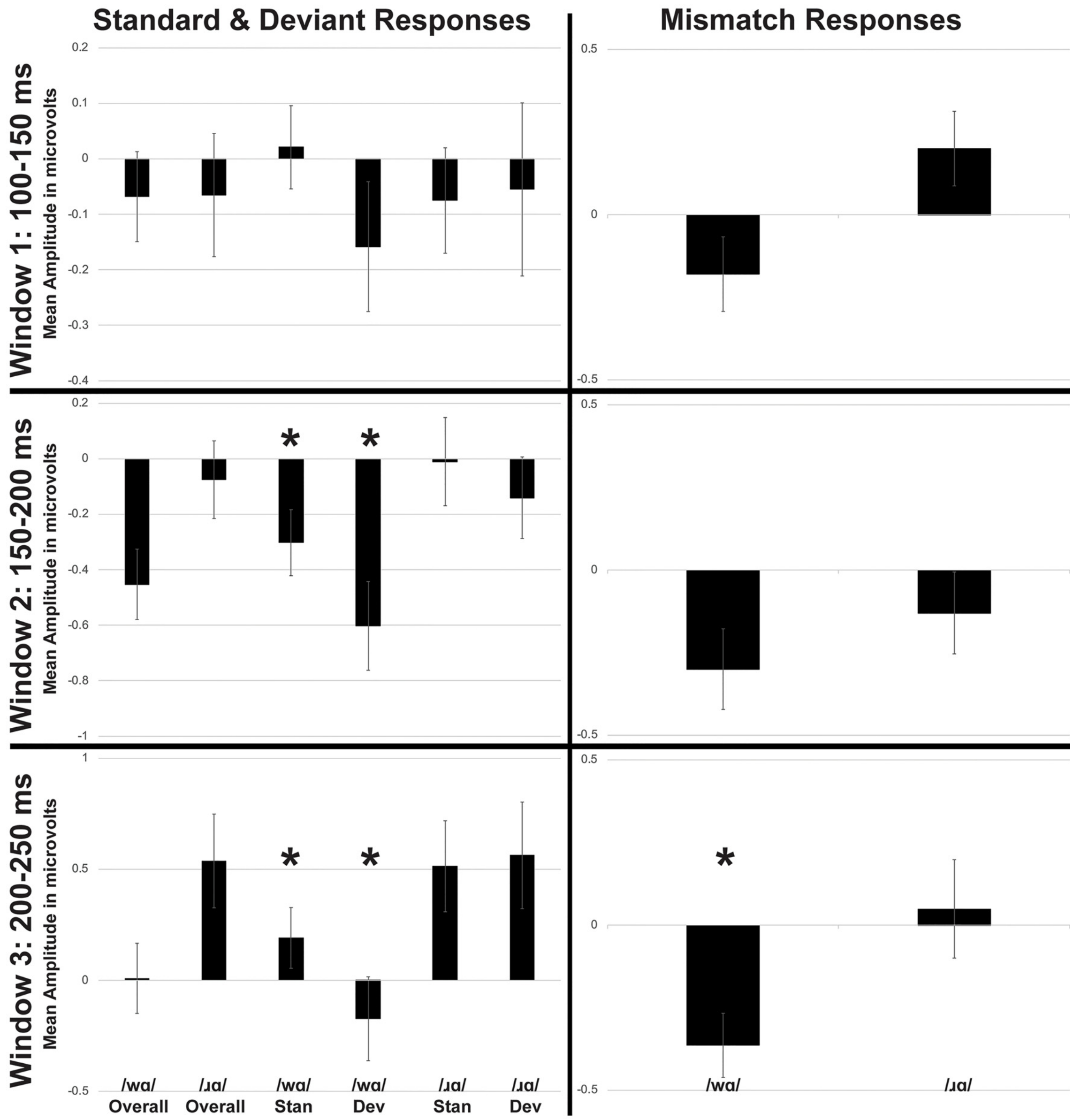


FIGURE 3 | Average mean amplitudes for standard and deviant ERPs (left side) and mismatch responses measured in identity difference waves (right side) across three time windows: (1) 100–150 ms post-syllable onset; (2) 150–200 ms post-syllable onset; and (3) 200–250 ms post-syllable onset. Error bars represent SEM. Time Windows 1 and 2 broadly captured the auditory N1 response, while Time Window 3 captured the auditory P2 response. The Mismatch Negativity (MMN) was present in all three time windows. The /wɑ/ deviants were significantly more negative than the /wɑ/ standards during Window 3. The MMN responses elicited by /wɑ/ were significantly more negative than those elicited by /ɹɑ/ during Window 3. *Significant effects.

smallest significant t-score was: $t_{(14)} = 2.149$, $p < 0.05$. When the /ɹɑ/ and /wɑ/ deviant syllables were contrasted, a broadly distributed cluster extending from 136 to 253 ms signified a time period during which the /wɑ/ deviants elicited more negative (i.e., larger) ERP responses than the /ɹɑ/ deviants (**Figure 4C**); the smallest significant t-score was: $t_{(14)} = 2.158$, $p < 0.005$.

Cluster Permutation Analysis Summary

FUL predicts a larger MMN will be elicited by an underspecified phoneme, as compared to a specified phoneme. Consistent with the ERP analyses, this prediction was confirmed. The MMN appeared in the difference waveforms between 100 and 300 ms post-syllable onset. The effects seen in the /wɑ/ stimuli extended far beyond the traditional timeline of the MMN[9]. This result was unexpected. As no phoneme type differences were observed in the standard trial and deviant trial analyses,

[9]Note that while only the mismatch responses elicited by /wɑ/ were found to be significant *via* the cluster permutation analyses, mismatch responses were also present in the /ɹɑ/ difference wave (**Figure 2**).

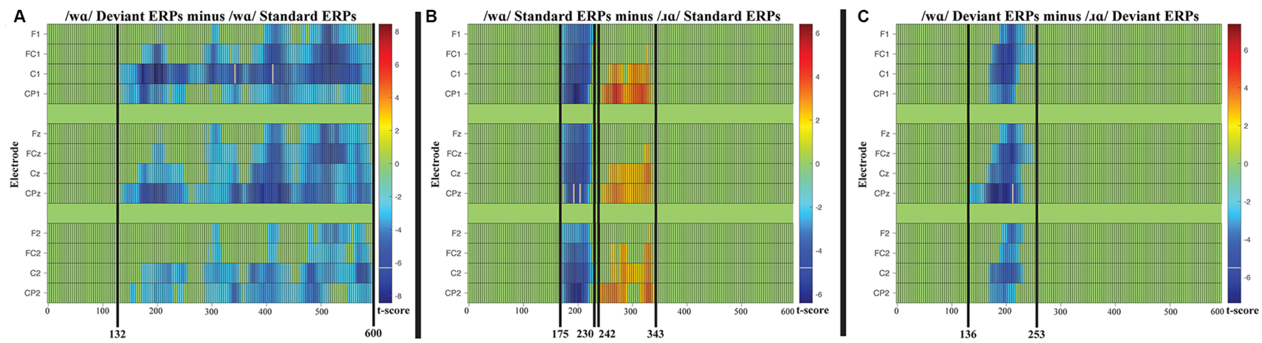


FIGURE 4 | (A) Raster diagram illustrating differences between the /wɑ/ deviants and standards, which extended from 132 ms post-syllable onset to the end of the analysis window (600 ms). **(B)** Raster diagram illustrating differences between /wɑ/ standards and /ɹɑ/ standards, one cluster extended from 175 to 230 ms post-syllable onset and the second cluster extended from 243 to 343 ms post-syllable onset. **(C)** Raster diagram illustrating differences between /wɑ/ deviants and /ɹɑ/ deviants, which extended from 136 to 253 ms post-syllable onset. There were no reliable clusters for the comparison of /ɹɑ/ deviants and standards. Note: for the raster diagrams, colored rectangles indicate electrodes/time points in which the ERPs to one stimulus are significantly different from those to another. The color scale dictates the size of the *t*-test result, with dark red and blue colors being more significant. Green areas indicate electrodes/time points at which no significant differences were found. Note that the electrodes are organized along the *y*-axis somewhat topographically. Electrodes on the left and right sides of the head are grouped on the figure's top and bottom, respectively; midline electrodes are shown in the middle. Within those three groupings, the *y*-axis top-to-bottom corresponds to scalp anterior-to-posterior.

this effect appears to be specific to the contrast of the /wɑ/ standard and deviant trials. Visual analysis of the cluster permutation (**Figure 4A**) suggests that there were potentially three parts to the /wɑ/ effect, ∼132–∼275, ∼300–400, and ∼400–600 ms. Thus, the first part could be attributed to the MMN, the second part could represent the deviance-related or novelty N2 (Folstein and Van Petten, 2008), while the third part could be attributed to a late MMN or Late Negativity (LN)[10]. Previous studies have identified the LN as a secondary index of speech perception and discrimination (Korpilahti et al., 1995; Čeponienė et al., 1998; Cheour et al., 1998; Shafer et al., 2005; Datta et al., 2010; Hestvik and Durvasula, 2016). However, there is currently insufficient information in the underspecification literature to further interpret this finding.

Consistent with the ERP analyses, phoneme type differences in the standard and deviant trials were observed. The two clusters in the analysis of standard trials were consistent with the auditory N1 and P2 ERP responses. That is, in the first period, the /wɑ/ standards elicited a larger auditory N1 than did the /ɹɑ/ standards. During the second period, the /ɹɑ/ standards elicited a larger auditory P2 than did the /wɑ/ standards. Similarly, in the analysis of deviant trials, the identified cluster almost exclusively encompassed the auditory N1 ERP response. As the MMN is derived from the subtraction of the standard stimulus from the deviant stimulus, the responses elicited by /wɑ/ are consistent with the prediction that the underspecified phoneme should elicit larger (i.e., more negative) responses than the more specified /ɹɑ/. Thus, the cluster permutation analyses provide converging evidence for the underspecification of /wɑ/.

[10]Note that while only the mismatch responses elicited by /wɑ/ were found to be significant *via* the cluster permutation analyses, mismatch responses were also present in the /ɹɑ/ difference wave (**Figure 2**).

ERSP Results

Theta Band (4–7 Hz) 100–300 ms

Individual participants' mean theta band responses to /ɹɑ/ and /wɑ/ standards and deviants are presented in **Supplementary Figure 5**. Theta responses elicited by /ɹɑ/ were significantly greater than those elicited by /wɑ/ ($F_{(1,14)}$ = 4.571, p = 0.05, η^2 = 0.246; **Figures 5**, **6**). A significant electrode laterality effect was found ($F_{(1,14)}$ = 14.053, p < 0.003, η^2 = 0.501), as the electrodes closer to midline (1- and 2-level electrodes; M = 0.235, SEM = 0.059) elicited greater theta activity than did the far lateral electrodes (three- and four-level electrodes; M = 0.140, SEM = 0.043).

Low Gamma Band (25–35 Hz) 50–300 ms

Individual participants' mean low gamma band responses to /ɹɑ/ and /wɑ/ standards and deviants are presented in **Supplementary Figure 6**. Low gamma activation varied across variables and time windows (**Figures 5**, **7**). The laterality of low gamma activation patterns changed over time, as significantly less gamma band activation was found across left hemisphere electrodes as compared to right hemisphere electrodes from 50 to 100 ms (Left: M = –0.028, SEM = 0.023; Right: M = 0.011, SEM = 0.016; $F_{(1,14)}$ = 5.042, p < 0.05, η^2 = 0.265) and from 100 to 150 ms (Left: M = –0.042, SEM = 0.028; Right: M = 0.012, SEM = 0.021; $F_{(1,14)}$ = 6.030, p < 0.03, η^2 = 0.301).

The trial type × laterality interaction was significant from 50 to 100 ms ($F_{(1,14)}$ = 6.019, p < 0.03, η^2 = 0.301) and 100 to 150 ms ($F_{(1,14)}$ = 7.589, p < 0.02, η^2 = 0.352). The deviants elicited significantly less gamma activation over the left hemisphere than over the right during the 50–100 ms window ($F_{(1,14)}$ = 6.152, p < 0.03, η^2 = 0.305) and 100–150 ms window ($F_{(1,14)}$ = 7.434, p < 0.02, η^2 = 0.348), while no laterality difference was observed for the standards. This effect was driven primarily by the /ɹɑ/ deviant responses elicited over the left hemisphere. Specifically, low gamma activation elicited by /ɹɑ/ over the left hemisphere

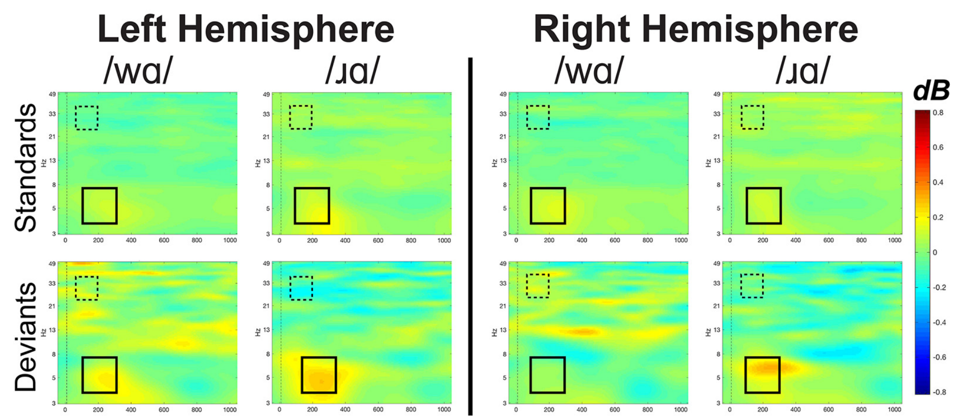


FIGURE 5 | Event-related spectral perturbation (ERSP) activation patterns (in dB) elicited by /wɑ/ and /ɹɑ/ in the standard and deviant stimuli averaged across the eight left hemisphere electrodes and eight right hemisphere electrodes for theta (4–7 Hz) and low gamma (25–35 Hz) bandwidths. Time is on the *x*-axis and frequency is on the *y*-axis. Theta band window of interest is highlighted by the solid black box while the low gamma band window of interest is highlighted by the dashed box. Overall, /ɹɑ/ elicited greater neural synchrony (i.e., more activation) in the theta band than did /wɑ/. The /ɹɑ/ deviant elicited less neural synchrony over the left hemisphere, as compared to the right, in the low gamma band.

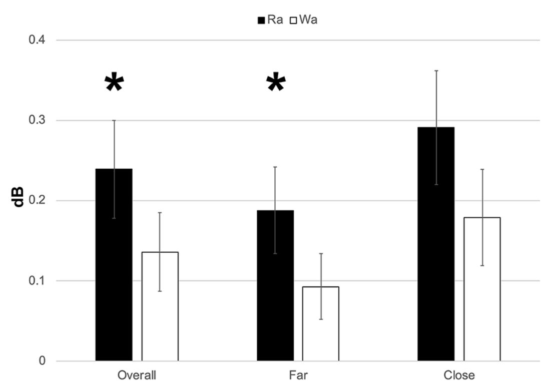


FIGURE 6 | ERSP activation (in dB) elicited by /wɑ/ and /ɹɑ/ for the theta (4–7 Hz) bandwidth in the 100–300 ms time window. The /ɹɑ/ elicited greater neural synchrony (i.e., more activation) in the theta band than did /wɑ/. The electrodes closer to midline (e.g., F1 and F2) elicited greater theta activation than did the electrodes further away from midline (e.g., F3 and F4). *Significant effects.

was significantly less than low gamma activation recorded over the right hemisphere during the 100–150 ms window ($F_{(1,14)} = 5.575$, $p < 0.04$, $\eta^2 = 0.285$; **Figure 7**). No laterality differences were noted for the /wɑ/ deviant responses. Moreover, there was a strong trend for the /ɹɑ/ deviants to elicit less low gamma activation than the /wɑ/ deviants over the left hemisphere during both the 50–100 ms ($F_{(1,14)} = 2.899$, $p < 0.11$, $\eta^2 = 0.172$) and 100–150 ms windows ($F_{(1,14)} = 3.063$, $p < 0.10$, $\eta^2 = 0.180$); no phoneme differences were noted over the right hemisphere.

ERSP Summary

The ERSP analyses were exploratory, as previous underspecification work has not addressed this aspect of phonological processing. Thus, the findings are preliminary. Theta band activation was examined to measure syllable-level processing while the low gamma band was examined to measure phoneme-level processing. At the syllable level, /ɹɑ/ elicited greater theta activation than did /wɑ/. At the phoneme level, low gamma activation was significantly lower for /ɹɑ/ vs. /wɑ/ deviants over the left hemisphere, as compared to the right.

DISCUSSION

This study provides the first neural evidence for [consonantal] underspecification in English-speaking adults. Two phonemes differing in their specification of the [consonantal] feature were contrasted: /ɹ/ and /w/. As /w/ is not specified for [consonantal] while /ɹ/ is, it was hypothesized that asymmetrical speech processing differences would be apparent. Indeed, mean amplitude measurements and cluster permutation analyses both showed that /wɑ/, as an oddball in a sequence of /ɹɑ/, elicited significantly larger MMN responses than did the reciprocal stimulus set—namely, /ɹɑ/ oddballs embedded within frequently occurring instances of /wɑ/. Characterizing the theta and low gamma band neural oscillation patterns provided further evidence for underspecification. The more specified /ɹɑ/ elicited increased activation, or neural synchrony, in the theta bandwidth as compared to /wɑ/. Moreover, the /ɹɑ/ deviants elicited less low gamma activation over the left hemisphere, as compared to the right hemisphere. As neural oscillation patterns have not previously been discussed concerning underspecification, these ERSP analyses identified potentially new indices of phonological underspecification.

ERP Evidence For [Consonantal] Underspecification

Consistent with previous reports of phonological underspecification (Diesch and Luce, 1997; Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013; Scharinger et al., 2012; Schluter et al., 2016; Cummings et al., 2017), ERP

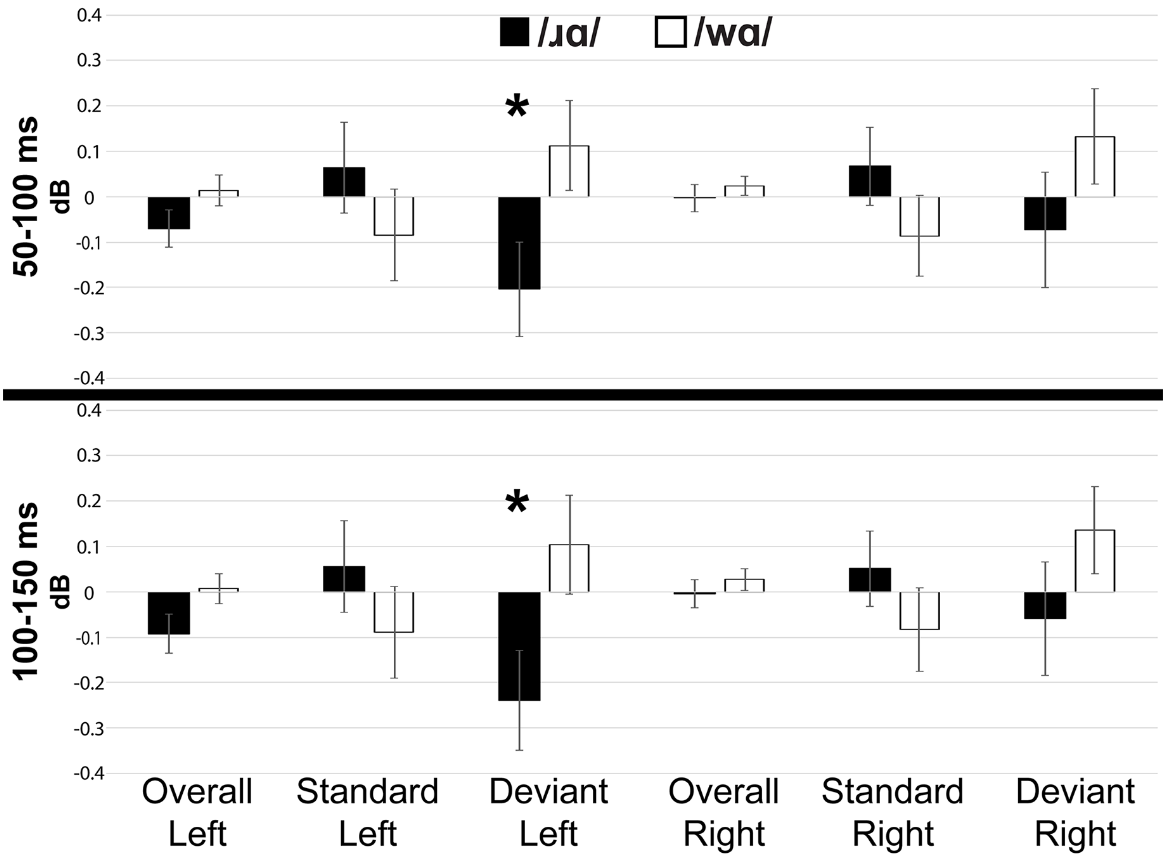


FIGURE 7 | Event-related spectral perturbation activation (in dB) elicited by /wɑ/ and /ɹɑ/ for the low gamma (25–35 Hz) bandwidth across five 50 ms time windows: 50–100 ms, 100–150 ms, 150–200 ms, 200–250 ms, and 250–300 ms. The /ɹɑ/ deviants elicited significantly less low gamma neural synchrony over the left hemisphere, as compared to the right. *Significant effects.

evidence for phonological underspecification was observed. The underspecified /wɑ/ elicited larger neural responses than did more specified /ɹɑ/. Moreover, the cluster permutation analyses identified a significant difference between the /wɑ/ standards and deviants, indicative of a reliable MMN response. No significant difference was observed between the /ɹɑ/ standards and deviants. Thus, the /wɑ/ deviant response (elicited within the /ɹɑ/ standard) appeared to drive the phoneme underspecification differences. These findings were consistent with the underspecification logic of FUL (Eulitz and Lahiri, 2004) which predicts an underspecified phoneme deviant (i.e., /wɑ/) presented within a stream of specified phoneme standards (i.e., /ɹɑ/) would elicit a large mismatch response due to the contrast in [consonantal] feature specification.

While [consonantal] was the obvious feature that differentiated /ɹ/ and /w/, these two phonemes also differ in terms of their place of articulation, with /ɹ/ being characterized as [coronal: +distributed] and /w/ being characterized as [labial] by (Clements and Hume, 1995; **Figure 1**). Given the previous investigations of [coronal] underspecification (Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013; Scharinger et al., 2012; Cummings et al., 2017), possibly the place of articulation of these phonemes would affect the neural response patterns.

Based on previous FUL work, /ɹ/ could have arguably constituted the underspecified phoneme due to its [coronal] place. However, [coronal] also has the assigned daughter [+distributed] within the Clements and Hume (1995) model[11]. This dependent feature is on a lower level of the feature tree than that of /w/'s [labial]. As features lower on the tree are more specified than those higher up in the tree (Core, 1997), in the Clements and Hume (1995) model, /ɹ/ is more specified in terms of place of articulation [coronal: +distributed], as well as in the manner of articulation [+consonantal]. If the [labial] of /w/ was considered to be underspecified as compared to the [coronal: +distributed] of /ɹ/, this would be contrary to all previous work proposing [coronal] underspecification. As a result, it is hypothesized that the place of articulation was not the target feature contrast of /ɹ/ and /w/. However, the multiple features that are underspecified in /w/ (i.e., [−consonantal] and [labial]), make it unclear as to what exactly might have been the feature that was driving the observed MMN asymmetry.

Additional studies contrasting liquid and glide phonemes are necessary to further test [consonantal] underspecification.

[11]The LN is a negativity that follows the MMN and typically peaks between 300 and 500 ms at fronto-central electrode sites.

Since /ɹ/ and /w/ differ not only in terms of [consonantal] but also in terms of place of articulation, a contrast that only varies [consonantal] is needed. This contrast is possible in /l/ and /j/. Both phonemes are [coronal] in nature, thus their main feature distinction is [consonantal], with /l/ being [+consonantal] ([−vocoid]) and /j/ [−consonantal] ([+vocoid]). Importantly, similar to /ɹ/, prevocalic /l/ often undergoes the phonological process of liquid gliding during typical and atypical phonological development, with /l/ substituted with [j] and/or [w]. For example, young children commonly produce "like" (i.e., /lɑIk/) as [jɑIk]. Thus, both liquid phonemes in American English are commonly observed to undergo liquid gliding during phonological development. These developmental and clinical observations provide additional evidence for the possibility that both American English glides, /w/ and /j/, are underspecified as compared to the American English liquids /ɹ/ and /l/. Replication of the present study with /l/ and /j/ would provide important converging evidence for the underspecification of [consonantal] in glide phonemes.

ERSP Evidence For [Consonantal] Underspecification

Since EEG neural oscillation patterns drive ERP responses, it was hypothesized that they could be additional indices of phonological underspecification. Exploratory analyses were conducted to examine whether specified and underspecified phonemes elicited distinct patterns of neural activity. Indeed, significant differences in neural oscillation patterns were elicited by /ɹɑ/ and /wɑ/. In the theta band, /ɹɑ/ elicited more spectral power than did /wɑ/. It has been proposed that inherent, resting-state oscillations in the primary auditory cortex undergo phase resetting—particularly in the theta range—in response to speech stimuli (Ghitza, 2011; Giraud and Poeppel, 2012). Thus, the enhanced theta activities between 100 and 300 ms to /ɹɑ/ relative to /wɑ/ likely reflect the impact of specification on this phase resetting process. Within the FUL framework, a more specified phoneme contains more exact phonetic feature information in its phonological representation—which may drive more precise theta phase-locking to the presentation of /ɹɑ/ syllables, yielding a stronger evoked response as compared to an underspecified phoneme that does not contain the same degree of robust featural specification.

As theta activities are proposed to capture syllable-level processing (Ghitza, 2011; Giraud and Poeppel, 2012), a secondary interpretation of the theta band results is acoustic in nature. That is, /ɹɑ/ may have been acoustically more distinct, with a clearer syllable onset boundary, than was /wɑ/. As the sharpness of a syllable's acoustic edges affects how easily the stimulus can be parsed into chunks (Prendergast et al., 2010; Ding and Simon, 2014; Doelling et al., 2014), /ɹɑ/ was able to elicit greater theta neural synchrony than /wɑ/. To explain further, /ɹ/ is a more preferable syllable onset consonant than /w/ due to the sounds' sonority differences (Clements, 1990). Specifically, listeners prefer syllables with strong consonant onsets that are clearly differentiated from the vowel nucleus (e.g., the Head Law; Vennemann, 1988). Since /w/ is nearly as sonorous as vowels, it does not provide a clear differentiated onset; thus, syllable-initial /ɹ/ is preferred over /w/ cross-linguistically (Dziubalska-Kołaczyk, 2001). This acoustic interpretation is still consistent with the idea of feature specification, as the [consonant] aspect of /ɹ/ is what arguably makes it a stronger syllable onset than that of /w/. Thus, while /w/ can function as a syllable onset (Bernhardt and Stoel-Gammon, 1994), /ɹɑ/ is a better-formed syllable than /wɑ/ because of its specified [consonant] feature.

While theta band activity has been correlated with syllable-level processing, low gamma band has been correlated with more rapid information sampling, analysis, and decoding (Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012), likely linked to the binding of different acoustic features needed to derive phonological representations from incoming speech signals. Notably, low gamma band responses have not been consistently observed in auditory paradigms (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Luo et al., 2010), potentially due to the stimuli used (Luo and Poeppel, 2012). The present study provided an ideal situation for eliciting distinguishing gamma responses, as [consonantal] was the contrasting feature between the phonemes.

Our findings revealed less low gamma activation over the left hemisphere as compared to the right hemisphere overall. Specifically, the low gamma activation in response to the /ɹɑ/ deviants was reliably less over the left hemisphere, as compared to the right, whereas the /wɑ/ deviants did not elicit laterality differences. Moreover, /wɑ/ elicited greater low gamma activation over the left hemisphere as compared to /ɹɑ/, while no phoneme differences were observed over the right hemisphere. Thus, /ɹɑ/ appeared to elicit a distinct pattern of activation over the left hemisphere. Interpreting this finding is challenging, given the lack of prior findings. However, a general interpretation could be similar to that of the MMN results. Namely, /wɑ/ elicited greater low gamma activation over the left hemisphere due to its underspecified nature. Future studies will need to continue to test the relationship between underspecification and low gamma activation.

Alternative Interpretations and Study Limitations

While the data in the present study provide evidence of [consonantal] underspecification, other interpretations are possible. For example, a memory/usage-based account of language (UBA; Pierrehumbert, 2006; Bybee, 2010) addresses how the neighborhood density of phonemes affects processing. That is, the larger a phoneme's phonological neighborhood, the more difficult it is to identify and differentiate a specific phoneme from others within the neighborhood. Within UBA, the [+consonantal] category contains many more consonants (21: /p b t d k g f v θ ð s z ʃ ʒ ʧ ʤ m n ŋ l ɹ/) than does the unspecified [−consonantal] category (3: /w j h/). Thus, /ɹ/ has a denser phonological neighborhood than does /w/. When considering MMN responses in the context where /ɹɑ/ is the standard and /wɑ/ is the deviant, UBA would predict that the large phonological neighborhood of /ɹ/ would negatively impact the system's ability to create a strong feature prediction of [+consonantal]. Without clear feature specification, this situation should result in a no mismatch situation and a small/no MMN being elicited by the /wɑ/ deviant. Conversely, in the context where /wɑ/ is the standard and /ɹɑ/ is the deviant, UBA would predict that the small phonological neighborhood of /w/ would allow the system to establish a strong feature prediction. This should result in a true mismatch situation—and also in a large MMN being elicited by the /ɹɑ/ deviant. However, neither one of these proposed results was observed in the present study. Instead, the exact opposite MMN response patterns were observed. Thus, it does not appear that UBA can account for the present study's findings.

The frequency occurrence of sounds in the ambient language environment could have unintentionally affected the MMN responses observed in the present study. Specifically, the MMN can reflect the phonotactic probability of phoneme combinations (Bonte et al., 2005; Näätänen et al., 2007). That is, the statistical regularity of sound combinations in a language can modulate the size of the MMN response. For example, nonwords with high phonotactic probability have been found to elicit larger MMN responses than nonwords with low phonotactic probability (Bonte et al., 2005). Bonte et al. (2005) suggested that the frequent co-occurrence of certain phoneme combinations could result in enhanced auditory cortical responses. In the present study, the phonotactic probability of the /ɹɑ/ syllable in English was greater than that of /wɑ/ (**Table 1**). Thus, following the results of Bonte et al. (2005), the more frequently occurring /ɹɑ/ should have elicited a larger MMN than did /wɑ/, which was not observed. The same general argument could be made for the frequency of occurrence of single phonemes, with /ɹ/ occurring much more frequently in English than /w/ (**Table 1**). However, again, the high frequency of /ɹ/ did not elicit larger MMN responses than did the less commonly occurring /w/. While the findings of the present study do not appear to be driven by the frequency of occurrence of the phonemes, this will remain a possible interpretation until this prediction is directly tested. Fully-crossed stimulus sets with similar individual phoneme and syllable phonotactic probabilities should be used to elicit responses from high and low frequency phonemes and syllables.

The present study included identity difference waves to control for basic differences in acoustic detail present in the /ɹɑ/ and /wɑ/ stimuli. However, it is still a possibility that the study design and/or stimuli did not test phonological representations, but rather tested the phonetic differences between the stimuli. It has been suggested that a single-standard MMN experiment can only capture the phonetic differences between speech sounds. That is, if the standards are not varied, the established memory trace is based on the consistent phonetic makeup of the standard. It has been argued that a variable-standards MMN experimental design (e.g., /t/ produced with multiple voice onset time allophones) is necessary instead to establish a true phonemic MMN (Phillips et al., 2000; Hestvik and Durvasula, 2016). For example, Hestvik and Durvasula (2016) only observed an underspecification MMN asymmetry using a variable-standards paradigm; symmetrical MMN responses were elicited with a single-standards paradigm.

While the possibility remains that the present study only captured phonetic differences between /ɹɑ/ and /wɑ/, the data suggest that the phonological level of representation was tested. The previous MMN studies accessing phonological representations only used a single deviant within their multiple-standard presentations (Phillips et al., 2000; Hestvik and Durvasula, 2016). Although the present study used a single-standard paradigm, it did incorporate three phoneme deviants. The three deviants were included to maximize the MMN responses. That is, the response to a deviant is reduced not only when it is preceded by itself, but also when it is preceded by other similar stimuli (Sams et al., 1984; Näätänen et al., 2007; Symonds et al., 2017). However, the reduction in MMN amplitude can be reduced if the second of two successive deviants differs from the standards in a different attribute/feature than the first deviant (Nousak et al., 1996; Müller et al., 2005; Näätänen et al., 2007). The two unused deviants in the present study, /bɑ/ and /dɑ/, were chosen in part because they were phonetically distinct from /ɹ/ and /w/. Thus, the presentation of multiple deviants, and the phonetic distinctiveness of the stimuli, could have allowed for phonological categorization to occur. Indeed, unlike Hestvik and Durvasula (2016), asymmetrical MMN responses were found in the present study, indicative of phonological-level processing.

A basic stimulus difference could also explain why a phonological mismatch asymmetry was elicited, rather than the symmetrical phonetic mismatch response predicted by previous studies. That is, previous studies used synthetic speech, while the present study used naturally-produced syllables. The acoustic-phonetic structure of synthetic speech conveys less information (per unit of time) than that of natural speech (Nusbaum and Pisoni, 1985). As a result, synthetic speech is considered to be perceptually impoverished as compared to natural speech because basic acoustic-phonetic cues are obscured, masked, or physically degraded in some way. Natural speech is highly redundant at the level of acoustic-phonetic structure, with many acoustic cues being present in the signal. As limited acoustic information is present in synthetic speech, some phonetic feature distinctions are minimally cued. This means that a single cue presented within a single synthetic stimulus might not be enough to convey a particular level of feature distinction. As a result, multiple different tokens of a synthetic phoneme might need to be presented to fully establish a phonemic category. This hypothesis is supported by the results of the previous studies (Phillips et al., 2000; Hestvik and Durvasula, 2016). Alternatively, the spectral variation and redundancy found in the naturally produced speech tokens of the present study might have been enough to accurately establish phonemic categories.

Thus, the naturally produced standard and deviants in the present study could have allowed for phonological categorization of all the stimuli, much like the variable standard presentation of synthetic speech did in previous studies. That said, it is still a possibility that the memory trace tested in the present study was a detailed acoustic/phonetic representation rather than a phonemic representation. Future studies that systematically vary the phonetic allophonic productions and phonemic categories of both standards and deviants are needed to address how best to access phonological representations. Additional studies contrasting synthetic and naturally-produced speech will also provide information regarding how specified and unspecified features are stored and accessed.

As discussed previously concerning theta band activities, possibly the acoustic differences of /ɹɑ/ and /wɑ/ alone were responsible for the observed MMN response asymmetry. That is, the intrinsic physical differences between stimuli could elicit different MMN response patterns (Näätänen et al., 2007). For example, the larger sonority difference between /ɹ/ and /ɑ/ made it an acoustically more distinctive syllable than that of /wɑ/[12]. In other words, the /w/ is perceptually more similar to /ɑ/ than is /ɹ/. Thus, if acoustic distinctiveness and clarity were the underlying mechanisms driving the MMN responses, hearing the deviant /ɹɑ/ within a stream of the /wɑ/ standards should have elicited a larger MMN response than hearing the deviant /wɑ/ in the stream of /ɹɑ/ standards. Yet, the opposite MMN response pattern was observed. The less acoustically distinct /wɑ/ deviant elicited a larger MMN than did the acoustically preferable /ɹɑ/. Moreover, the MMN response elicited by both syllables was larger over the left hemisphere, as compared to the right, which is indicative of feature-level processing; acoustical change detection would have been indicated by similar bilateral MMN responses (Näätänen et al., 1997).

While underspecification is presumed to be a language universal phenomenon, possibly the specification of features can vary across languages. For example, voiced stops are underspecified in English, while voiceless stops are underspecified in Japanese (Hestvik and Durvasula, 2016; Hestvik et al., 2020). In terms of /ɹ/, Natvig (2020) proposed that liquids, and rhotics in particular, are underspecified consonantal sonorants due to the multiple variations of "r-sounds" that occur in languages such as German, Arabic, Hawaiian, New Zealand Maori, Malayalam, and Norwegian. While it is beyond the scope of this study to address whether /ɹ/ is specified or not in languages other than English, cross-linguistic differences in [consonantal] underspecification are possible.

The decision to use /ɹ/ and /w/ here was driven by the need to better understand the clinical observation of particular speech error patterns observed during phonological development. Specifically, young typically developing children, as well as older children with speech sound disorders, often have difficulty producing /ɹ/ with adequate palatal and pharyngeal constriction, resulting in an incorrect [w] production. Thus, it was hypothesized that constriction [i.e., (consonantal)] is the primary distinguishing feature of /ɹ/ and /w/, at least in American English. Clements and Hume (1995) feature geometry theory was used to address the underlying differences in the phonological representations of /ɹ/ and /w/. Alternative explanations, including usage-based phonology, phonotactic probability, and sonority/acoustics/phonetics were explored. However, none of the predictions made by these approaches fit with the data. Moreover, while it was possible that [labial] underspecification of /w/ elicited the observed results, that explanation would not be consistent with the many previous studies showing [coronal] to be the underspecified place of articulation. As a result, the presence or absence of [consonantal] in the phonological representations of /ɹ/ and /w/, respectively, is the current best explanation of the results. Future work can either further confirm and extend our proposal, or correct it as needed.

FUL's underspecification predictions, tested within an oddball paradigm, provide a clear framework within which to examine feature encoding and the specification of phonological representations. By contrasting single phonemes, different patterns of neural responses can be associated with distinctive features. The identification of individual features' neural patterns is a necessary first step in understanding how speech perception and processing lead to language comprehension and production. However, as pointed out by a reviewer, the use of individual phonemes and/or syllables in the oddball paradigm does not capture the complexity of parsing phonemes (and features) and their subsequent mapping onto lexical items in single words or continuous speech (Gwilliams et al., 2018, 2020; Dikker et al., 2020). To further understand how phonological underspecification improves the efficiency of speech processing, studies involving naturalistic language tasks are an important next step.

Underlying Neural Mechanisms for Underspecification

From its theoretical inception, underspecification has been proposed as a mechanism to improve the efficiency[13] of speech processing (Chomsky and Halle, 1968; Kiparsky, 1985; Archangeli, 1988; Mohanan, 1991; Clements and Hume, 1995; Steriade, 1995; Eulitz and Lahiri, 2004). That is, an underspecified feature is the default in a phonological representation. It is efficient to assume a feature is underspecified unless evidence is presented to the contrary. The predictability of that default status allows for ease of phonological processing.

The hallmark neural index of underspecification in electrophysiological studies has been a larger MMN to underspecified phonemes, as compared to specified ones. However, few proposals have been made to address the underlying neural mechanisms of this underspecification response. The size of the MMN has been associated with ease of discrimination (Tiitinen et al., 1994; Näätänen et al., 1997). The large underspecification MMN response would thus suggest that it is easier to discriminate an underspecified feature in a phoneme within a stream of specified phonemes, as compared to contrasting a specified feature within a stream of underspecified phonemes. But what does this large MMN response characterize at a neural level?

From a neurophysiological standpoint, one possibility is that the size of the MMN reflects the tuning characteristics of the

[12]It should be noted that this is a different feature assignment than what would be found in FUL (Lahiri & Reetz, 2002), as there are no articulator dependents in FUL; this is what allows for [coronal] to be underspecified. The dependents in the Clements and Hume (1995) make it difficult, or impossible, for [coronal] to be underspecified.

[13]It is important to note that the acoustic aspects of cannot be fully differentiated from language experience and phonotactics as honeme combinations that are perceptually more distinct tend to occur more often in and across languages (Bonte et al., 2005).

responding neural populations. That is, the specification of a feature could lead to the recruitment of specialized neural populations that are tuned to only respond to that feature. Conversely, if a phoneme is not specified for a feature, other less-specialized populations of neurons could be recruited to respond. These less-specialized neurons could be weakly tuned for phonetic-acoustic content. By having weaker encoding, these neurons might be more flexible in their perceptual responses and would likely respond to more types of features at the same time. As a result, the responses elicited by the less-specified neurons could be larger than those of the specifically tuned neurons because they are coded to respond to more types of acoustic-phonetic information. Besides, since the less-specified neurons might be activated more frequently due to their lack of feature specification, their responses could be more highly tuned/practiced, which could also result in larger responses.

In regards to the present study, perhaps the underspecified [–consonantal] feature in /wɑ/ could activate the weakly-coded neurons that were tuned to respond to a variety of phonetic-acoustic content. This broad phonetic-acoustic tuning could elicit a large neural response due to the many different cues that might be summed together in the response. Alternatively, the specified feature in /ɹɑ/ could access neuronal populations that were explicitly coded for a single feature, [+consonantal]. Thus, the neurons would respond, but only to that specific feature and ignore all other features. This could result in a small neural response.

Neuroimaging studies have provided some evidence in support of this proposal. For example, very small populations of neurons (characterized by single electrodes or voxels) have been found to encode and respond to linguistically meaningful information, such as formant frequencies (e.g., low-back vowels), phonetic features (e.g., obstruent, plosive, voicing), and/or entire phonemes (Mesgarani et al., 2014; Arsenault and Buchsbaum, 2015; de Heer et al., 2017; Gwilliams et al., 2018; Yi et al., 2019). Also, phonemes and features elicited activation across multiple electrodes and voxels, suggesting that responses were not constrained to a single neural population. Thus, there is evidence for highly tuned neural populations to respond to one, or many, features, while also working in conjunction with other neural populations.

To our knowledge, previous studies of underspecification have not directly discussed the neural implications of underspecification, and rightfully so, given the limited spatial resolution of scalp-level EEG recordings (Luck, 2014). The present study proposes some possible neural-level interpretations of its results. Future collaborative work with researchers using spatially sensitive neuroimaging techniques will be necessary to further define the underlying neural mechanisms of underspecification.

Summary and Conclusions

The less specified /wɑ/ elicited a large MMN, whereas a much smaller MMN was elicited by the more specified /ɹɑ/. This outcome reveals that the [consonantal] feature follows the underspecification predictions of FUL previously tested with the place of articulation and voicing features. Thus, this study provides new evidence for the language universality of underspecification by addressing a different phoneme feature. Moreover, left hemisphere low gamma activation characterized distinct phoneme-specific feature processing patterns for /ɹ/ and /w/, revealing a potentially novel index of underspecification. Examining theta and/or low gamma bandwidths in future studies could provide further support for the claims of underspecification.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Idaho State University Human Subjects Committee and the University of North Dakota Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AC created the stimuli, tested participants, prepared and analyzed the data, and helped write the manuscript. DO and YW analyzed data and helped write the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by NIH grant R15DC013359 (from the National Institute on Deafness and Other Communication Disorders) awarded to the first author (AC). This funding paid for research program equipment, participant research payments, and student research assistants. The open access fees were paid for by Idaho State University start-up funds awarded to the first author. The second author (YW) was supported by NSF grant 1540943.

ACKNOWLEDGMENTS

We would like to thank Kari Lehr for her assistance during the testing of participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.585817/full#supplementary-material.

REFERENCES

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology* 5, 183–207.

Arsenault, J. S., and Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *J. Neurosci.* 35, 634–642. doi: 10.1523/JNEUROSCI.2454-14.2015

Bernhardt, B. (1992). The application of nonlinear phonological theory to intervention with one phonologically disordered child. *Clin. Linguist. Phon.* 6, 283–316. doi: 10.3109/02699209208985537

Bernhardt, B., and Gilbert, J. (1992). Applying linguistic theory to speech-language pathology: the case for nonlinear phonology. *Clin. Linguist. Phon.* 6, 123–145. doi: 10.3109/02699209208985523

Bernhardt, B., and Stoel-Gammon, C. (1994). Nonlinear phonology: introduction and clinical application. *J. Speech Hear. Res.* 37, 123–143.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., and Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clin. Neurophysiol.* 116, 2765–2774. doi: 10.1016/j.clinph.2005.08.012

Broen, P. A., Strange, W., Doyle, S. S., and Heller, J. H. (1983). Perception and production of approximant consonants by normal and articulation-delayed preschool children. *J. Speech Hear. Res.* 26, 601–608. doi: 10.1044/jshr.2604.601

Brown, J. C. (2004). Eliminating the segmental tier: evidence from speech errors. *J. Psycholinguist. Res.* 33, 97–101. doi: 10.1023/b:jopr.0000017222.24698.73

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253

Bybee, J. (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Čeponienė, R., Cheour, M., and Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: evidence for multiple generators. *Electroencephalogr. Clin. Neurophysiol.* 108, 345–354. doi: 10.1016/s0168-5597(97)00081-6

Cheour, M., Čeponienė, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* 1, 351–353. doi: 10.1038/1561

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper and Row.

Clements, G. (1990). "The role of the sonority cycle in core syllabification," in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, eds J. Kingston and M. Beckman (Cambridge: Cambridge University Press), 283–333.

Clements, G., and Hume, E. (1995). "The internal organization of speech sounds," in *The Handbook of Phonological Theory*, ed. J. A. Goldsmith (Cambridge, MA: Blackwell Publishing), 245–306.

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice.* Cambridge, MA: MIT Press.

Core, C. (1997). *Feature Geometry, Underspecification and Child Substitutions.* Miamai, FL: Florida International University. Thesis.

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

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862

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

Cummings, A., Seddoh, A., and Jallo, B. (2016). Phonological code retrieval during picture naming: influence of consonant class. *Brain Res.* 1635, 71–85. doi: 10.1016/j.brainres.2016.01.014

Datta, H., Shafer, V. L., Morr, M. L., Kurtzberg, D., and Schwartz, R. G. (2010). Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development. *J. Speech Lang. Hear. Res.* 53, 757–777. doi: 10.1044/1092-4388(2009/08-0123)

Davidson, D. J., and Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res.* 1158, 81–92. doi: 10.1016/j.brainres.2007.04.082

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., and Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *J. Neurosci.* 37, 6539–6557. doi: 10.1523/JNEUROSCI.3267-16.2017

Delattre, P., and Freeman, D. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics* 6, 29–68.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Diesch, E., and Luce, T. (1997). Magnetic mismatch fields elicited by vowels and consonants. *Exp. Brain Res.* 116, 139–152. doi: 10.1007/pl00005734

Dikker, S., Assaneo, M. F., Gwilliams, L., Wang, L., and Kösem, A. (2020). Magnetoencephalography and language. *Neuroimaging Clin. N. Am.* 30, 229–238. doi: 10.1016/j.nic.2020.01.004

Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465. doi: 10.1016/j.cub.2015.08.030

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035

Dziubalska-Kołaczyk, K. (2001). *Constraints and Preferences.* Berlin: Walter de Gruyter.

Edwards, M. L. (1983). Selection criteria for developing therapy goals. *J. Childhood Commun. Disord.* 7, 36–45.

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. doi: 10.1111/j.1469-8986.2007.00602.x

Fromkin, V. (1973). *Speech Errors as Linguistic Evidence.* The Hague: Mouton.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130

Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology* 16, 29–54.

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063

Greenberg, J. H. (1975). Research on language universals. *Ann. Rev. Anthropol.* 4, 75–94.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Grunwell, P. (1987). *Clinical Phonology, 2nd edition*. Baltimore, MD: Williams and Wilkins.

Gurariy, G., Killebrew, K. W., Berryhill, M. E., and Caplovitz, G. P. (2016). Induced and evoked human electrophysiological correlates of visual working memory set-size effects at encoding. *PLoS One* 11:e0167022. doi: 10.1371/journal.pone.0167022

Gwilliams, L., King, J.-R., Marantz, A., and Poeppel, D. (2020). Neural dynamics of phoneme sequencing in real speech jointly encode order and invariant content. *BioRxiv* [Preprint]. doi: 10.1101/2020.04.04.025684

Gwilliams, L., Linzen, T., Poeppel, D., and Marantz, A. (2018). In spoken word recognition, the future predicts the past. *J. Neurosci.* 38, 7585–7599. doi: 10.1523/JNEUROSCI.0065-18.2018

Halle, M., Vaux, B., and Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Linguist. Inq.* 31, 387–444. doi: 10.1162/002438900554398

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

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

Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511. doi: 10.1152/jn.00251.2010

Hume, E., and Odden, D. (1996). Reconsidering [consonantal]. *Phonology* 13, 345–376.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111, 1745–1758. doi: 10.1016/s1388-2457(00)00386-2

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* 33, 630–645.

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonol. Yearbook* 2, 85–138.

Korpilahti, P., Lang, H., and Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalogr. Clin. Neurophysiol.* 95:P96.

Lahiri, A., and Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245–294. doi: 10.1016/0010-0277(91)90008-r

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Labphon 7*, eds C. Gussenhoven and N. Warnereds (Berlin: Mouton), 637–676.

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phon.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38; discussion 38–75. doi: 10.1017/s0140525x99001776

Luck, S. (2014). *An Introduction to the Event-Related Potential Technique, 2nd edition*. Cambridge, MA: MIT Press.

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004

Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170

Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8:e1000445. doi: 10.1371/journal.pbio.1000445

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293. doi: 10.1016/0013-4694(93)90110-h

McLeod, S., and Crowe, K. (2018). Children's consonant acquisition in 27 languages: a cross-linguistic review. *Am. J. Speech Lang. Pathol.* 27, 1–26. doi: 10.1044/2018_AJSLP-17-0100

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Mohanan, K. P. (1991). On the bases of radical underspecification. *Nat. Lang. Linguist. Theory* 9, 285–325. doi: 10.1007/BF00134678

Müller, D., Widmann, A., and Schröger, E. (2005). Deviance-repetition effects as a function of stimulus feature, feature value variation and timing: a mismatch negativity study. *Biol. Psychol.* 68, 1–14. doi: 10.1016/j.biopsycho.2004.03.018

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Näätänen, R., and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* 125, 826–859. doi: 10.1037/0033-2909.125.6.826

Natvig, D. (2020). Rhotic underspecification: Deriving variability and arbitrariness through phonological representations. *Glossa* 5:48. doi: 10.5334/gjgl.1172

Nousak, J. M., Deacon, D., Ritter, W., and Vaughan, H. G. (1996). Storage of information in transient auditory memory. *Cogn. Brain Res.* 4, 305–317. doi: 10.1016/s0926-6410(96)00068-7

Nusbaum, H. C., and Pisoni, D. B. (1985). Constraints on the perception of synthetic speech generated by rule. *Behav. Res. Methods Instrum. Comput.* 17, 235–242. doi: 10.3758/bf03214389

Palmer, J. A., Makeig, S., Kreutz-Delgado, K., and Rao, B. D. (2008). "Newton method for the ICA mixture model," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing,* (Las Vegas, NV: IEEE), 1805–1808. doi: 10.1109/ICASSP.2008.4517982

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567

Picton, T. W., Alain, C., Otten, L., Ritter, W., and Achim, A. (2000). Mismatch negativity: different water in the same river. *Audiol. Neurootol.* 5, 111–139. doi: 10.1159/000013875

Pierrehumbert, J. (2006). The next toolkit. *J. Phon.* 34, 516–530. doi: 10.1016/j.wocn.2006.06.003

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

Prendergast, G., Johnson, S. R., and Green, G. G. R. (2010). Temporal dynamics of sinusoidal and non-sinusoidal amplitude modulation. *Eur. J. Neurosci.* 32, 1599–1607. doi: 10.1111/j.1460-9568.2010.07423.x

Preston, J., Benway, N., Leece, M., Hitchcock, E., and McAllister, T. (2020). Tutorial: motor-based treatment strategies for /r/ distortions. *Lang. Speech Hear. Serv. Sch.* 51, 966–980. doi: 10.1044/2020_LSHSS-20-00012

Sams, M., Alho, K., and Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology* 21, 434–441. doi: 10.1111/j.1469-8986.1984.tb00223.x

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

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *J. Speech Lang. Hear. Res.* 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of English fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Secord, W. A., Boyce, S. E., Donohue, J. S., Fox, R. A., and Shine, R. E. (2007). *Eliciting Sounds: Techniques and Strategies for Clinicians*, 2nd edition. Clifton Park, NY: Cengage Learning.

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., and Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *J. Cogn. Neurosci.* 17, 1168–1180. doi: 10.1162/0898929054475217

Shriberg, L. D. (1980). An intervention procedure for children with persistent /r/ errors. *Lang. Speech Hear. Serv. Sch.* 11, 102–110. doi: 10.1044/0161-1461.1102.102

Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.* 18, 555–586. doi: 10.1146/annurev.ne.18.030195.003011

Steriade, D. (1995). "Underspecification and markedness," in *The Handbook of Phonological Theory*, ed J. A. Goldsmith (Oxford and Cambridge, MA: Blackwell Publishing), 114–174.

Stoel-Gammon, C., and Dunn, C. (1985). *Normal and Disordered Phonology in Children.* Baltimore, MD: University Park Press.

Symonds, R. M., Lee, W. W., Kohn, A., Schwartz, O., Witkowski, S., and Sussman, E. S. (2017). Distinguishing neural adaptation and predictive coding hypotheses in auditory change detection. *Brain Topogr.* 30, 136–148. doi: 10.1007/s10548-016-0529-8

Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372, 90–92. doi: 10.1038/372090a0

Vennemann, T. (1988). *Preference Laws For Syllable Structure and the Explanation of Sound Change: With Special Reference to German, Germanic, Italian and Latin.* Berlin: Mouton de Gruyter.

Vitevitch, M. S., and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instrum. Comput.* 36, 481–487. doi: 10.3758/bf03195594

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn. Brain Res.* 7, 357–369. doi: 10.1016/s0926-6410(98)00039-1

Yi, H. G., Leonard, M. K., and Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron* 102, 1096–1110. doi: 10.1016/j.neuron.2019.04.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fnhum.2021.612345

Asymmetries in Accessing Vowel Representations Are Driven by Phonological and Acoustic Properties: Neural and Behavioral Evidence From Natural German Minimal Pairs

Miriam Riedinger[1]*, *Arne Nagels*[1], *Alexander Werth*[2] *and Mathias Scharinger*[2,3]

[1]*Department of English and Linguistics, Johannes Gutenberg University, Mainz, Germany,* [2]*Institute for German Linguistics, Philipps University, Marburg, Germany,* [3]*Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany*

OPEN ACCESS

Edited by:
Yang Zhang,
University of Minnesota Health Twin Cities, United States

Reviewed by:
Linda Polka,
McGill University, Canada
Miwako Hisagi,
California State University, United States

***Correspondence:**
Miriam Riedinger
miriedin@uni-mainz.de

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *30 September 2020*
Accepted: *26 January 2021*
Published: *18 February 2021*

Citation:
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. Front. Hum. Neurosci. 15:612345. doi: 10.3389/fnhum.2021.612345

In vowel discrimination, commonly found discrimination patterns are directional asymmetries where discrimination is faster (or easier) if differing vowels are presented in a certain sequence compared to the reversed sequence. Different models of speech sound processing try to account for these asymmetries based on either phonetic or phonological properties. In this study, we tested and compared two of those often-discussed models, namely the Featurally Underspecified Lexicon (FUL) model (Lahiri and Reetz, 2002) and the Natural Referent Vowel (NRV) framework (Polka and Bohn, 2011). While most studies presented isolated vowels, we investigated a large stimulus set of German vowels in a more naturalistic setting within minimal pairs. We conducted an mismatch negativity (MMN) study in a passive and a reaction time study in an active oddball paradigm. In both data sets, we found directional asymmetries that can be explained by either phonological or phonetic theories. While behaviorally, the vowel discrimination was based on phonological properties, both tested models failed to explain the found neural patterns comprehensively. Therefore, we additionally examined the influence of a variety of articulatory, acoustical, and lexical factors (e.g., formant structure, intensity, duration, and frequency of occurrence) but also the influence of factors beyond the well-known (perceived loudness of vowels, degree of openness) in depth *via* multiple regression analyses. The analyses revealed that the perceptual factor of perceived loudness has a greater impact than considered in the literature and should be taken stronger into consideration when analyzing preattentive natural vowel processing.

Keywords: vowel discrimination, mismatch negativity (MMN), reaction time (RT), multiple regression analysis, perceived loudness

INTRODUCTION

In recent years, much research has been done on the mental representations of vowels and on investigating which properties are involved in vowel discrimination. This article investigates the mental representations of vowels and compares two models that both make specific hypotheses regarding sound discrimination and mental representations of speech sounds, namely the Featurally Underspecified Lexicon (FUL) model (Lahiri and Reetz, 2002, 2010) and the Natural Referent Vowel (NRV) framework (Polka and Bohn, 2011). Based on the notions that spoken language has a sequential, serial structure and those earlier events that precede later events influence the recognition or discrimination of those later events, both models have in common that they are predicting directional asymmetries in the discrimination of speech sounds: discrimination of two speech sounds is easier in one direction than in the other; therefore, 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]). Both models assume that vowel discrimination is based on the nature of the mental representation and predict facilitated discrimination in one direction, but predictions about the easier presentation order are often competing. Furthermore, what separates the models are the substantially different assumptions about the features involved in discrimination processes and therefore in mental representations.

Within the FUL model, Lahiri and Reetz (2002, 2010) made a proposition for speech perception and lexical access suggesting that speech sounds can be described with the help of abstract and underspecified feature specifications (e.g., [HIGH] for high or close vowels, such as [i]). Importantly, they also describe sound processing based on those features. Crucially, this model assumes that there can be a discrepancy between the features contained in the signal and those stored in the mental lexicon, since mental representations may be underspecified and therefore do not contain all possible features. These assumptions of underspecified mental representations express both similarities and differences to other approaches of underspecification. In common with other underspecification theories, the underspecified sound descriptions are based on the notion of *minimalism*. In this respect, it is postulated that only a distinct set of sound descriptors are necessary for underlying representations. But in contrast to theories like Radical Underspecification (Archangeli, 1988) the underspecification approach in FUL is not only a theoretical means to describe certain linguistic phenomena (e.g., assimilation) but also constitutes mental representations of speech. Therefore underspecification is directly involved in speech perception and production. Additionally, in FUL sounds can be described solely with monovalent features. For example, in FUL it is believed that coronal segments (i.e., front vowels) are underspecified for a place of articulation information ([–]) in the mental representation, but the feature [COR] can be retrieved from the auditory signal. This underspecification approach, together with the specific proposed ternary mapping process, is the reason for the resulting directional asymmetries in sound discrimination. This mapping process includes a comparison of the features obtained from the signal with those stored in the mental lexicon. Due to the underspecification of redundant features, there are three possible outcomes: a match occurs if the feature extracted from the signal has the same equivalent feature in the mental lexicon (e.g., [DOR]—[DOR]: [u]—[o]). A mismatch occurs if the feature taken from the signal and the feature in the underlying representation are complementary and exclude each other (e.g., [HIGH]—[LOW]: [i]—[a]). Last but not least, a no-mismatch occurs if a feature extracted from the signal neither mismatches with a feature of the mental lexicon nor matches it. The last setup of the mapping process is crucial in the elicitation of directional asymmetries. For example, if [COR] is extracted from the signal, this feature produces a mismatch with [DOR] in the lexicon, but if [DOR] is extracted from the signal, the result is a no-mismatch due to the underspecification of the coronal place of articulation ([–]). These different results should become apparent when the discrimination of two vowels is tested in both possible presentation orders ([i]—[u] vs. [u]—[i]).

Several studies have shown that the presentation order with a mismatch as the result of the mapping process usually elicits larger effects than vice versa. Eulitz and Lahiri (2004) conducted an ERP study with German vowels [o], [ø], and [e], which differ mainly in place of articulation. When discriminating [o]—[ø], larger electrophysiological responses occurred because of the mismatching features [DOR]—[COR]. In the reverse direction, the effects were attenuated due to the underspecification of the coronal place of articulation. Similar results have been produced by Scharinger et al. (2012b) for tongue height oppositions using American English vowels for which the mid of tongue height is believed to be underspecified. They found larger effects if the mid vowel [ɛ] had to be discriminated from low vowel [æ] due to the mismatching features of [LOW] and [MID] compared to the reverse sequence, in which there is no feature mismatch due to underspecification. Similar evidence for this approach has been found not only for vowels (Lipski et al., 2007; de Jonge and Boersma, 2015) but also for consonants (Hestvik and Durvasula, 2016; Schluter et al., 2016, 2017; Cummings et al., 2017; Højlund et al., 2019; Hestvik et al., 2020) and suprasegmental elements like lexical tones (Politzer-Ahles et al., 2016). While most studies used isolated vowels or syllables, there is also evidence from complex stimuli like words (Friedrich et al., 2008; Scharinger et al., 2012b; Cornell et al., 2013; Lawyer and Corina, 2018).

The other model investigated in this article, the NRV framework, also predicts different discrimination performances as a function of presentation order. In contrast to the aforementioned model, NRV operationalizes phonetic properties of the speech signal which can be specified by acoustical or visual cues to explain directional asymmetries and predict different discrimination performances and proposes that "vowels with extreme articulatory-acoustic properties (peripheral in the vowel space;" Polka and Bohn, 2011, p. 474) are so-called referent vowels and are easier to discriminate. Polka and Bohn (2003, 2011) observed a universal perceptual bias favoring vowel discrimination from a more central to a more peripheral vowel in the vowel space in infants. They proposed that the vowels on the periphery of the vowel space (/i/, /a/, /u/, /y/) act as universal referent vowels in language development and vowel discrimination due to their more salient and extreme articulatory-acoustic properties. The vowel space periphery's perceptual advantage can be explained by the convergence of adjacent formants and therefore the stronger focalization of the referent vowels (Schwartz et al., 1997, 2005). Since this framework has been developed from the point of view of language acquisition and infant vowel discrimination, much work has been done on the investigation of the proposed perceptual bias in infants. There is evidence from an early cross-linguistic study with German- and English-learning infants that for English vowels /ɛ/ and /æ/, discrimination was easier for /ɛ/—/æ/ than in the reverse direction, regardless of the language background of the infants (Polka and Bohn, 1996). A similar bias with easier discrimination from a more central (less focal) to a more peripheral (more focal) vowel was shown in several studies (Bohn and Polka, 2001; Polka and Bohn, 2011; Pons et al., 2012; Simon et al., 2014). Additionally, there is some encouraging evidence that the perceptual bias preferring some sounds to others in discrimination in infants also could hold true in consonants (Nam and Polka, 2016). Concerning adult vowel perception and discrimination, within the framework, it was initially proposed that the perceptual bias is shaped by language experience. Therefore, the asymmetry only occurs if subjects are discriminating non-native vowel contrasts, while in native vowel contrasts, the perceptual bias disappears, and asymmetry occurs (Polka and Bohn, 2011). The assumption of experience-dependent asymmetries was found in some studies (Tyler et al., 2014; Kriengwatana and Escudero, 2017) while others also report universal biases in adults. In an AX discrimination test with Canadian-French and Canadian-English subjects using tokens of less focal English /u/ and more focal French /u/, Masapollo et al. (2017b) found that discrimination from less to more focalization produced better and faster results irrespective of language background. Therefore, the authors argued that there is a universal bias towards more focalized vowels in adults, too. These results have been replicated and extended in that the universal bias seems to have an impact not only on the auditory domain of speech processing but also on visual vowel discrimination (Masapollo et al., 2017a, 2018).

In recent research, mental representations of speech sounds have often been investigated with the help of electrophysiological methods, for example by using event-related potentials (ERPs). ERPs offer a means to investigate speech processing on a temporal axis with the accuracy of milliseconds. In the investigation of speech sound processing, one ERP component, the so-called Mismatch Negativity (MMN), has become prominent. The MMN can be defined as a specific electrophysiological detection change response of the brain when the repetitive presentation of one stimulus (standard) is interrupted occasionally and unpredictably by a different stimulus (Näätänen et al., 2007). The MMN component has often been used for the investigation of (speech) sound processing since this component can be elicited even when participants are not attending to the stimulation. It, therefore, reflects preattentive and automatic speech processing, making it possible to differentiate the neural responses of stimuli without attention effects and other perceptual and cognitive processes (for a review on the component, see Näätänen et al., 2007). This component usually peaks fronto-centrally between 100 and 250 ms after change onset and can be elicited by any discriminable change in the stimulation (Näätänen, 2001), for example in pure tones (e.g., Sams et al., 1985), with sensitivity for changes in frequency (for example Takegata and Morotomi, 1999; Tervaniemi et al., 2000), intensity and duration (e.g., Paavilainen et al., 1991) but also in more complex stimuli like speech sounds (Dehaene-Lambertz, 1997; Dehaene-Lambertz et al., 2000). Furthermore, several studies have shown that the latency of the component is usually linked to the complexity of the stimuli, while the amplitude of the MMN is correlated with the magnitude of deviation. The greater the differences between the standard and the deviant stimulus are, the greater the MMN (e.g., Sams et al., 1985; Savela et al., 2003). Moreover, it has been shown that the MMN component is sensitive for language-specific phonemic processing of speech sounds (e.g., Dehaene-Lambertz, 1997; Näätänen et al., 1997), which led to the interpretation that mental representations are of phonemic or phonetic nature (as opposed to auditory ones) and language-specific.

In this article, we tested both competing models 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 (Pulvermüller et al., 2001, 2004; Shtyrov and Pulvermüller, 2002; Endrass et al., 2004; Pettigrew et al., 2004a; Shtyrov et al., 2008). 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 a passive oddball paradigm and can lead to a stronger MMN response for words with higher lexical frequency in opposition to deviants with a lower or intermediate frequency of occurrence (Alexandrov et al., 2011; Shtyrov et al., 2011; Aleksandrov et al., 2017). Furthermore, it has been shown that phonotactic probabilities (sequential order of phonemes in words) influence MMN results with higher probability, accompanied by enhanced MMN effects (Bonte et al., 2005; Yasin, 2007; Emmendorfer et al., 2020). Concerning vowel perception, acoustic properties, like for example fundamental frequency, vowel duration or intensity (Aaltonen et al., 1994; Kirmse et al., 2007; Peter et al., 2010; Partanen et al., 2011), 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 concerning the first formant: high vowels have a low F1, while low vowels have a high F1 (Lahiri and Reetz, 2010). 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., 1997). While the common contributing articulatory-based factors of vowel perception have often been investigated, the influence of perceptual and psychoacoustic parameters (e.g., perceptual 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 the processing of natural and unmanipulated German long vowels in the best way; and (2) which factors influence vowel discrimination in natural German minimal pairs? The first question has been addressed on an electrophysiological level through measurement of MMN (Experiment 1) and on a behavioral level in means of reaction times (RT; Experiment 2). 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.

EXPERIMENT 1: MMN STUDY

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).

Participants

Nineteen participants (nine females, mean age 24.7, SD 3.4), graduate and undergraduate students of 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 before 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 analyzed the complete data of 17 participants.

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 concerning the 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 the frequency of occurrence with SUBTlex (Brysbaert et al., 2011), as seen in **Table 1**.

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 analyzed for F0, F1, F2, F3, as well as vowel duration and were all scaled to an intensity level of 70 dB within Praat (Boersma and Weenink, 2016). The five best tokens per word have been chosen as experimental stimuli. Phonetic parameters of

TABLE 1 | Phonetic and lexical parameters of the vowels.

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)	1,488 (59)	3,046 (77)	72.85 (0.27)	292	2.861
	Ziel ("target")	203 (5)	244 (28)	2,445 (61)	3,446 (74)	73.79 (0.92)	159	3.358
/e:/—/i:/	*Steg* ("bridge")	179 (10)	359 (6)	2,483 (58)	3,171 (113)	73,96 (0.67)	270	1.255
	Stieg ("climbed")	190 (3)	286 (11)	2,479 (51)	3,528 (21)	74.92 (0.31)	228	2.352
/a:/—/e:/	*Mahl* ("meal")	174 (5)	913 (10)	1,484 (47)	2,966 (44)	71.68 (0.22)	225	1.672
	Mehl ("flour")	191 (3)	341 (8)	2,566 (39)	3,377 (209)	71.74 (0.26)	195	1.857
/u:/—/i:/	*Stuhl* ("chair")	200 (8)	294 (37)	1,974 (212)	2,647 (99)	73.88 (0.44)	186	2.892
	Stiel ("handle")	199 (6)	275 (22)	2,475 (75)	3,579 (59)	73.67 (0.44)	176	1.756
/u:/—/y:/	*Sud* ("brew")	194 (2)	296 (40)	1,145 (140)	2,628 (93)	74.75 (0.20)	194	0.845
	Süd ("south")	180 (5)	292 (14)	2,085 (120)	2,633 (146)	74.43 (0.43)	223	1.771

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. The frequency of occurrence (as log-values) for each word is given in the last column.

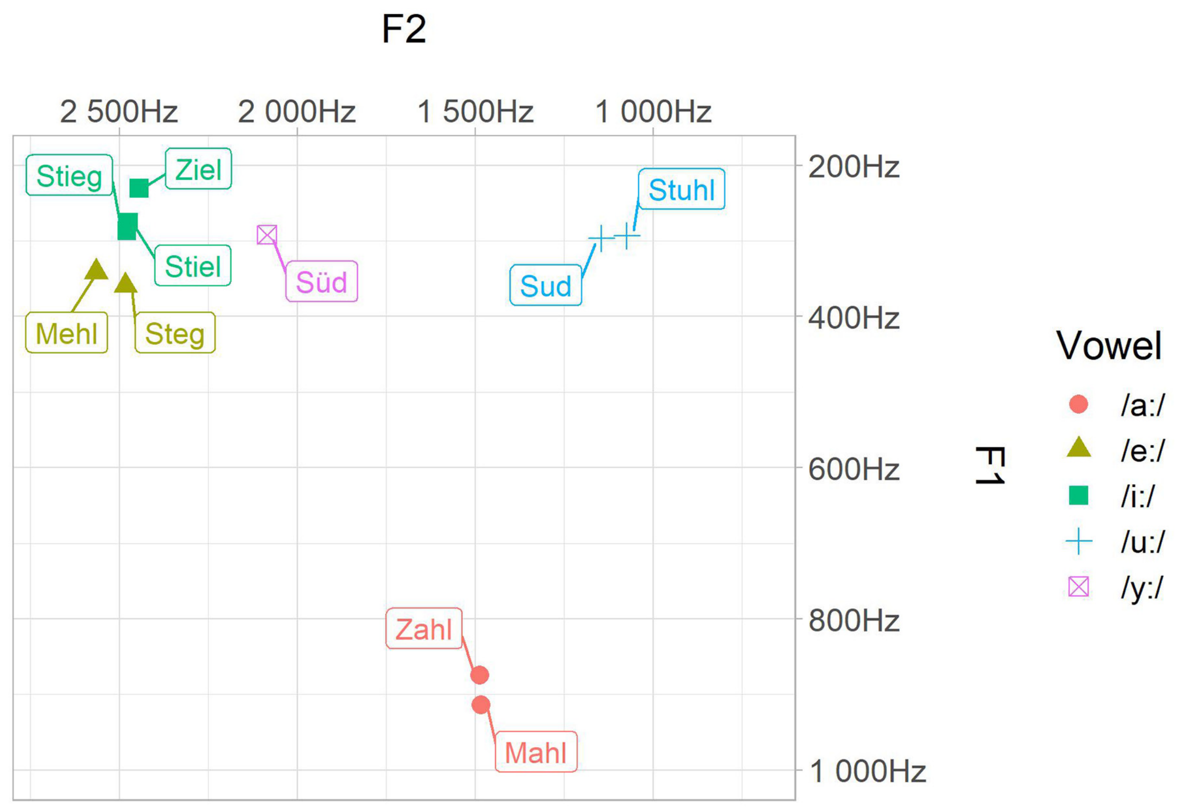


FIGURE 1 | Acoustic characteristic of the stimuli. Mean values of the first (F1) and the second (F2) formant are given per word category in Hertz.

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), but a more detailed description of the acoustic parameters can be found in **Supplementary Table 1**. 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**). We compared the formant values (F1, F2) to the ones of Sendlmeier and 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 representation to cope with inter-token variability (Phillips et al., 2000; Eulitz and Lahiri, 2004; Jacobsen et al., 2004). This is an important design feature since it mediates the likely collinearity between formant frequencies and acoustic- or articulatory-phonetic features.

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 2 h) within 15–20 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 2,000 stimuli for each vowel contrast. Within the blocks, stimuli were randomized, and the interval between two deviants randomly consisted of 4–11 standards. Blocks were randomized for both sessions. Blocks of the same condition never succeeded each other. The fixed ISI was 1,000 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 before the experiment and was kept equal across all subjects (based on the intensity level of 70 dB as manipulated in PRAAT).

Hypotheses

Since this article aims at comparing the two aforementioned models, hypotheses have been made on the assumptions of vowel discrimination following FUL as well as NRV. 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 and Bohn, 2011). The basic assumptions of both models regarding feature specifications and position in the vowel space for each investigated contrast are displayed in **Table 2**.

In accordance with the models, we predict the following MMN effects (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 models. In the other three contrasts, a symmetry should occur since both vowels are peripheral and act as referents within the framework.

EEG Recording and Analysis

EEG was recorded with 28 Ag/AgCl passive electrodes connected to a BrainAmp amplifier (Brain Products GmbH). Electrodes were arranged on an EasyCap in 10-20 positions. AFz served as the 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 and 30 Hz high- and low-pass 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 the correlation of EOG channels and ICA components. The calculation of the MMN component was based on 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).

Results

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

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 and Obleser, 2012; Straußet 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 1,000 randomizations, as implemented in fieldtrip (Oostenveld

TABLE 2 | Assumptions for feature specifications (FUL) and location in the vowel space (NRV) for each vowel contrast and hypotheses for mismatch negativity (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]	Mismatch	Stronger effect	Both peripheral	Symmetrical effect
		[LOW]—[HIGH]	Mismatch			
	Ziel—Zahl	[–]—[DOR]	No-mismatch	Weaker effect		
		[HIGH]—[LOW]	Mismatch			
/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]	Mismatch	Stronger effect	Peripheral—central	Weaker effect
		[LOW]—[MID]	Mismatch			
	Mehl—Mahl	[–]—[DOR]	No-mismatch	Weaker effect	Central—peripheral	Stronger effect
		[–]—[LOW]	No-mismatch			
/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		

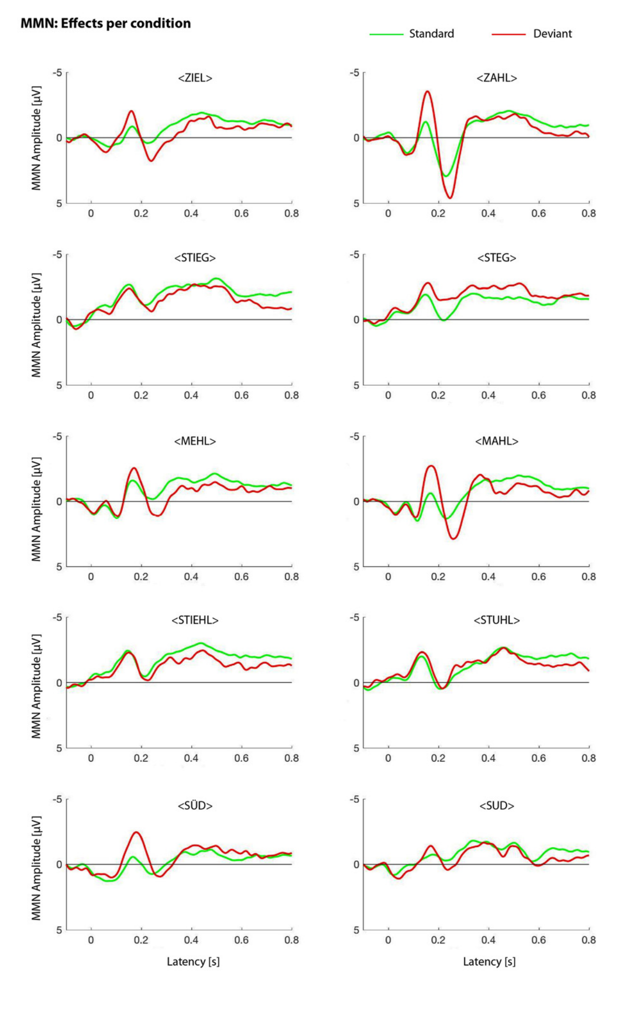


FIGURE 2 | Identity mismatch negativity (MMN) effects per condition. MMN waveforms for all word pairs, in both presentation orders, are shown.

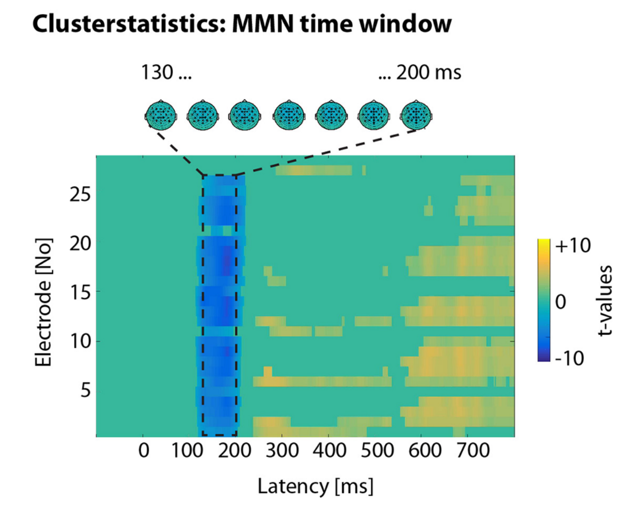


FIGURE 3 | Clusterstatistics. In an electrode × time cluster, deviants elicited more negative responses than standards in the time window between 130 and 200 ms post vowel onset.

et al., 2011), estimated type I-error controlled cluster significance probabilities (at $p < 0.05$). In an electrode × time cluster (with Fz, Cz, CPz, between 130 and 200 ms post vowel onset), deviants elicited a significantly more negative response than standards (see **Figure 3**).

To analyze 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). Electrodes were chosen by cluster statistics.

In the /i:/—/a:/ contrast, we found both main effects for word ($F_{(1,16)} = 7.286$, $p = 0.016$) with a larger MMN for *Zahl* ($M = -1{,}721$, SEM = 0.36; *Ziel*: $M = -0.586$, SEM = 0.253) and for electrode ($F_{(2,32)} = 14.634$, $p < 0.001$) with strongest effect at Cz ($F_{(1,16)} = 5.890$, $p < 0.05$). In the vowel contrast /e:/—/a:/, there was not only a highly significant main effect for electrode ($F_{(2,32)} = 12.307$, $p < 0.001$) but also an interaction word × electrode ($F_{(2,32)} = 4.942$, $p = 0.013$). *Post hoc* analysis of the interaction showed a significant effect on Cz ($F_{(1,16)} = 5.039$, $p < 0.05$). 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 ($F_{(2,32)} = 12.349$, $p < 0.001$) with a marginal effect on CPz ($F_{(1,16)} = 4.265$, $p = 0.055$). 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: $F_{(1,16)} = 1.687$, $p = 0.212$, electrode: $F_{(2,32)} = 2.367$, $p = 0.110$). The same is also true for the comparison of /i:/—/u:/ (word: $F_{(1,16)} = 0.294$, $p = 0.595$, electrode: $F_{(2,32)} = 0.725$, $p = 0.492$).

Discussion

In summary, we found no clear evidence 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 in line with 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 good 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 at first sight not compatible within the FUL paradigm (Lawyer and Corina, 2018). The same is true for German umlauting back vowels. In our case, when deriving the plural of the 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 the 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 the place of articulation information, asymmetry has to occur when it is compared to /i:/, which is also underspecified.

Contrary to this, the results of the remaining two contrasts are somewhat 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:/, an asymmetric 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 et al., 2004a,b).

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, asymmetry 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 in other perceptual and psychoacoustic parameters. There is evidence that, for example, the perceived loudness of speech stimuli varies for vowel quality. That is, lower front vowels are perceived louder despite equal intensity (Glave and Rietveld, 1975, 1979) and vocal effort (Eriksson and 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 using multiple regression analyses in "Explorative Analysis for Additional Influential Factors in MMN and log RT Data" section.

EXPERIMENT 2: REACTION TIMES

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 RT 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 low-level information like variations in duration and intensity (Näätänen et al., 1989; Paavilainen et al., 1991; Schröger, 1996; Jacobsen et al., 2003) or acoustic distance of stimuli (Savela et al., 2003; Deguchi et al., 2010). 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 et al., 2004a). Therefore, RT 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 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 before the experiment.

The stimuli used in Experiment 2 were the same as in Experiment 1. In contrast to the prior experiment, we had to reduce the number of tested vowel contrasts in order to shorten the session length (approximately testing 45 min). 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.

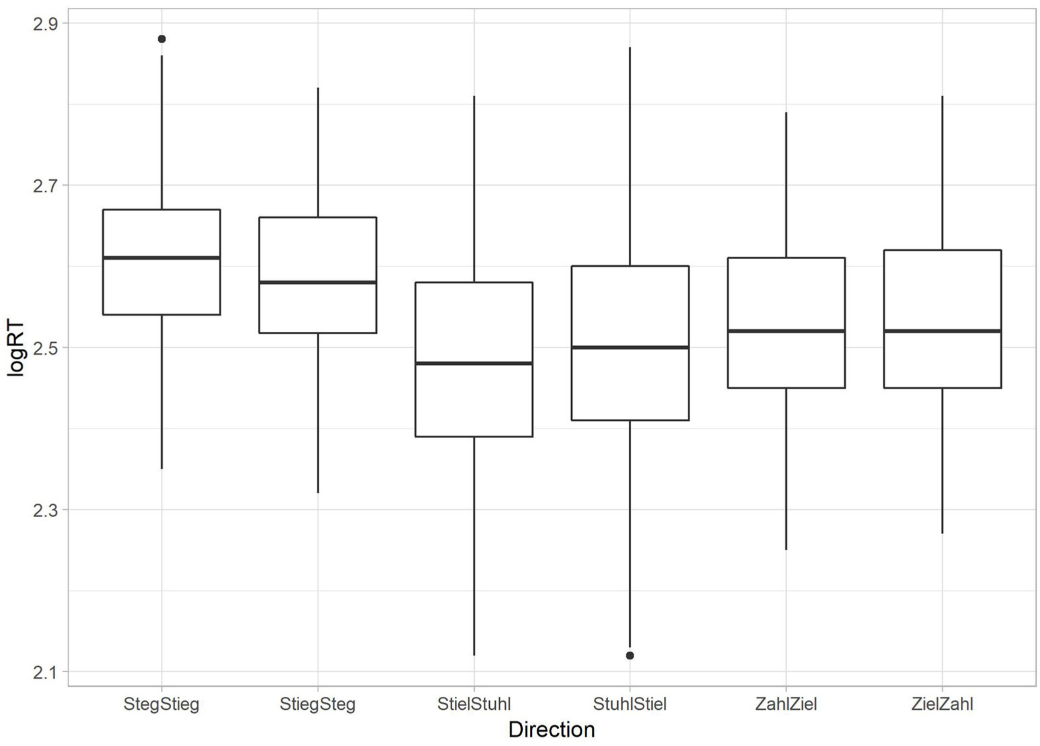


FIGURE 4 | Reaction time results per condition. Reaction time results are given as log values per presentation direction of words with whiskers indicating the variance of the data and small dots representing outliers (but not extreme values) which were beneath the ±2 SD cut-off.

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 (version 16.4)[1] 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 4–11 standards between two deviants. Blocks of the same condition never followed each other.

Analysis and Results

The reaction time analysis was based on correct responses only (98% of data points included). RT were corrected for the onset cluster of each stimulus. Thus, measurement of RT began on the vowel onset. RT faster than 100 ms and slower than 1,000 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 behavioral asymmetries. Here, we found a highly significant main effect ($F_{(5,2,320)} = 107.811$, $p < 0.001$). *Post hoc* analysis revealed asymmetric patterns in two of the tested vowel contrasts. /i:/—/e:/ was significantly faster than vice versa ($F_{(1,464)} = 22.234$, $p < 0.001$). The same was true for /i:/—/u:/ ($F_{(1,464)} = 13.550$, $p < 0.001$). However, in the vowel contrast /i:/—/a:/ a symmetrical pattern of RT occurred. Here, the *post hoc* analysis shows no difference between presentation directions ($F_{(1,464)} = 0.793$, $p = 0.374$; see **Figure 4**).

Discussion

The behavioral study aimed to investigate the basis for some of the electrophysiological found effects in Experiment 1. 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:/

[1]http://www.neurobs.com

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 RT indicate that in an active discrimination paradigm, German natural word stimuli were discriminated phonemically, based on higher-order abstract phonological features.

The RT for the vowel contrast /i:/—/e:/ and the observed asymmetrical pattern match with the predictions of FUL. The faster RT obtained when /e:/ was deviant seem to be due to the underspecification of [COR]. In this contrast, abstract representations may help to discriminate concerning 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 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, RT should be faster when subjects are presented with a fully specified vowel (e.g., /u:/) followed by an underspecified vowel (e.g., /i:/). But concerning the present study, we found the opposite effect for the obtained RT. The hypothesis of NRV for this contrast seems equally unsuitable: it states that asymmetric effect should occur because both vowels are reference vowels. In the case of /u:/ as deviant, one possible explanation for our findings 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, 2001a,b; Wolff and 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). Besides., the effect of stronger STS activations was also seen in reaction time measures whereby reaction time decreased with increasing feature number (Scharinger et al., 2016). Furthermore, in a MEG study, it was shown that N1m amplitudes increased when feature number increased (Scharinger 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 et al., 2011b). Under the assumption of an additivity effect of the phonological feature [LAB], we argue that the underspecification approach still holds 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 good. 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 **Figure 1**), evidence from theoretical analysis (Wiese, 2000), and also neurobiological evidence (Obleser et al., 2004) that Standard German /a:/ is likely not specified for a 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, asymmetry has to occur.

In conclusion, it seems that participants use phonological and phonemic cues in vowel discrimination within natural German words. But the effects found in Experiments 1 and 2 are different although the same experimental paradigm has been applied. The reason for different effect patterns in the electrophysiological and behavioral 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.

EXPLORATIVE ANALYSIS FOR ADDITIONAL INFLUENTIAL FACTORS IN MMN AND LOG RT DATA

Because the interpretation of the MMN and RT data with common models is challenging, and because both models failed to explain the found patterns comprehensively, 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 (Näätänen et al., 1989; Paavilainen et al., 1991; Schröger, 1996; Jacobsen et al., 2003; Peter et al., 2010; Pakarinen et al., 2013). For 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 and Reetz, 2010), they also automatically imply an acoustic difference. Moreover, when words are used as stimuli, lexical features like frequency of occurrence (Alexandrov et al., 2011; Shtyrov et al., 2011) or phonotactic probability (Bonte et al., 2005; Yasin, 2007; Emmendorfer et al., 2020) are known to interfere in speech perception and vowel discrimination. Especially in our approach, where 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 behavioral 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.

Preparation: Rating of Implicit 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 and 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 the 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). 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 and 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.

Materials, Subjects, and Procedure

To test the word stimuli of Experiments 1 and 2 for differences in the perceived (or implicit) loudness, we conducted a rating study. Here, 10 subjects (seven females, three males) participated, all of the 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 before 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 (version 16.4)[1], 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. Afterward, each trial started with a 1,500 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

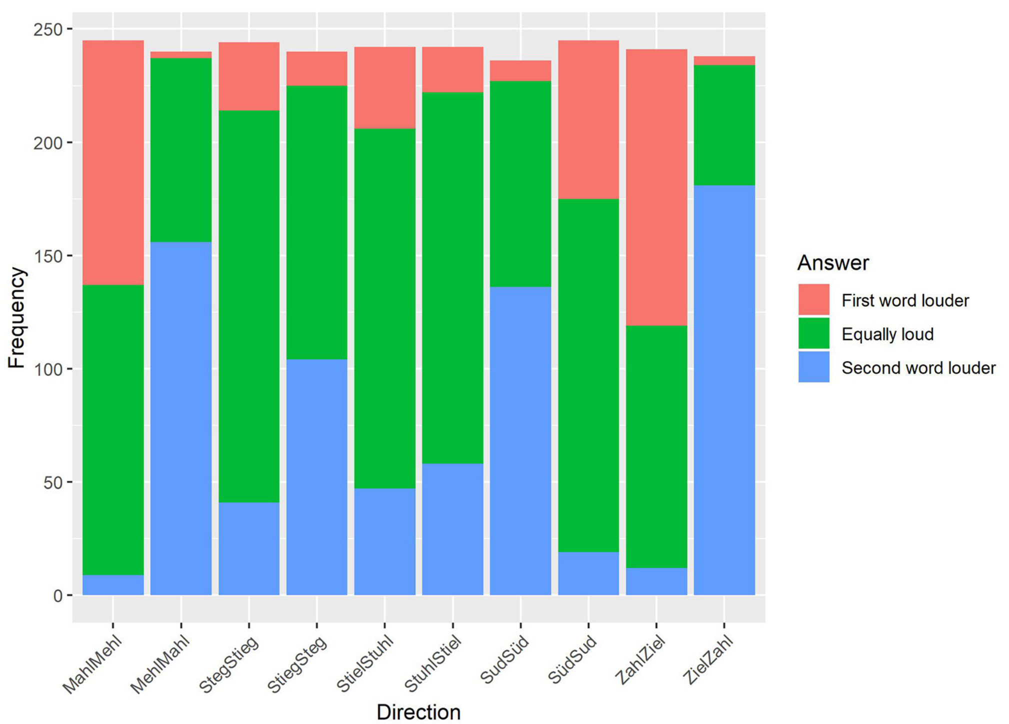


FIGURE 5 | Results of the perceived loudness rating. The results are plotted for each presentation direction (*x*-axis) in relation to the frequency of the given responses (*y*-axis).

TABLE 3 | Distribution of the answers for the perceived loudness of words (in percent) with the most given answer per presentation order in bold.

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**

question marks with a timeout of 2,500 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.

Analysis and Results

Having collected the responses of all participants, frequency values of the three answer categories (first, second, and 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 5**. 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 ($\chi^2_{(18)} = 998.986, p < 0.001$).

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

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, implicit loudness could have a potential influence on the 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 the 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, the result for *Zahl* as deviant. It may be that the higher implicit 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 the 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 implicit loudness affected 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* and *Stieg*, *Stuhl* and *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 influenced the neural effects once more.

Additionally, and following the feedback of participants, it can be hypothesized that implicit 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 the 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.

Explorative Analysis *via* Multiple Regressions

Because of the challenging and unexpected electrophysiological effects that do not match the behavioral results in addition to the descriptive identification of a potential influence of the implicit 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 behavioral level (log RTs).

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 taken into account. 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 concerning both evaluated models in Experiments 1 and 2. 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 to the assumption of NRV that peripheral vowels may be 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 may be 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 a categorical variable (discrimination from a less to a more focal vowel or from a more focal to a less focal vowel).

The empirically motivating factors have been chosen from the 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 the 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, the *degree of openness* (increasing openness of the mouth, decreasing openness, equal openness) was also taken into account.

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 behavioral 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 good 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*: the difference between mean f0 of the deviant minus mean f0 of the standard; *vowel duration*: the difference between vowel duration of the deviant minus the vowel duration of the standard; *intensity*: the difference between mean vowel intensity of the deviant minus mean vowel intensity of the standard).

Last but not least, the *contrast* has been included as a controlling factor, since the iMMN and RT effects were different for the tested vowel oppositions. Here, both presentation orders of each minimal pair are combined.

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 reasonable factors to be included in the final multiple regression analysis.

Correlation analysis showed very strong correlation between the factors *F2* and *specificity* ($\tau = -0.907$, $p < 0.001$) and *F1* and *degree of openness* ($\tau = 0.866$, $p < 0.001$), *vowel duration* and *degree of openness* ($\tau = 0.856$, $p < 0.001$), as well as *F2* and *F3* ($\tau = 0.867$, $p < 0.001$). 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 and 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* ($\tau = -0.764$, $p < 0.001$), *F3* and *specificity* ($\tau = -0.760$, $p < 0.001$), *F1* and *vowel duration* ($\tau = 0.764$, $p < 0.001$), *vowel duration* and *f0* ($\tau = -0.778$, $p < 0.001$), *specificity* and *peripherality* ($\tau = 0.559$, $p < 0.001$), *bigram frequency* and *peripherality* ($\tau = 0.676$, $p < 0.001$), *vowel duration* and *intensity* ($\tau = -0.689$, $p < 0.001$), *f0* and *intensity* ($\tau = 0.556$, $p < 0.001$), *focality* and *F1* ($\tau = -0.603$, $p < 0.001$), *F2* ($\tau = 0.507$, $p < 0.001$) as well as F3 ($\tau = 0.686$, $p < 0.001$), *focality* and *f0* ($\tau = 0.686$, $p < 0.001$), *focality* and *vowel duration* ($\tau = -0.745$, $p < 0.001$), with *intensity* ($\tau = 0.566$, $p < 0.001$), with *degree of openness* ($\tau = -0.731$, $p < 0.001$) and *implicit loudness* ($\tau = -0.654$, $p < 0.001$), *implicit loudness* and *f0* ($\tau = -0.775$, $p < 0.001$), *implicit loudness* and *F1* ($\tau = 0.667$, $p < 0.001$), *implicit loudness* and *vowel duration* ($\tau = 0.660$, $p < 0.001$), *implicit loudness* and *degree of openness* ($\tau = 0.603$, $p < 0.001$), and *implicit loudness* and *intensity* ($\tau = -0.488$, $p < 0.001$).

Single factor regression (with MMN data) revealed the influence of only five significant factors with reasonable R^2 and adjusted R^2 values: *implicit loudness* ($\Delta R^2 = 0.023$, $p < 0.001$), *contrast* ($\Delta R^2 = 0.019$, $p < 0.001$), *intensity* ($\Delta R^2 = 0.016$, $p < 0.001$), *f0* ($\Delta R^2 = 0.012$, $p \leq 0.001$), *vowel duration* ($\Delta R^2 = 0.013$, $p < 0.001$).

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

TABLE 4 | Multiple regression models for the MMN dataset.

Model#		b	SE B	β	p
1	Constant	−0.363 (−0.592, −0.133)	0.117		0.002
	Loudness	−0.436 (−0.607, −0.265)	0.087	−0.155	0.000
$R^2 = 0.024$, $\Delta R^2 = 0.023$, $p < 0.001$					
2	Constant	−0.703 (−0.971, −0.434)	0.137		0.000
	Loudness	−0.436 (−0.605, −0.267)	0.086	−0.155	0.002
	Contrast	0.170 (0.098, 0.242)	0.037	0.143	0.000
$R^2 = 0.044$, $\Delta R^2 = 0.043$, $p < 0.001$					

are approximately an equal fit for the explanation of results. In this study, implicit loudness seems to influence the neural results of the MMN study (see **Figure 6**).

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 behavioral 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 behavioral pattern in the reaction time experiment (see **Figure 7**).

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 of 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 a 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 did not influence the behavioral patterns; therefore, the effects are more likely driven by the traditional models (namely FUL and NRV) discussed before.

GENERAL DISCUSSION AND CONCLUSION

In this article, we reported the results of an electrophysiological and a behavioral 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 was 2-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 behavioral levels. For this purpose, we conducted an in-depth analysis delving into possible confounds to a degree that has not been investigated so far.

To summarize the results of the electrophysiological experiment concerning the first research question, we found MMN evidence for discrimination and perceptual asymmetries (or symmetries) in vowel perception according to the NRV model. 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 behavioral and electrophysiological studies (e.g., Masapollo et al., 2015, 2017b; Zhao et al., 2019) that also report easier discrimination from a more central to a more peripheral vowel. Only one contrast could be explained

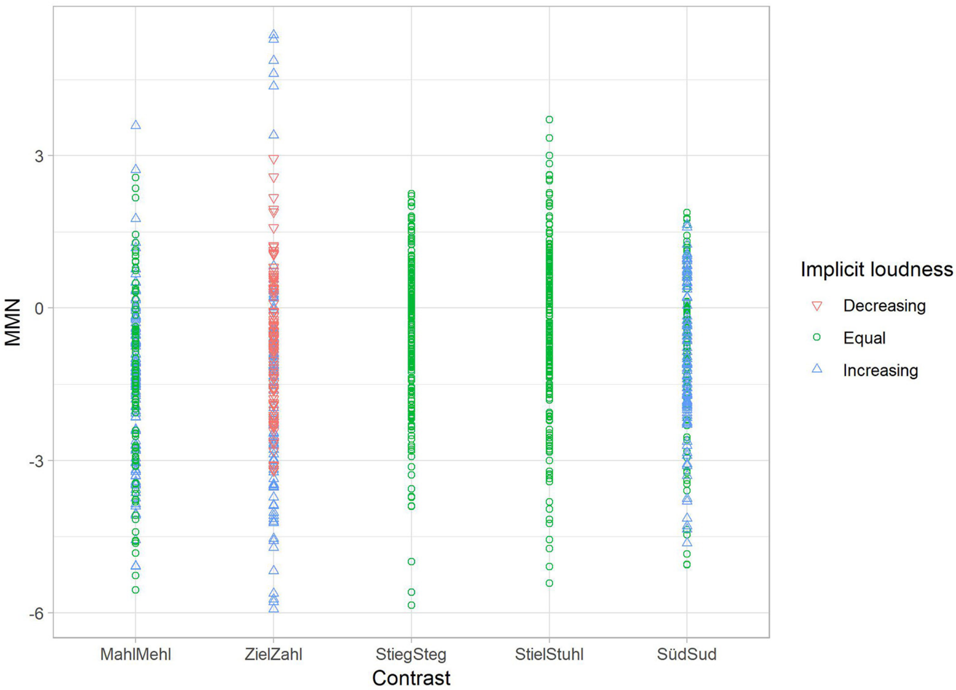


FIGURE 6 | Scatterplot for the regression analysis of the given iMMN data from Experiment 1. MMN difference values of each participant (*y*-axis) per vowel contrast (*x*-axis) in relation to implicit loudness. Increasing loudness (deviant louder than standard) is shown as a blue triangle, decreasing 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).

TABLE 5 | Multiple regression models for the reaction time (RT) dataset (n.s. = no significance).

Model#		b	SE B	β	*p*
1	Constant	2.547 (2.522, 2.573)	0.013		0.000
	Loudness	0.001 (−0.021, 0.024)	0.011	0.010	n.s.
$R^2 = 0.000$, $\Delta R^2 = -0.006$, n.s.					
2	Constant	2.558 (2.535, 2.595)	0.015		0.000
	Loudness	0.001 (−0.020, 0.023)	0.011	0.127	n.s.
	Contrast	−0.021 (−0.033′6, −0.005)	0.008	−0.204	0.011
$R^2 = 0.041$, $\Delta R^2 = 0.029$, $p < 0.05$					

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 et al., 2004b) 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 the subjects were not instructed to perform phonemic discrimination, but to ignore the stimulation (Johnson, 2015). To test those challenging results, we investigated these contrasts with a behavioral 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 behavioral 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 and Lahiri, 2004; Lahiri and Reetz, 2010; Scharinger et al., 2012a,b).

In summary, the results of Experiments 1 and 2 are challenging in two ways. First, the neural pattern cannot be explained comprehensively by either of the two models. Second, the neural and behavioral patterns do not match. The lack of compliance between electrophysiological and behavioral 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

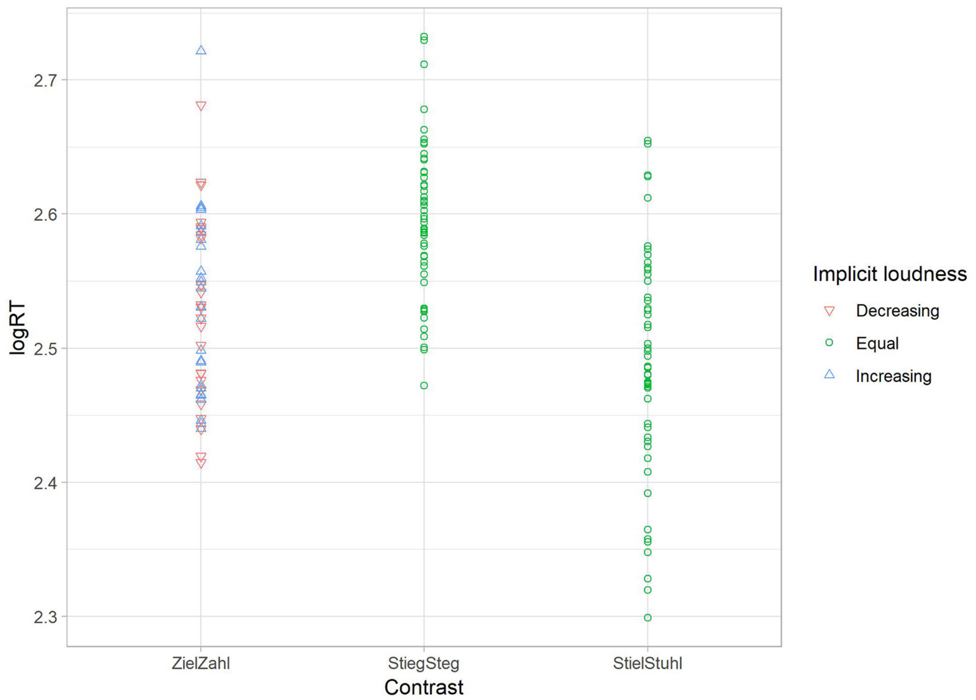


FIGURE 7 | Scatterplot for the regression of the obtained reaction time (RT) data from Experiment 2. Mean log RTs of each subject (*y*-axis) are depicted per vowel contrast (*x*-axis) in relation to implicit loudness. RT results (log-values) are not scaled by the perceived loudness of the stimuli.

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 concerning the richness of mental representation accessible during discrimination as a function of attention were also found by Tuomainen et al. (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 behavioral results indicated phonemic discrimination of the vowels, the preattentive MMN patterns were more driven by acoustic differences than phonemic representations. Concerning our electrophysiological and behavioral 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 following 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 behavioral datasets. We included several theoretical, lexical, and phonotactic 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; Yasin, 2007; Emmendorfer et al., 2020) or the lexical frequency of words (Alexandrov et al., 2011; Shtyrov et al., 2011; Aleksandrov et al., 2017), we cannot provide evidence for those factors neither on the electrophysiological nor on the behavioral 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 behavioral 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 behavioral data. Once again, these results can be interpreted as evidence that in preattentive processing, more perceptual and acoustic features are responsible for the 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), 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 the degree of openness of the vowels and changes in F1. We conclude that perceived loudness differences could be guided by differences in the 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 and 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 and 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 MMN data regarding natural vowel processing. We propose that the perceptual, or implicit, the 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 underinvestigated 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 analyzing and interpreting neural and behavioral data.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Society of German Linguistics (DGfS). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MR contributed to the design of the work, acquisition, analysis, and interpretation of the data, as well as drafting of the work. AW contributed to the design of the work, acquisition, and revising the manuscript. AN contributed to the design of the work, acquisition and analysis of the data, as well as the revising of the manuscript. MS contributed to the design of the work, analysis and interpretation of the data, as well as drafting and revising of the manuscript. All authors gave final approval of the version to be submitted and agreed to be accountable for all aspects of the work ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the article and approved the submitted version.

FUNDING

The research presented here was supported by the Hessen State Ministry of Higher Education, Research and the Arts, Landesoffensive Zur Entwicklung Wissenschaftliche-ökonomischer Excellenz (LOEWE), Research Focus: exploring fundamental linguistic categories grant, Research Project: phonological word—Constituents of the Phonological Word, and the Johannes Gutenberg University Mainz.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.612345/full#supplementary-material.

REFERENCES

Aaltonen, O., Eerola, O., Lang, A. H., Uusipaikka, E., and Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *J. Acoust. Soc. Am.* 96, 1489–1493. doi: 10.1121/1.410291

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

Aleksandrov, A. A., Memetova, K. S., Stankevich, L. N., and Uplisova, K. O. (2017). Effects of Russian-language word frequency on mismatch negativity in auditory event-related potentials. *Neurosci. Behav. Phys.* 47, 1043–1050. doi: 10.1007/s11055-017-0510-3

Alexandrov, A. A., Boricheva, D. O., Pulvermüller, F., and Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS One* 6:e22999. doi: 10.1371/journal.pone.0022999

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

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

Bohn, O.-S., and Polka, L. (2001). Target spectral, dynamic spectral and duration cues in infant perception of German vowels. *J. Acoust. Soc. Am.* 110, 504–515. doi: 10.1121/1.1380415

Bonte, M., Mitterer, H., Zellagui, N., Poelmans, H., and Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clin. Neurophysiol.* 116, 2765–2774. doi: 10.1016/j.clinph.2005.08.012

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., and Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp. Psychol.* 58, 412–424. doi: 10.1027/1618-3169/a000123

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862

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

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

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

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8, 919–924. doi: 10.1097/00001756-199703030-00021

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

Endrass, T., Mohr, B., and Pulvermuller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *Eur. J. Neurosci.* 19, 1653–1660. doi: 10.1111/j.1460-9568.2004.03247.x

Eriksson, A., and Traunmüller, H. (1999). Percetion of vocal effort and speaker distance on the basis of vowel utterances. *Percept. Psychophys.* 64, 2469–2472. doi: 10.3758/bf03194562

Eriksson, A., and Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Percept. Psychophys.* 64, 131–139. doi: 10.3758/bf03194562

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Fletcher, H., and Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J. Acous. Soc. Am.* 5, 82–108. doi: 10.1002/j.1538-7305.1933.tb00403.x

Friedrich, C. K., Lahiri, A., and Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1545–1559. doi: 10.1037/a0012481

Glave, R. D., and Rietveld, A. C. M. (1975). Is the effort dependence of speech loudness explicable on the basis of acoustical cues? *J. Acoust. Soc. Am.* 58, 875–879. doi: 10.1121/1.380737

Glave, R. D., and Rietveld, A. C. M. (1979). Bimodal cues for speech loudness. *J. Acoust. Soc. Am.* 66, 1018–1022. doi: 10.1121/1.383320

Hamamura, M., and Iwamiya, S. (2016). Relationship of the differences in perceived loudness of sound and the optimum listening level between males and females. *Acoust. Sci. Tech.* 37, 40–41. doi: 10.1250/ast.37.40

Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U S A* 109, 20095–20100. doi: 10.1073/pnas.1213390109

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

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

Højlund, A., Gebauer, L., McGregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. *Lang. Cogn. Neurosci.* 34, 545–560. doi: 10.1080/23273798.2019.1572204

Hsu, C.-H., Evans, J. P., and Lee, C.-Y. (2015). Brain responses to spoken F 0 changes: is H special? *J. Phon.* 51, 82–92. doi: 10.1016/j.wocn.2015.02.003

Jacobsen, T., Horenkamp, T., and Schröger, E. (2003). Preattentive memory-based comparison of sound intensity. *Audiol. Neurootol.* 8, 338–346. doi: 10.1159/000073518

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., Winkler, I., et al (2004). Pre-attentive auditory processing of lexicality. *Brain Lang.* 88, 54–67. doi: 10.1016/s0093-934x(03)00156-1

Johnson, K. (2015). Vowel perception asymmetry in auditory and phonemic listening. *UC Berkeley PhonLab Annual Report*, 11:26. Available online at: https://escholarship.org/uc/item/21t337gh.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schroger, E., and 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. *Int. J. Psychophysiol.* 67, 131–143. doi: 10.1016/j.ijpsycho.2007.10.012

Kriengwatana, B. P., and Escudero, P. (2017). Directional asymmetries in vowel perception of adult nonnative listeners do not change over time with language experience. *J. Speech Lang. Hear. Res.* 60, 1088–1093. doi: 10.1044/2016_JSLHR-H-16-0050

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Laboratory Phonology*, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 637–676.

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phon.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

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

Lawyer, L. A., and Corina, D. P. (2018). Putting underspecification in context: ERP evidence for sparse representations in morphophonological alternations. *Lang. Cogn. Neurosci.* 33, 50–64. doi: 10.1080/23273798.2017.1359635

Lipski, S. C., Lahiri, A., and Eulitz, C. (2007). Differential height specification in front vowels for German speakers and Turkish-German bilinguals: an electroencephalographic study," in *Proceedings of the International Congress of Phonetic Sciences XVI*, Saarbrücken, 809–812.

Masapollo, M., Polka, L., and Ménard, L. (2015). Asymmetries in vowel perception: effects of formant convergence and category "goodness." *J. Acoust. Soc. Am.* 137, 2385–2385. doi: 10.1016/S0167-6393(02)00105-X

Masapollo, M., Polka, L., and Ménard, L. (2017a). A universal bias in adult vowel perception—by ear or by eye. *Cognition* 166, 358–370. doi: 10.1159/000512876

Masapollo, M., Polka, L., Molnar, M., and Ménard, L. (2017b). Directional asymmetries reveal a universal bias in adult vowel perception. *J. Acoust. Soc. Am.* 141, 2857–2869. doi: 10.1121/1.4981006

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., and Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: the roles of oral-facial kinematics, orientation and configuration. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1103–1118. doi: 10.1037/xhp0000518

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

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

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, 1–21. doi: 10.1017/s0048577201000208

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

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

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

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

Obleser, J., Lahiri, A., and Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.* 16, 31–39. doi: 10.1162/089892904322755539

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9. doi: 10.1155/2011/156869

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., and Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalogr. Clin. Neurophysiol.* 78, 466–479. doi: 10.1016/0013-4694(91)90064-b

Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., et al. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity (MMN). *Int. J. Psychophysiol.* 87, 103–110. doi: 10.1016/j.ijpsycho.2012.11.010

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

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

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., et al. (2004a). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear Hear.* 25, 284–301. doi: 10.1097/01.aud.0000130800.88987.03

Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., et al. (2004b). Processing of English words with fine acoustic contrasts and simple tones: a mismatch negativity study. *J. Am. Acad. Audiol.* 15, 47–66. doi: 10.3766/jaaa.15.1.6

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567

Politzer-Ahles, S., Schluter, K., Wu, K., and Almeida, D. (2016). Asymmetries in the perception of mandarin tones: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1547–1570. doi: 10.1037/xhp0000242

Polka, L., and Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J. Acoust. Soc. Am.* 100, 577–592. doi: 10.1121/1.415884

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *J. Phon.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Pons, F., Albareda-Castellot, B., and 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 Dev.* 83, 965–976. doi: 10.1111/j.1467-8624.2012.01740.x

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage* 14, 607–616. doi: 10.1006/nimg.2001.0864

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

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

Ratcliff, R. (1993). Methods for dealing with reaction times outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510

Sams, M., Paavilainen, P., and Alho, K. (1985). Auditory freqeuncy discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., and Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Cogn. Brain Res.* 16, 250–256. doi: 10.1016/s0926-6410(02)00280-x

Scharinger, M. (2009). Minimal representations of alternating vowels. *Lingua* 119, 1414–1425. doi: 10.1177/0023830908099071

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., and Obleser, J. (2012a). A sparse neural code for some speech sounds but not for others. *PLoS One* 7, e409531–e4095311. doi: 10.1371/journal.pone.0040953

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

Scharinger, M., Idsardi, W. J., and Poe, S. (2011a). A comprehensive three-dimensional cortical map of vowel space. *J. Cogn. Neurosci.* 23, 3972–3982. doi: 10.1162/jocn_a_00056

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

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

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012b). Asymmetries in the processing of vowel height. *J. Speech Lang. Hear. Res.* 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of English, fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Schluter, K. T., Politzer-Ahles, S., Al Kaabi, M., and Almeida, D. (2017). Laryngeal features are phonetically abstract: mismatch negativity evidence from Arabic, English, and Russian. *Front. Psychol.* 8:746. doi: 10.3389/fpsyg.2017.00746

Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology* 32, 55–65. doi: 10.3390/ph14020079

Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalogr. Clin. Neurophysiol.* 100, 517–526. doi: 10.1016/S0168-5597(96)95576-8

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., and Vallée, N. (2005). Asymmetries in vowel perception, in the context of the dispersion-focalisation theory. *Speech Commun.* 45, 425–434. doi: 10.1016/j.specom.2004.12.001

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *J. Phonetics* 25, 255–286. doi: 10.1006/jpho.1997.0043

Sendlmeier, W. F., and Seebode, J. (2006). *Formantkarten des deutschen Vokalsystems*. Available online at: http://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf.

Shtyrov, Y., Kimppa, L., Pulvermüller, F., and 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, 658–668. doi: 10.1016/j.neuroimage.2010.12.002

Shtyrov, Y., Osswald, K., and Pulvermuller, F. (2008). Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. *Cereb. Cortex* 18, 29–37. doi: 10.1093/cercor/bhm028

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

Simon, E., Sjerps, M. J., and Fikkert, P. (2014). Phonological representations in children's native and non-native lexicon. *Biling. Lang. Cognit.* 17, 3–21. doi: 10.1017/S1366728912000764

Strauß, A., Kotz, S. A., Scharinger, M., and Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *NeuroImage* 97, 387–395. doi: 10.1177/1740774519877851

Szymanski, M. D., Yund, E. W., and Woods, D. L. (1999). Phonemes, intensity and attention differential effects on the mismatch negativity (MMN). *J. Acoust. Soc. Am.* 106, 3492–3505. doi: 10.1121/1.428202

Takegata, R., and Morotomi, T. (1999). Integrated neural representation of sound and temporal features in human auditory sensory memory: an event-related potential study. *Neurosci. Lett.* 274, 207–210. doi: 10.1016/s0304-3940(99)00711-9

Takegata, R., Paavilainen, P., Näätänen, R., and Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neurosci. Lett.* 266, 109–112. doi: 10.1016/s0304-3940(99)00267-0

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

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

Tervaniemi, M., Schröger, E., Saher, M., and Näätänen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans—a mismatch negativity study. *Neurosci. Lett.* 290, 66–70. doi: 10.1016/s0304-3940(00)01290-8

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

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

Tyler, M. D., Nil, M. D., Nil, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21. doi: 10.1159/000356237

Whelan, R. (2008). Effective analysis of reaction time data. *Psychol. Rec.* 58, 475–482. doi: 10.1007/BF03395630

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

Wolff, C., and Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double and triple deviations as revealed by mismatch negativity additivity. *Neurosci. Lett.* 311, 37–40. doi: 10.1016/s0304-3940(01)02135-8

Yanushevskaya, I., Gobl, C., and Ní Chasaide, A. (2013). Voice quality in affect cueing: does loudness matter? *Front. Psychol.* 4:335. doi: 10.3389/fpsyg.2013.00335

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

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., and 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 Lang.* 194, 77–83. doi: 10.1016/j.bandl.2019.05.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/fnhum.2021.609898

Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants

Zhanao Fu[1] *and Philip J. Monahan*[1,2,3]*

[1] *Department of Linguistics, University of Toronto, Toronto, ON, Canada,* [2] *Department of Language Studies, University of Toronto Scarborough, Toronto, ON, Canada,* [3] *Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada*

OPEN ACCESS

Edited by:
Valerie L. Shafer,
The City University of New York, United States

Reviewed by:
Yan H. Yu,
St. John's University, United States
Hia Datta,
Molloy College, United States

Correspondence:
Philip J. Monahan
philip.monahan@utoronto.ca

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *24 September 2020*
Accepted: *23 February 2021*
Published: *24 March 2021*

Citation:
Fu Z and Monahan PJ (2021) Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants. Front. Hum. Neurosci. 15:609898. doi: 10.3389/fnhum.2021.609898

How speech sounds are represented in the brain is not fully understood. The mismatch negativity (MMN) has proven to be a powerful tool in this regard. The MMN event-related potential is elicited by a deviant stimulus embedded within a series of repeating standard stimuli. Listeners construct auditory memory representations of these standards despite acoustic variability. In most designs that test speech sounds, however, this variation is typically intra-category: All standards belong to the same phonetic category. In the current paper, inter-category variation is presented in the standards. These standards vary in manner of articulation but share a common phonetic feature. In the standard retroflex experimental block, Mandarin Chinese speaking participants are presented with a series of "standard" consonants that share the feature [retroflex], interrupted by infrequent non-retroflex deviants. In the non-retroflex standard experimental block, non-retroflex standards are interrupted by infrequent retroflex deviants. The within-block MMN was calculated, as was the identity MMN (iMMN) to account for intrinsic differences in responses to the stimuli. We only observed a within-block MMN to the non-retroflex deviant embedded in the standard retroflex block. This suggests that listeners extract [retroflex] despite significant inter-category variation. In the non-retroflex standard block, because there is little on which to base a coherent auditory memory representation, no within-block MMN was observed. The iMMN to the retroflex was observed in a late time-window at centro-parieto-occipital electrode sites instead of fronto-central electrodes, where the MMN is typically observed, potentially reflecting the increased difficulty posed by the added variation in the standards. In short, participants can construct auditory memory representations despite significant acoustic and inter-category phonological variation so long as a shared phonetic feature binds them together.

Keywords: mismatch negativity (MMN), retroflex, Chinese, EEG – electroencephalogram, speech perception, phonology, phonetics, phonetic features

INTRODUCTION

Speech is a variable and continuous signal. Despite this, successful spoken word recognition requires listeners to identify and extract meaningful linguistic units. Different models rely on different linguistic units, from syllables (Greenberg, 1999) to phonological features (Stevens, 2002) to a combination of both (Hickok, 2014). In particular, features play a central role in several models of speech processing (Halle and Stevens, 1962; McClelland and Elman, 1986; Stevens, 2002; Gow, 2003; Poeppel and Monahan, 2011). Stevens (2002) proposed that listeners utilize features to identify major landmarks in the speech signal. The identification of these features is critical for word segmentation and lexical access. Despite their central role in speech processing models and phonological theory (Clements and Hume, 1995; Halle, 2002), evidence that the perceptual system or human brain utilizes features or feature-like representations has been difficult to establish. Their best support has arisen from neurophysiology (see Monahan, 2018 for a review).

Individual speech sounds are a complex constellation of articulatory and acoustic properties. Phonological theory has long represented these properties with distinctive features (Jakobson et al., 1961; Chomsky and Halle, 1968; Trubetskoy, 1969; Clements, 1985; McCarthy, 1988). Features encode the relation between an aspect of a speech sound's articulation and the corresponding acoustic signature (Halle, 1983; Baković, 2014). Moreover, they also serve to denote active natural classes, that is, sets of sounds that pattern together in the phonological grammar. Initially, features were binary in nature, and each feature had a polarity: A consonant was either [+obstruent] or [−obstruent] (Chomsky and Halle, 1968). These binary feature systems, however, wrongly predicted that both positive and negative specifications should denote active natural classes in the grammar (van der Hulst, 2016). For example, some languages have word-final devoicing of voiced obstruents (e.g., German, Dutch), while others allow both voiced and voiceless obstruents in word-final position (e.g., English). No language, as far as we know, employs a word-final voicing rule (Reiss, 2017; although see Blevins et al., 2020 for a potential recent counterexample).

As such, privative, or monovalent, features were proposed (Clements, 1985; Sagey, 1986; Harris and Lindsey, 1995; Lombardi, 1995). In a privative system, a nasal segment contains the feature [nasal], whereas a non-nasal segment completely lacks a representation for nasality in memory. Underspecification accounts go one step further and posit that only predictable privative features are stored (Mester and Itô, 1989; Steriade, 1995). As an example, the place of articulation for the coronal nasal segment [n] is often determined by its local phonotactic context, as it assimilates in place to match the following consonant; otherwise, it has the putative default place of articulation [coronal]. Given this predictability, the feature [coronal] is argued to be underspecified in memory (Archangeli, 1988; Avery and Rice, 1989).

Neurophysiological measures [e.g., electroencephalography (EEG), electrocorticography (ECoG), magnetoencephalography (MEG)] have been used to assess the nature of speech sound representations (Näätänen, 2001; Mesgarani et al., 2014). An extensively used method is the mismatch negativity (MMN, mismatch field (MMF/MMNm) in MEG; Näätänen et al., 2007; Näätänen and Kreegipuu, 2012). The MMN is a negative deflection in the event-related potential (ERP) to an infrequent deviant stimulus embedded within a sequence of repeating standard stimuli. It peaks between 100 ms and 400 ms post-stimulus onset and in EEG, is largest over fronto-central electrode sites. Auditory cortex is the cortical source of the auditory MMN, and its precise location depends on the property of the deviant that differs from that of the standards (e.g., frequency, intensity, duration; see Alho, 1995 for a review). For speech stimuli, the MMN localizes to supratemporal auditory cortex (Aulanko et al., 1993). In studies of speech perception, the "varying standards" paradigm is often utilized. There, different acoustic tokens of the standard are used that all belong to the same speech sound category. This encourages participants to construct auditory memory representations of the standards that are not based solely on acoustic properties but instead reflect phonetic or phonological categories (Phillips et al., 2000; Kazanina et al., 2006; Hestvik et al., 2020).

A listener's native language phonology modulates the size and/or presence of the MMN (Näätänen et al., 1997; Winkler et al., 1999, 2003; Sharma and Dorman, 2000; Kazanina et al., 2006; Nenonen et al., 2003; Ylinen et al., 2006; K. Yu et al., 2019). Additionally, various MMN results support a role for phonological features during speech processing (Eulitz and Lahiri, 2004; Scharinger et al., 2012; Cornell et al., 2013; Hestvik and Durvasula, 2016; Scharinger et al., 2016b; Schluter et al., 2016). In a number of these studies, an asymmetric MMN is observed. In most MMN designs, two categories are tested, and participants are presented with two experimental blocks separated by a short break in a single testing session. In the first experimental block, one category is the standard while the other is the deviant. This role is reversed in the second experimental block. The asymmetry is that one deviant elicits a larger MMN than the other deviant. These asymmetries are often taken to reflect the underlying featural content of the two categories consistent with underspecified representations (Lahiri and Reetz, 2002, 2010). A larger MMN is observed when the standard is specified for a given feature and the deviant mismatches with that feature. When the standard is underspecified, there is no mismatch between the standard and deviant, and as such, a smaller or no MMN is observed. Asymmetric MMN results have been observed for vowels (Eulitz and Lahiri, 2004; Cornell et al., 2011; Scharinger et al., 2012, 2016b), consonants (Cornell et al., 2013; Hestvik and Durvasula, 2016; Schluter et al., 2016; Hestvik et al., 2020) and lexical tones (Politzer-Ahles et al., 2016).

In all these studies, however, a single category is used for the standards. In the current paper, we present multiple different phonetic categories in the standards that all share the common phonetic feature [retroflex]. Gomes et al. (1995) reported that when sinusoidal standards shared the same duration but varied in intensity and frequency, an MMN was still observed to deviants that differed along all three parameters. These findings suggest that listeners can extract single cues, in this case duration, from the standards and build an auditory memory representation based on a single cue. Most phonetic and phonological features refer to a single cue that denotes natural classes of sounds (Halle, 2002), while some features refer to multiple acoustic cues that denote a single natural class (e.g., retroflex; Hussain et al., 2017). The question this paper addresses is whether listeners also extract features from standards that belong to the same natural class and use those features to construct auditory memory representations.

Standard Mandarin Chinese has a relatively rich set of retroflex consonants at the coronal place of articulation. In particular, Mandarin Chinese has the fricative [ʂ], affricate [tʂ], aspirated affricate [tʂʰ], and a final category [ʐ∼ɻ]. This last category has been argued to be a voiced fricative (Duanmu, 2007), while others have argued that phonetically, it is an approximant (Lee and Zee, 2003; Lee-Kim, 2014; Lin, 2007). Mandarin Chinese also has the non-retroflex coronal counterparts for each of these categories: [s], [ts], [tsʰ] and [l], respectively. Here, each of these sound categories is presented both as standards and deviants, which makes the current study a unique departure from traditional MMN studies, where intra-category stimulus tokens are used.

The current paper describes the results of a single MMN experiment using EEG with Mandarin Chinese retroflex consonants. Most previous MMN studies of features assume a privative feature system and argue for underspecified featural representations (see above). Here, we spell out the predictions for both binary and privative accounts assuming listeners can extract constant phonetic properties despite inter-category variation in the standards. Overall, we attempt to determine whether the property (or lack thereof) retroflex, which gives speech sounds an "r"-color, is extracted by Mandarin listeners and used to construct an auditory memory representation. In the retroflex standard experimental block, listeners heard the standards [ʂ tʂ tʂʰ ɻ] interrupted by an occasional deviant stimulus, e.g., [s ts tsʰ l]. In the non-retroflex standard experimental block, the standard-deviant relationship was reversed. All segments used in the experiment are [coronal] and as such, the feature [coronal] is insufficient to explain the presence of an MMN. In a binary feature account, listeners should extract [+retroflex] in the retroflex standard block and [−retroflex] in the non-retroflex standard block. Equal-sized MMNs are predicted for both blocks as positive and negative feature valences are equally informative. Under a privative account, however, only retroflex consonants are stored with the feature [retroflex] in their auditory memory representation. Then, we predict an asymmetric MMN. We anticipate an MMN only in the retroflex standard block as listeners extract the feature [retroflex]. Because the standards in the non-retroflex standard block do not share a common feature to the exclusion of the deviants under a privative account, we do not anticipate observing a clear MMN.

METHODS

Participants

Thirty-three right-handed native Mandarin speakers participated in the experiment. All subjects were recruited from the University of Toronto Scarborough. The entire experimental session was conducted in Mandarin Chinese. No participant reported any hearing, language, or neurological deficits. Data from seven participants were excluded due to technical issues during the recording sessions. This left 26 participants (17 females, mean ($\bar{x}$) age = 19.7 years, standard deviation (s) = 0.7 years, $\bar{x}$ age of arrival = 16.8 years, s = 2.3 years). All participants also spoke English. Nine participants reported proficiency in additional languages (i.e., Cantonese, Shanghainese, and Japanese). All participants self-reported 10/10 on listening proficiency and at least 8/10 in speaking proficiency in Mandarin, except for one participant, who self-rated 7/10 in listening proficiency. The Mandarin participants reported consistently lower speaking and listening English self-ratings (speaking: $\bar{x}$ = 6.54, s = 1.64; listening: $\bar{x}$ = 7.43, s = 1.56) and only reported using English on average 23.7% (s = 18%) of the time in their daily lives. The mean length of stay in Canada was 3.02 years (s = 2.63 years). The experiment was approved by the University of Toronto Research Ethics Board. All participants provided written informed consent and received course credit.

Stimuli

Stimuli included eight [Cɤː4] syllables. The C represents a consonant from four retroflex/non-retroflex consonant pairs: [s]/[ʂ], [ts]/[tʂ], [tsʰ]/[tʂʰ], [l]/[ɻ]. The eight consonants represent every retroflex consonant in the Mandarin inventory and their non-retroflex counterparts. The high-mid back unrounded vowel [ɤ] and Tone 4 were chosen to form the frame because they yield phonotactically legal syllables with all consonants in the experiment, and all were real words of Mandarin Chinese. For example, [sɤː4] corresponds to the words 色 "color" or 涩 "astringent". Some [Cɤː4] stimuli corresponded to multiple lexical items. The decision was taken to use real words. The relatively limited Mandarin syllable inventory made it impossible to obtain a set of CV syllables where each was a phonotactically legal pseudoword in the language across all eight consonant, vowel and tone combinations. The alternatives were to use a set of stimuli that contained a mixture of words and pseudowords or a set of eight phonotactically illegal syllables. Choosing a mixture of words and pseudowords would potentially result in evoking a set of processes (e.g., lexical access) that would be present for some items but not others. Meanwhile, using phonotactically illegal syllables would require participants to employ repair strategies that are not part of natural language processing in their native language (Dehaene-Lambertz et al., 2000).

Stimuli were produced by a male native speaker of Mandarin Chinese. The tokens were recorded with an Audio-Technica AT3035 cardioid microphone onto a MixPre-3 digital recorder (Sound Devices LLC, United States). The stimulus recording session occurred in a sound-attenuated booth. The audio files were recorded with a 44.1 kHz sampling frequency at 16-bit depth. The retroflex stimuli had a mean duration of 429 ms (s = 9 ms). The non-retroflex stimuli had a mean duration of 459 ms (s = 81 ms). **Table 1** provides the syllable and consonant durations for the stimuli in our experiment, the number of words and senses for each syllable, and their corpus frequencies. The number of words and senses are obtained from Xinhua dictionary (The Commercial Press, 2009). Word frequencies are obtained from DoWLS (Neergaard et al., 2016), which is based on SUBTLEX-CH (Cai and Brysbaert, 2010) and includes phonetic transcriptions. Cosine^2 offset ramps were applied to the final 10 ms of each stimulus. Stimulus intensity was normalized to 70 dB SPL.

As demonstrated in **Figure 1**, the distinctive acoustic feature between retroflex/non-retroflex fricative is the range of spectral energy. Retroflection is associated with lower spectral energy ranges in both fricatives and affricates (Lee, 1999). In the approximant pair [l]/[ɻ], retroflection leads to a lower F3, which is similar to the differences between prevocalic /l/ and /r/ in American English (Polka and Strange, 1985). A lower F3 correlates with more "r"-color. The retroflex [ɻ] also has a larger difference between F4 and F5 than [l].

EEG Acquisition

Subjects were seated in a sound-attenuated cabin. They were instructed to watch a silent movie while passively listening to the stimuli during the EEG recording (Tervaniemi et al., 1999; Scharinger et al., 2016b). Experimenters communicated with participants in Mandarin, and all experiment materials (i.e., instructions, recruitment and debriefing materials) were provided in Mandarin. Stimuli were presented in an auditory oddball paradigm. The experiment consisted of two blocks. One block contained retroflex standards and non-retroflex deviants. The other block had non-retroflex standards and retroflex deviants. The order of blocks was randomized for each subject. In each block, each of the four tokens of the deviant category was presented 20 times, totaling 80 deviant tokens per block. The order of deviant tokens was randomized. Prior to each deviant stimulus, a random number (drawn from a uniform distribution between 4 and 10) of standard tokens were presented. This resulted in an approximate standard-to-deviant ratio of 7-to-1. There were approximately 560 standard tokens per block. The standard tokens were also randomly sampled from the four stimuli of its category. The duration of the interstimulus interval was randomly sampled from a uniform distribution between 1.25 and 1.75 s. These values were selected to reinforce phonological-level processing (Werker and Logan, 1985; Yu Y. H. et al., 2017).

Continuous EEG signals were acquired with 32-channel ActiCAP active electrodes (Brain Products GmbH, Germany) and an actiCHamp (Brain Products GmbH, Germany) amplifier. The data were digitalized at 1000 Hz with a 0.01–500 Hz online bandpass filter. Electrodes were placed according to the international 10-20 system and positions include Fp1/2, F3/4, F7/8, FC1/2, FC5/6, FT9/10, C3/4, T7/8, CP1/2, CP5/6, TP9/10, P3/4, P7/8, O1/2, Oz, Fz, Cz, and Pz. A ground electrode was placed at Fpz. The continuous EEG signal was referenced to the left mastoid (TP9) online.

To ensure precise stimulus-digital trigger timing, auditory stimuli were first passed through a StimTrak device (Brain Products GmbH, Germany), which is engineered specifically for EEG trigger precision. In our configuration, the StimTrak device forward the auditory signal simultaneously to the amplifier and headphones. The auditory signal sent to the amplifier is recorded as an additional EEG channel. This provides the moment at which the auditory stimulus is presented to the participants. The auditory stimuli were delivered to subjects through BeyerDynamic DT 770 PRO headphones.

EEG Analysis

Data analysis was conducted in MATLAB (Mathworks, Inc) using the EEGLAB toolbox (Delorme and Makeig, 2004). First, we corrected for any offset delays between the trigger and the auditory stimulus presentation to ensure millisecond-precise stimulus-digital trigger synchrony. This was done by cross-correlating the original stimuli sound files and the audio track in the EEG recording delivered by StimTrak. Subsequently, triggers were aligned with the onset of audio file in the EEG recording. Trigger-stimulus onset synchrony was checked in the raw continuous EEG signal. Next, the EEG signal was re-referenced to the linked mastoids, which provide the most robust MNN responses (Mahajan et al., 2017). The EEG signal was then filtered with a Hamming windowed sinc FIR filter. The signal was first high-pass filtered at 0.1 Hz (transition bandwidth 0.1 Hz), then low-pass filtered at 70 Hz (transition bandwidth 17.5 Hz). The filtered signal was downsampled to 250 Hz. The PREP pipeline (Bigdely-Shamlo et al., 2015) was subsequently used to remove line noise and bad channels. Then, the artifact subspace reconstruction (ASR[1] algorithm was applied to remove stationary artifacts. On average, 2.3 channels were removed as bad channels. Next, previously removed channels were interpolated. An independent component analysis (ICA) decomposition was done by Adaptive Mixture Independent Component Analysis (AMICA; Palmer et al., 2012). Dipoles of the independent components were localized with the DIPFIT plug-in (Oostenveld and Oostendorp, 2002)[2]. We manually inspected the topography, fitted dipole locations, waveforms, and residual variances to identify the independent components. Independent components that correspond to eye movements (e.g., blinks, saccades) or widely distributed artifacts on the scalp were removed. Fewer than six ($\bar{x}$ = 2.1) independent components were removed for each subject. Next, the continuous EEG data were epoched with a 1-second pre-onset period and a 2-second post-onset period. Epochs with voltages $\pm$ 75 μV were rejected. Combined with the window rejection from the application of ASR, fewer than 3% of trials were removed for each subject. Among the standard trials, the two trials immediately after each deviant trial were excluded from further analysis. This was done to ensure that only trials where participants heard a sequence of standards prior to a deviant were included in the analysis. After preprocessing, each subject had more than 357 standards and more than 76 deviants in each of the two blocks for the final analysis.

Within-block MMN analyses compare the evoked potentials to the deviant and standard stimuli in the same experimental block. Different properties of the standards and deviants could elicit different ERPs even without the oddball frequency differences. This could potentially confound the MMN analysis. An alternative is to analyze the identity MMN (iMMN). In this analysis, the ERP to the deviant is compared with the ERP to

[1] https://sccn.ucsd.edu/wiki/Artifact_Subspace_Reconstruction_(ASR)
[2] https://sccn.ucsd.edu/wiki/A08:_DIPFIT

TABLE 1 | Acoustic properties and lexical information of the stimuli.

	Syllable duration (ms)	Consonant duration (ms)	Number of words and senses corresponding to the syllable	Frequency per million (independent appearances)	Frequency per million (all appearances)
Retroflex stimuli					
ʂɤː4	428	200	12, 26	111.4	892.0
tʂɤː4	417	85	7, 9	8554.9	18156.6
tʂ^{h}ɤː4	434	151	5, 8	20.7	144.7
ɻɤː4	439	98	1, 7	90.2	317.1
Average (s)	430 (9)	134 (53)	6.25 (4.57), 12.5 (9.04)	2194.3 (4240.6)	4877.6 (8858.4)
Non-retroflex stimuli					
sɤː4	577	253	7, 15	87.1	840.9
tsɤː4	432	75	3, 5	0.1	0.1
ts^{h}ɤː4	434	176	6, 14	40.1	405.7
lɤː4	393	76	8, 17	72.1	659.3
Average (s)	459 (81)	145 (86)	6 (2.16), 12.75 (5.32)	49.8 (38.5)	476.5 (364.3)

The first two columns show the syllable and consonant durations of the stimuli. The third column is the number of words and senses corresponding to the stimuli according to Xinhua Dictionary (The Commercial Press, 2009). The last two columns show word frequencies obtained from DoWLS (Neergaard et al., 2016). The fourth column is the frequency for the syllable when it occurs as an independent word, and the fifth column is the frequency for all appearances of the syllable, that is, including when the syllable forms a compound.

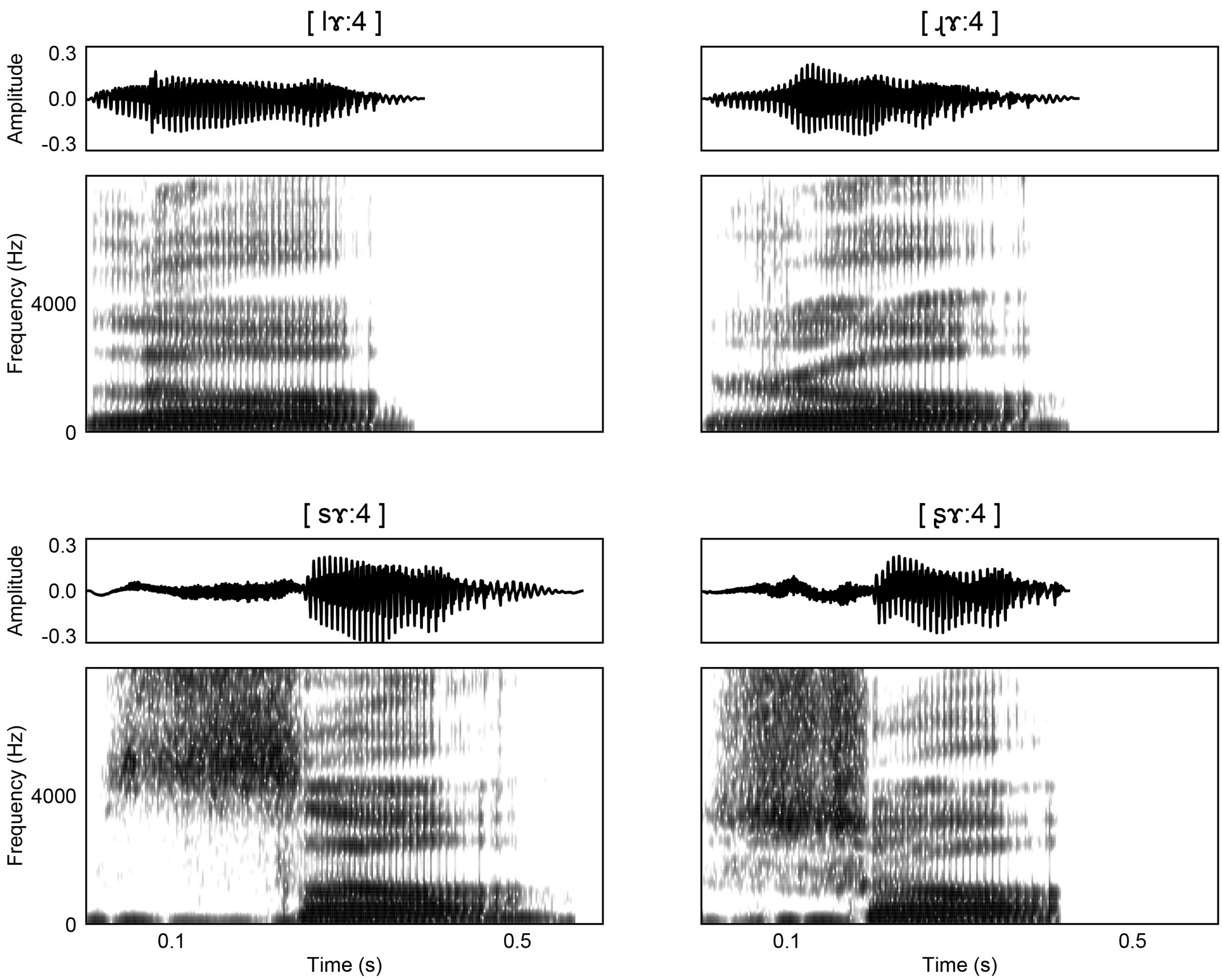


FIGURE 1 | Waveforms and spectrograms for the (top) approximant pair [lɤː4]/[ɻɤː4] and the (bottom) fricative pair [sɤː4]/[ʂɤː4].

the standard version of the same stimulus (Pulvermüller and Shtyrov, 2006; Peter et al., 2010; Hestvik and Durvasula, 2016). In the current study, we include both the within-block (see section "Within-Block MMN") and iMMN (see section "Identity MMN") analyses. Because MMNs are largest over fronto-central scalp areas when referenced to linked mastoids (Näätänen et al., 2007), the average potential of four fronto-central electrode sites (i.e., Cz, Fz, FC1/2) is used to calculate average ERPs. Statistical analyses are conducted in EEGLAB with permutation tests on the t-statistic and an FDR correction for multiple comparisons. Differences with $pFDR < 0.05$ are reported as statistically significant. In the ERP analysis, only significant differences longer than 30 ms in duration are reported and discussed. Because MMNs are sometimes accompanied by a polarity reversal at the mastoid sites (Schröger, 1998), we also computed and visually inspected the averaged difference waves at mastoid electrodes and fronto-central electrodes after average-referencing the data. In the topographic analysis, permutation tests for both the within-block and iMMN comparisons are conducted in each time window on each electrode site (with an FDR correction).

RESULTS

Within-Block MMN

In the retroflex standard block, the permutation test in the −100–600 ms time window shows significant differences at 256–380 ms and 476–540 ms over fronto-central electrode sites. See **Figure 2**. The ERP to the deviant is more negative than to the standard, as is typically observed in MMN paradigms. In the non-retroflex standard block, the permutation test in the −100–600 ms time window shows significant differences between their grand average ERPs at 320–364 ms and 396–416 ms (**Figure 2**); however, the ERPs to retroflex deviants are more positive than the ERPs to non-retroflex standards at 320–364 ms and more negative at 396–416 ms. The positive difference to retroflex deviants at 320–364 ms is not consistent with the characteristics of an MMN.

To examine the nature of this positive difference to the retroflex deviant, we compared the ERPs in the block where the retroflex stimuli are the deviant to the ERPs in the block where the retroflex stimuli are the standard. The ERPs to the retroflex stimuli are similar across the two blocks, as are the ERPs to the non-retroflex stimuli. Thus, both the negative deflection to the deviant in the retroflex standard block and the positive deflection in the non-retroflex standard block may be at least partially caused by intrinsic differences in responses to the retroflex and non-retroflex sounds without the effect of presentation frequency. In other words, regardless of being the standard or the deviant, retroflex sounds elicited more positive ERPs around 320–364 ms, which may potentially confound the within-block MMN analysis. This observation motivated us to examine the iMMNs. Here, the ERP to a deviant is compared with the ERP to the same stimulus in the other block where it serves as the standard. This is done to ensure that potentially different ERPs to the two types of stimuli are controlled. Finally, we observed the typical mastoid reversal in auditory MMN studies; however, this was present in the retroflex standard block only and apparent between 252–388 ms, as well as 440–548 ms (see **Figure 2**).

Identity MMN

Based on visual inspection, the average ERPs over front-central electrode sites to non-retroflex deviants are more negative than that to non-retroflex standards at most time points within the 250–550 ms time window. The ERPs to retroflex standards and deviants have a less defined relative positivity and do not differ throughout the time window; however, the permutation test in the −100–600 ms time window shows no significant differences in either the non-retroflex deviant/standard contrast or the retroflex deviant/standard contrast (**Figure 3**). The absence of significant negative deflections to non-retroflex deviants and positive deflections to retroflex deviants support the analysis that both deflections observed in the within-block MMN comparison are at least partially caused by the different ERPs to the retroflex and non-retroflex stimuli and not the standard-deviant relationship. Here, there is no such well-defined mastoid reversal as seen in the within-block MMN analysis. For non-retroflex stimuli, mastoid potentials trend positive, while frontal cluster potentials trend negative between 250–400 ms and 430–540 ms (see **Figure 3**).

Topographic Comparison

Topographic comparisons were conducted in four time-windows: 150–250 ms, 250–400 ms, 400–450 ms, and 450–550 ms. The selection of these time-windows was largely based on visual inspection of the grand averaged evoked potentials. The 150–250 ms time-window is when MMNs typically occur. The 250–400 ms time-window is when we first observe differences in the within-block MMNs, as well as a negativity in non-retroflex iMMN. Moreover, it also includes the common time window for P300 (Pedroso et al., 2012). The 400–450 ms time-window is when changes in the within-block MMN polarities occur. Finally, the 450–550 ms time-window is the last time window with significant differences in within-block MMN comparisons. ERP topographies in the 250–400 ms time-window are provided in **Figure 4**. ERP topographies for other time windows are provided in the Supplementary Material. At 250–400 ms, permutation tests revealed significant differences elicited by non-retroflex deviants in both the within-block MMN and iMMN comparisons, as well as a significant difference elicited by retroflex deviants in the within-block MMN comparison. Non-retroflex deviants elicit more negative ERPs than both baseline conditions (i.e., retroflex standards and non-retroflex standards). This suggests that the negativities to non-retroflex deviants are not uniquely caused by differences between the ERPs to retroflex and non-retroflex stimuli, as the iMMN results suggest. Regarding the spatial distribution of the negativity, in the within-block MMN comparison, the negativity to non-retroflex deviants is distributed across all electrode sites except for FT9/10 and T8. In the iMMN non-retroflex comparison, the maximum negativity is distributed over left posterior sites and significant at fronto-central to occipital sites (i.e., Fz, FC1/2, Cz, C3/4, CP1/2, CP5, Pz, P3/4, P7, Oz, and O1/2). In Section 3.2, the iMMN comparison only considered fronto-central electrodes; as

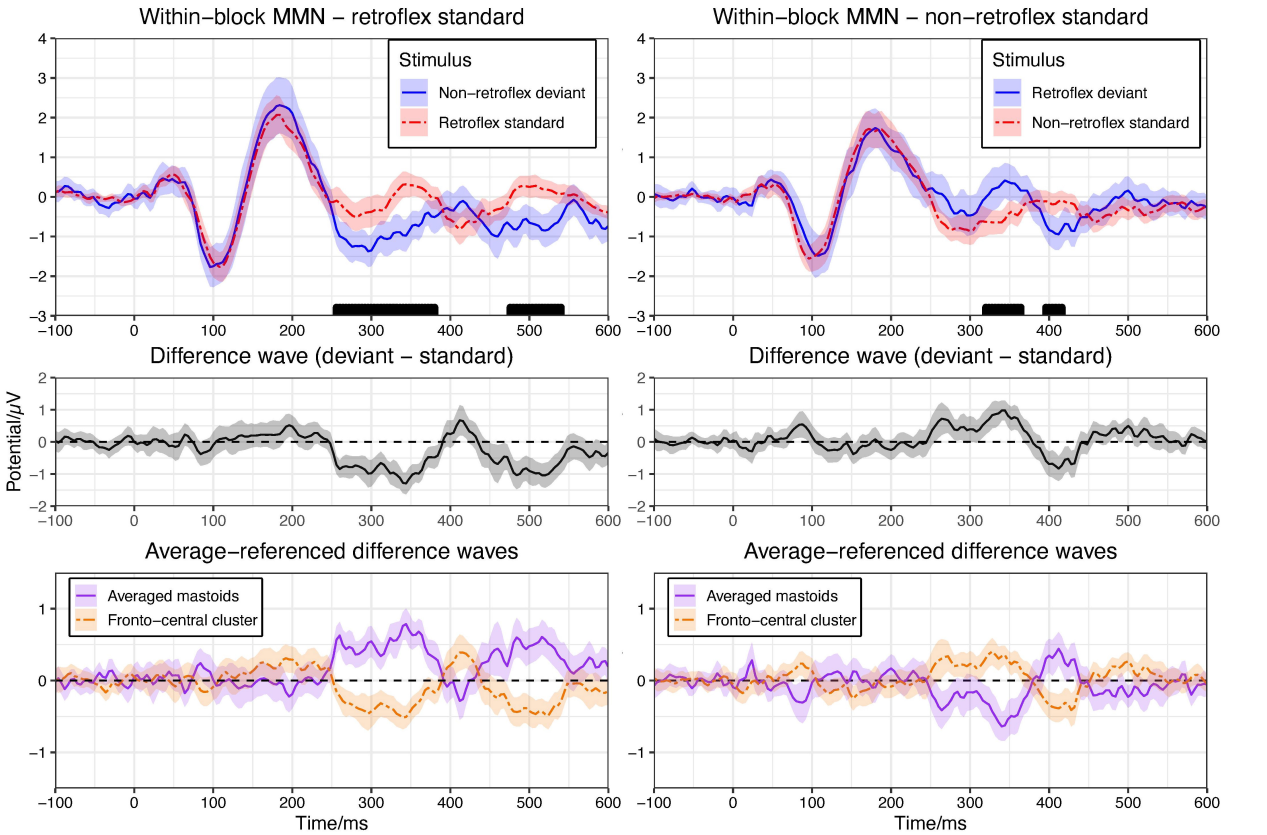


FIGURE 2 | Upper panel: means of grand average ERPs at electrodes Cz, Fz, FC1/2. Ribbons represent 95% confidence intervals. Rug plots along the x-axes mark time points with significant differences for greater than 30 ms in duration (pFDR < 0.05). Plots are arranged to show the within-block MMNs. Left panel compares the ERPs to the retroflex standard and the non-retroflex deviant; right panel compares the ERPs to the non-retroflex standard and the retroflex deviant. Middle panel: difference waves of grand average ERPs at the fronto-central cluster (Cz, Fz, FC1/2). Lower panel: difference waves of grand average ERPs of averaged mastoids and the fronto-central cluster (Cz, Fz, FC1/2) with an average reference.

such, it would not have captured this more posterior distribution and could contribute to why we did not observe a significant difference in the non-retroflex comparison (see **Figure 3**). In other time-windows, the only significant difference in the iMMN comparisons is between 400–450 ms, where the non-retroflex deviant has a larger negative potential at Oz.

DISCUSSION

In the current experiment, Mandarin Chinese participants were presented with a varying standards paradigm that included inter-category variation. There were two experimental blocks. In the retroflex standard block, participants heard varying standards that differed in manner of articulation and shared only the feature [retroflex]. Deviants were the non-retroflex counterparts. In the non-retroflex standard block, the standards were the non-retroflex categories, and the deviants were the retroflex categories. No MMN studies, to our knowledge, have employed a varying standards paradigm wherein multiple phonetic categories are used as the standards. For an MMN to be elicited, the brain must extract the one common feature from the series of standards despite the significant inter-category variation. A binary feature account predicted equal-sized MMNs in both blocks, whereas a privative feature account predicted an asymmetric MMN. Assuming a privative account, we anticipated observing an MMN only in the retroflex standard block, where there is a common feature [retroflex] to be extracted. We also predicted no MMN in the non-retroflex standard block. This design allowed us to test whether listeners can identify and extract the common feature from a natural class set, which is a hallmark of feature behavior in phonetic and phonological systems.

In the within-block MMN analysis, we observed significant negative deflections in the retroflex standard block in the 256–380 ms and 476–540 ms time windows. In the non-retroflex standard block, we observed a more positive deviant response in the 320–364 ms time window. This polarity is opposite to the typical MMN response. In the iMMN analysis, we did not observe differences over fronto-central electrode sites between the responses to standards and deviants. We did, however, observe a difference in the non-retroflex standard-deviant comparison over central-parietal-occipital electrode sites in the 250–400 ms time window. These results are consistent with privative feature accounts and previous MMN studies of features (Eulitz and

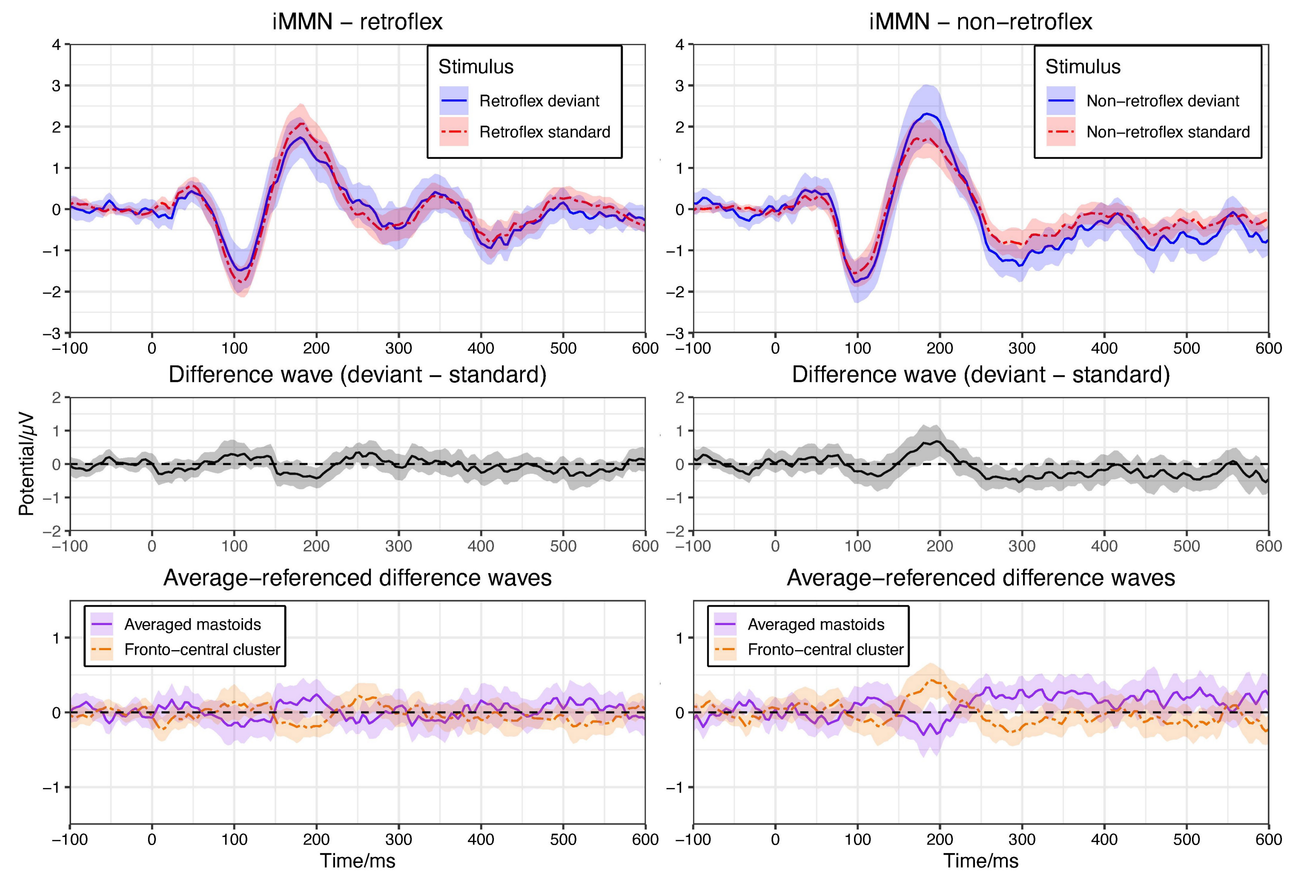


FIGURE 3 | Upper panel: means of grand average ERPs at electrodes Cz, Fz, FC1/2. Ribbons represent 95% confidence intervals. In both panels, there is no significant difference between the ERPs at any time point (pFDR < 0.05). Plots are arranged to show iMMNs. Left panel compares the ERPs to the retroflex standard and deviant; right panel compares the ERPs to the non-retroflex standard and deviant. Middle panel: difference waves of grand average ERPs at the fronto-central cluster (Cz, Fz, FC1/2). Lower panel: difference waves of grand average ERPs of averaged mastoids and the fronto-central cluster (Cz, Fz, FC1/2) with an average reference.

Lahiri, 2004; Scharinger et al., 2012; Cornell et al., 2013; Hestvik and Durvasula, 2016; Scharinger et al., 2016b; Schluter et al., 2016). In the following discussion, we first discuss issues related to stimulus selection in the many-to-many paradigm. Next, we discuss methodological considerations regarding the use of the within-block MMN and iMMN analyses and compare our results from these two analysis methods. Then, we discuss the delayed latency and the parietal scalp distribution of the negativity in the topographic analysis. Finally, we conclude with the broader implications for the feature [retroflex], specifically and phonological representations, more generally.

Stimuli Variation

The current study utilizes a many-to-many oddball paradigm. The goal was to test Mandarin listeners' ability to extract the [retroflex] feature from a series of standards that included inter-category variation. All retroflex consonants in Mandarin and their non-retroflex counterparts ([ʂ]/[s], [tʂ]/[ts], [tʂ^{h}]/[ts^{h}], [ɻ]/[l]) were included. This inter-category variation highlights a few issues worth discussing.

First, the average duration of retroflex stimuli is shorter and less variable than non-retroflex stimuli. In a pairwise comparison, the shorter duration of the retroflex stimuli is largely due to the pairs [ʂɤː4] /[sɤː4] and [tʂɤː4] /[tsɤː4]. This observation might reflect the state of affairs in natural speech or be due to chance in stimulus creation. The larger duration variation in the non-retroflex stimuli is principally due to [sɤː4] and [lɤː4], which have the longest and shortest durations, respectively, of all our items. These length differences could affect ERP latencies to our two stimulus types. In **Figure 2**, the N1 to retroflex stimuli is later than to non-retroflex stimuli in both blocks. Moreover, the larger duration variation in the non-retroflex stimuli could lead to a larger variation in the latency of peaks and troughs in ERPs and produce reduced amplitudes in the grand average ERPs. Both duration and its variation could influence the within-block MMNs. This is discussed in Section 4.2.

Second, our stimuli have different levels of frication. The approximants ([ɻ]/[l]) have a lower level of air turbulence than other stimuli (see **Figure 1**). It should be noted that this contrast in frication level is possible to elicit mismatch responses.

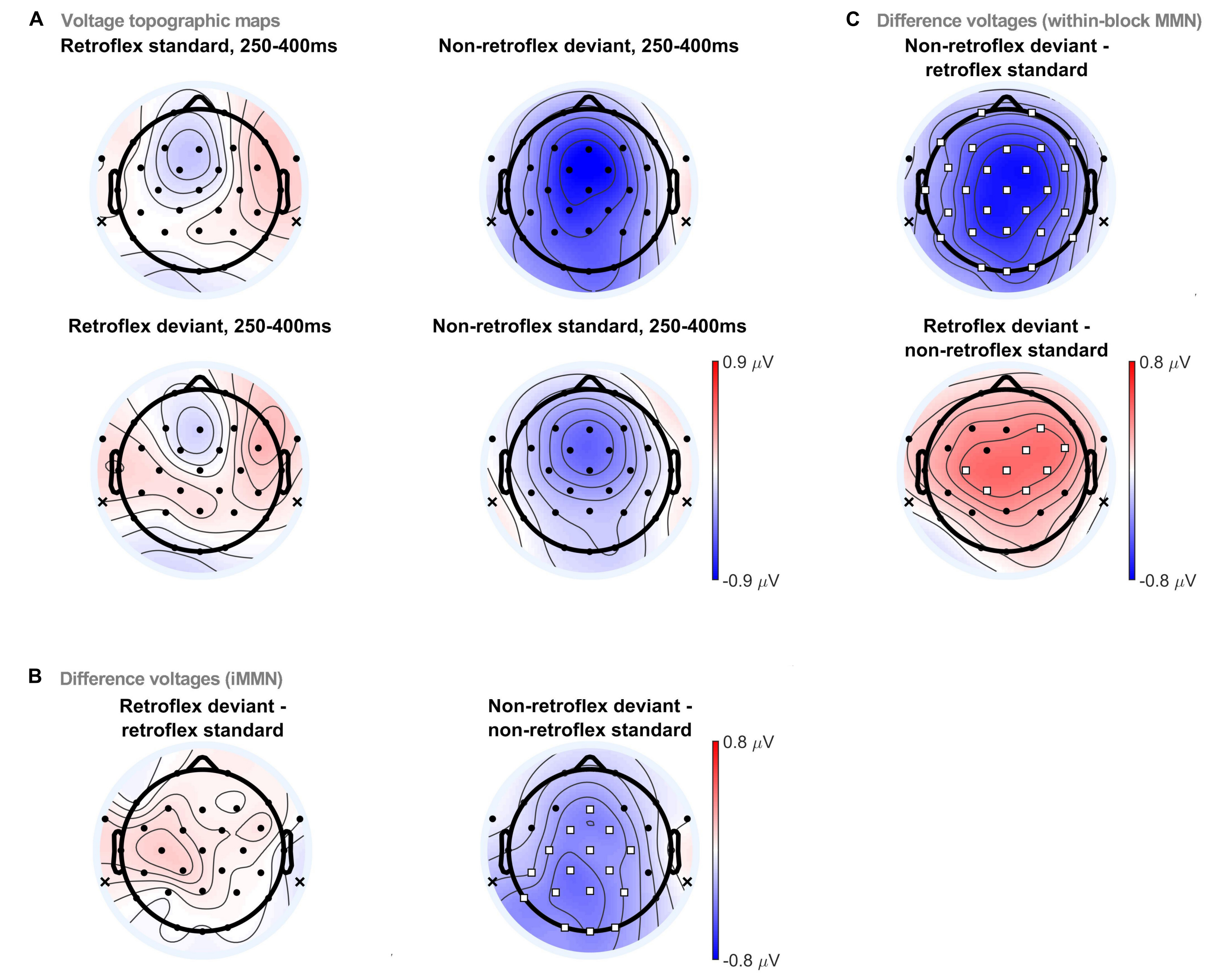


FIGURE 4 | (A) Voltage topographic maps for retroflex/non-retroflex standard/deviant at 250–400 ms. Difference voltage topographic maps and the permutation test results for **(B)** within-block MMNs, **(C)** iMMNs. Electrodes highlighted in white squares denote electrode sites with significant differences using a permutation test (pFDR < 0.05). Topographic analyses were carried out with a linked mastoid reference. The reference locations are marked with an 'x' on the topographic plots.

The current study has a high ratio (1:3) of approximant to non-approximant stimuli and an uncontrolled number of intervening syllables between approximants in the experiment. This configuration renders a weak oddball ratio and reduces the likelihood of observing an MMN. These potential effects are left for future research. Another potential consideration is that [ɻ] is sometimes argued to be a fricative (Duanmu, 2007). Under this categorization, [l] would be the only approximant in our study and the approximant to non-approximant stimuli ratio would be a strong oddball ratio (1:7). There are few points to note, however. First, we do not observe clear frication in the [ɻ] in the current study (see **Figure 1**). This is consistent with previous observations that the Mandarin Chinese [ɻ] has little to no frication (Fu, 1956; Wang, 1979; Duanmu, 2007; Lee-Kim, 2014). Even if we were to assume that [ɻ] contains more frication than [l], the difference in frication between our two approximant stimuli (i.e., [l, ɻ]) would be far less than the difference between [ɻ] and the fricatives/affricates. Thus, it is unlikely that listeners will group [ɻ] with the fricative/affricate stimuli, as opposed to treating it as an approximant.

Finally, each of our syllables corresponds to different numbers of Mandarin words with varying lexical frequency. Overall, the retroflex stimuli correspond to a slightly higher number of words and have a much higher frequency. This is largely due to the unaspirated affricate pair ([tʂ]/[ts]), as [tʂɤː4] is the Mandarin demonstrative 这 'this'. Because this stimulus pair constitute one-fourth of the stimuli in the experiment and are randomly presented with other stimuli, the potential effect of occurrence frequency might be reduced. Lexical frequency has been shown to affect the MMN amplitude: High-frequency deviants elicit larger MMN amplitudes than low-frequency deviants (Alexandrov et al., 2011); however, the opposite pattern is observed in our experiment. The low-frequency (non-retroflex) stimuli elicit larger negativities. This likely suggests that the asymmetry in negativities in the current study is not due to lexical frequency.

Within-Block MMN and iMMN

The within-block MMN design compares the ERPs to deviants and standards in the same block. The standard and deviant are drawn from different categories, for example, retroflex versus non-retroflex. If the ERPs to these different stimuli categories are intrinsically distinct and there is no effect of being a standard or deviant, then this could potentially confound the interpretation of the MMN. To control for this potential confound, some studies calculate an iMMN, which compares the standard and deviant versions of the same stimulus across experimental blocks (e.g., Kraus et al., 1995; Chandrasekaran et al., 2007; Peter et al., 2010; Hestvik and Durvasula, 2016). An alternative method to obtain the iMMN is to present the deviant stimulus alone in a separate control block and subtract the average ERP in this control block from the ERPs to the same deviant stimulus in an experimental block (Kraus et al., 1995; Sharma and Dorman, 2000; Pettigrew et al., 2004).

In our experiment, the feature [retroflex] is physically manifested differently across phonetic categories (e.g., lower distribution of spectral energy in fricatives and affricates, lower F3 and a larger difference between F4 and F5 in approximants, etc.). This results in varying amounts of acoustic differences between the retroflex stimuli and their non-retroflex counterparts. For instance, [ɻ] and [l] differ along more acoustic dimensions than the fricatives and affricates. The acoustic differences between standards and deviants in this experiment are shown to elicit different ERPs and produce unusual patterns (e.g., the unexpected positivity in the MMN time window) in the within-block MMN analysis. Therefore, we also conducted an iMMN analysis. With the assumption that [retroflex] is a privative feature, it is predicted that the retroflex iMMNs are absent or weaker than the non-retroflex iMMNs.

Although the iMMN calculation can eliminate the influence of intrinsic differences between responses to comparing stimuli of different categories, iMMNs are susceptible to the repetition effect of the same or similar stimuli. In the *adaption to standards* and *predictive coding* accounts of the MMN (see Fitzgerald and Todd, 2020 for a review), ERPs to standards might be modulated as a result of repetition. For example, the number of repetitions could influence ERPs to the standard in such a way that the "repetition positivity" (RP, a positive deflection in the ERP around 50 to 250 ms to the standard stimuli) increases with more repetitions (Haenschel, 2005). This increased RP leads to a larger MMN because the RP occurs at a similar window as the MMN, and the relatively more positive standard response results in a more negative difference wave. The RP, however, has only been consistently reported in the roving oddball paradigm. In a roving oddball paradigm, each block contains different trains of stimuli, and each train has a different standard and deviant. The deviant in the preceding train is the standard of the following train (Haenschel, 2005). In our experiment, because the roving oddball paradigm was not used, we do not *a priori* expect to observe the RP.

Besides the RP, another potential repetition effect is the refractory state of frequency-specific neurons (Jacobsen et al., 2003; Näätänen et al., 2005), which could reduce N1 amplitudes and contribute to the calculated MMNs. In the current study, retroflex stimuli tend to have higher energy in the 2000–4000 Hz frequency range, thus the neurons sensitive to this frequency range might generate a smaller N1 response when the retroflex stimuli are repeated as the standard. In turn, this could increase the amplitude of the calculated iMMN in the N1 time window. Assuming the refractory effect is present in our data, we can predict that retroflex standards elicit a smaller N1 relative to retroflex deviants. The result shows no significant positive deflections or any other significantly different responses to the retroflex standard relative to the retroflex deviant. Thus, the refractory effect—or any other repetition effect to the retroflex standard—is not observed in the ERPs to the retroflex standard. Note that although we focused on the RP wave and the refractory effect, the repetition of standards could also influence brain responses in other ways that should be considered when interpreting the iMMN, such as enhancing β-band oscillatory power (Scharinger et al., 2016b).

In summary, the positive difference to the retroflex deviant in the within-block MMN comparison and its reduction in amplitude/disappearance in the iMMN comparison confirmed that the brain responds to retroflex/non-retroflex stimuli differently and that it is insufficient to use the within-block MMN as the sole evidence for the negativity to the non-retroflex deviants; however, the survival of significant differences in the iMMN analysis for non-retroflex stimuli from multiple electrode sites in the topographic analysis suggests that non-retroflex deviants indeed elicit a negative response. The existence of negative deflections that only associate with the non-retroflex stimuli is also consistent with the observation that in a within-block MMN analysis (see **Figure 2**), the difference between the non-retroflex deviant and retroflex standard is larger than the difference between the retroflex deviant and the non-retroflex standard.

Latency and Distribution of the Negativities

Negative deflections in this study occurred around 250–450 ms post-stimulus onset. The time window is later than the normal MMN window. Negativities in oddball paradigms distributed at a late time window have been reported as late MMNs (e.g., Korpilahti et al., 1995; Zachau et al., 2005) or late discriminative negativities (LDNs; Cheour et al., 2001; Čeponienė et al., 2002; Martynova et al., 2003; Strotseva-Feinschmidt et al., 2015; Hestvik and Durvasula, 2016). The LDN normally appears with the traditional MMN, but it can also appear independently (Strotseva-Feinschmidt et al., 2015). The LDN is more frequently found in children than in adults and is shown to decrease (Bishop et al., 2011) or disappear (Strotseva-Feinschmidt et al., 2015) with development into adulthood. In adults, the LDN spatial distribution is difficult to characterize. In the limited number of adult LDN reports, it has been observed in various locations, including anterior sites (Zachau et al., 2005; Hestvik and Durvasula, 2016), fronto-central sites (Korpilahti et al., 1995; Shafer et al., 2004), right and central sites (Peter et al., 2012), and parieto-central sites (Hestvik and Durvasula, 2016). Thus, discerning the presence of an LDN in adults based on topographic patterns is not straightforward.

The rarity of LDNs in adults leads to the question of what property of the stimuli elicits them. The cause of the LDN in adults is insufficiently studied and yet unknown. Bishop et al. (2011) suggest that LDNs might appear as a result of additional processing required by certain features of stimuli that are difficult to detect. Their hypothesis is based on the finding that in contrast to MMNs, LDNs are larger for smaller differences between standard and deviant stimuli (see also Čeponienė et al., 2004). This hypothesis can also account for the decrease of LDNs with age considering the maturation of brain and the exposure to language. For adults, LDNs can be elicited by both speech sounds (Hestvik and Durvasula, 2016; Monahan et al., in prep.) and non-speech sounds (Zachau et al., 2005; Peter et al., 2012). The experiments that elicited adult LDNs along with our experiment all use the many-to-many oddball paradigm (i.e., multiple unique stimuli for both standards and deviants). In our current paradigm, the variation in both standards and deviants demands a more abstract grouping for deviant detection. This would agree with the potential relationship between the presence of LDNs and processing difficulty. Thus, it is possible that a certain level of complexity in the stimuli is a necessary condition to elicit adult LDNs. Additionally, the [retroflex] feature might be more difficult to process due to its low frequency in languages. Shafer et al. (2004) conducted an MMN study with retroflex (i.e., /ɖa/) and bilabial stimuli (i.e., /ba/). Hindi speakers showed an MMN with a later peak latency when /ɖa/ was the standard (∼200–300 ms), compared to when /ba/ was the standard (∼100–250 ms). The later MMN latency when retroflex stimuli are the standard might be comparable to the long-latency negativities in the retroflex-standard block in our experiment. Moreover, their observed asymmetric MMN is also consistent with a monovalent [retroflex] feature. That being said, in Shafer et al. (2004), there exist two changes between the standard and deviant: a place change and manner of articulation change. This makes exclusively interpreting the role of [retroflex] difficult.

Besides the late timing, the negativity in the iMMN analysis also occurs at an unexpected distribution that spreads from central to occipital electrode sites with a maximum negativity at parietal sites, instead of the fronto-central region where the MMN is usually observed to have the largest amplitude. One possible explanation for the more parietal distribution is that the negativity to the deviants in this experiment is generated at different neural sources than conventional MMNs. According to the dual-generator model (Näätänen et al., 1978), MMNs originate from a principal source in primary auditory cortex that is responsible for the memory component in the deviant detection, and a secondary prefrontal source responsible for the additional attention directed to the deviation; however, because the standard stimuli vary across phonetic categories in our design, primary areas of auditory cortex alone might not be sufficient for the identification of the retroflex feature. Thus, the formation and violation of the memory trace for the retroflex feature might need to be completed at a later stage in speech processing, for example, at a location closer to the superior temporal sulci where phonological information is processed (Okada and Hickok, 2006; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Vaden et al., 2010). Using fMRI, Scharinger et al. (2016a) observed that underspecified vowels (i.e., [e]) following specified vowels (i.e., [o]) in same-different word pairs resulted in stronger blood oxygenation level dependent (BOLD) responses in bilateral superior-temporal sulcus (STS). This is in comparison to when the first member of the same-different word pair included an underspecified vowel, and the second member included a specified vowel. These results place a locus of feature processing in STS. Moreover, as they note, these findings mimic the typical pattern observed in asymmetric MMN responses to specified and underspecified speech sounds: A larger MMN is observed to underspecified deviants following specified standards compared to the opposite orientation. In the current experiment, a larger MMN was observed when the standard was specified for [retroflex], as compared to when the standard was underspecified. Given the relatively sparse electrode array (32 channels) used in the current experiment, source analyses are not possible; however, previous combined EEG and hemodynamic experiments are potentially useful in linking the current results with STS activity.

Both the latency of the negativities and their more posterior distribution in the iMMN subtraction resemble those of the N400 response. The N400 response is typically described as a negative deflection to semantically incongruent stimuli at posterior electrode sites around 200–600 ms post-stimulus onset (Kutas and Federmeier, 2011). Unlike the low-level processing of auditory information in the MMN elicitation, the N400 elicitation requires access to higher-level semantic information, which is reflected in the N400's different neural generators. Studies have shown that brain regions supposed to be related to semantic processing, such as the middle and superior temporal areas, the medial temporal lobe, and the prefrontal areas are involved in the N400 responses (Kutas and Federmeier, 2011). The similarities in the latency and distribution of the negativities in our experiment and the N400 raise the question of whether the many-to-many oddball paradigm elicits negativities in a similar manner as the elicitation of the N400. Recently, the N400 has been accounted for within predictive coding frameworks for language processing, akin to extant models of the MMN (Bornkessel-Schlesewsky and Schlesewsky, 2019). Moreover, while the choice of the longer ISI in our design was intended to reinforce phonological processing (Werker and Logan, 1985; Yu Y. H. et al., 2017), it is also possible that it permitted greater influence of lexical factors than initially intended (see Čeponienė et al., 1999, who argued that shorter ISIs in children lead to stronger auditory memory traces and consequently larger MMNs for putatively phonological contrasts). That being said, each of our items corresponds to approximately six different words and more than 12 senses, on average. As such, it is difficult to know which particular lexical item is activated by the participant on a given trial, making the observation of consistent N400 effects at the lexical level unlikely in our design. We leave this possibility for future research.

Asymmetric MMNs and Underspecification

As with previous MMN studies that have identified a role for features in speech processing (Eulitz and Lahiri, 2004; Scharinger et al., 2012; Cornell et al., 2013; Hestvik and Durvasula, 2016; Scharinger et al., 2016b; Schluter et al., 2016; Hestvik et al., 2020), we observed an asymmetric MMN. In each of these previous studies, an MMN was observed when the category with a specified feature was the standard and either no MMN or a reduced MMN when the category underspecified for that feature was used as the standard. In the current paper, we did not test an underspecified relationship *per se*, but the presence or absence of a common feature, i.e., [retroflex], in a series of standards that varied in their phonetic category. In this sense, the current experiment is similar to these previous findings. For example, Cornell et al. (2013) tested [ɡ] versus [d]. As [coronal] is thought to be underspecified in the lexicon (Archangeli, 1988; Avery and Rice, 1989; Paradis and Prunet, 1991; Lahiri and Reetz, 2010; Cummings et al., 2017), [d] is underspecified for its place feature. The place feature [dorsal], however, is specified, and so [ɡ] has a specified place feature in the lexicon. Cornell et al. (2013) compared experimental blocks when [d] was the standard and [ɡ] was the deviant versus when [ɡ] was the standard and [d] was the deviant. A larger MMN was observed when [ɡ] is the standard, as [ɡ] is specified for its place feature. As such, when a privative feature is specified in the standards, and the deviant does not contain that feature, a mismatch occurs, resulting in an MMN. When the feature is underspecified, there is nothing upon which an auditory memory representation can be constructed that would contrast with the deviant. In these cases, there is no mismatch, and either no MMN or a smaller MMN is observed.

The current experiment is similar in that there is a single feature that can be extracted from the varying standards in the retroflex standard block, assuming that [retroflex] is a privative feature. And if the deviant mismatches with that feature, then an MMN and an iMMN is predicted, as we observed. When there is no shared feature, the deviant will not be able to mismatch with the auditory memory representation of the standard and again, either no MMN or a reduced MMN is predicted. We observed no MMN in both the within-block MMN and iMMN analyses in the non-retroflex standard block. It might be argued that predicting an asymmetric MMN is problematic because our non-retroflex standards were all [coronal] in their place of articulation. And as such, they shared a common feature, which should elicit an MMN to the retroflex deviants in the context of our coronal non-retroflex standards. Thus, the MMNs to the retroflex and non-retroflex stimuli should be symmetrical. There are two potential responses. First, each of our retroflex categories was also [coronal] in their place of articulation. Consequently, this is not a distinguishing feature between the retroflex and non-retroflex categories. That is, the feature [coronal] will not create a mismatch, which is necessary for an MMN (Lahiri and Reetz, 2010). Second, it has been demonstrated that when the standards are [coronal] and that is the distinguishing feature with the deviants, either no or reduced MMNs are observed, as [coronal] is underspecified in the mental lexicon (Eulitz and Lahiri, 2004; Cornell et al., 2013; Cummings et al., 2017). In light of this, symmetric MMNs based on [coronal] to retroflex and non-retroflex stimuli are unlikely. In short, we conclude that the feature [retroflex] is privative in Mandarin Chinese, and this feature can be extracted from a series of standards with inter-category variation that all share this feature.

CONCLUSION

The goal of the current paper was to determine whether listeners can extract a common phonetic feature from a series of standards with inter-category variation in an MMN paradigm. Observing an MMN would suggest that listeners extracted the relevant phonetic feature, and more broadly, that listeners access and represent speech sounds in terms of features. No previous work, to the best of our knowledge, has employed inter-category variation in the standards. In particular, we presented Mandarin-speaking participants with two blocks in an auditory oddball paradigm. In one block, the standards all shared the retroflex feature and the deviants were non-retroflex consonants. In the other block, the standards were all non-retroflex. A binary model for distinctive features would predict symmetric MMNs across the two blocks. A privative model for distinctive features would predict an MMN only in the retroflex standard block and no MMN in the non-retroflex standard block, as there would be no feature to bind the standard stimuli together. We found late MMNs/LDNs in the retroflex standard block only. This result suggests that first, Mandarin speakers extract the privative feature [retroflex] from varying stimuli, and an asymmetric MMN was observed. This supports a privative model of distinctive features. Second, the later differences in the ERP responses in a many-to-many paradigm might suggest the extraction of a feature from inter-category stimuli requires additional effort. In summary, listeners can extract phonetic features and construct auditory memory traces based on these features—despite significant acoustic and inter-category phonological variation—so long as a shared feature binds the standards together.

DATA AVAILABILITY STATEMENT

The stimuli and raw EEG datasets can be found in the Open Science Framework https://osf.io/uyg9e/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Toronto Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZF and PM designed the experiment, created the stimuli, analyzed the data and co-wrote the manuscript. ZF programmed the experiment and ran participants, supported by research assistants. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada (RGPIN-2017-06053) to PM.

ACKNOWLEDGMENTS

We thank Alejandro Pérez for assistance with various aspects of the data analysis and Mathias Scharinger and Rachel Soo for insightful comments on a previous version of this manuscript. All errors are our own.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.609898/full#supplementary-material

REFERENCES

Alexandrov, A. A., Boricheva, D. O., Pulvermüller, F., and Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS ONE* 6:e22999. doi: 10.1371/journal.pone.0022999

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear.* 16, 38–51. doi: 10.1097/00003446-199502000-00004

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

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R., and Sams, M. (1993). Phonetic invariance in the human auditory cortex. *Neuroreport* 4, 1356–1358. doi: 10.1097/00001756-199309150-00018

Avery, P., and Rice, K. (1989). Segment structure and coronal underspecification. *Phonology* 6, 179–200. doi: 10.1017/s0952675700001007

Baković, E. (2014). Phonemes, segments and features. *Lang. Cogn. Neurosci.* 29, 21–23. doi: 10.1080/01690965.2013.848992

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* 9:16. doi: 10.3389/fninf.2015.00016

Bishop, D. V. M., Hardiman, M. J., and Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? a study using time-frequency analysis of mismatch responses from 7 years to adulthood: is auditory discrimination mature? *Dev. Sci.* 14, 402–416. doi: 10.1111/j.1467-7687.2010.00990.x

Blevins, J., Egurtzegi, A., and Ullrich, J. (2020). Final obstruent voicing in lakota: phonetic evidence and phonological implications. *Language* 96, 294–337. doi: 10.1353/lan.2020.0022

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298

Cai, Q., and Brysbaert, M. (2010). SUBTLEX-CH: chinese word and character frequencies based on film subtitles. *PLoS One* 5:e10729. doi: 10.1371/journal.pone.0010729

Čeponienė, R., Lepistö, T., Soininen, M., Aronen, E., Alku, P., and Näätänen, R. (2004). Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology* 41, 130–141. doi: 10.1111/j.1469-8986.2003.00138.x

Čeponienė, R., Service, E., Kurjenluoma, S., Cheour, M., and Näätänen, R. (1999). Children's performance on pseudoword repetition depends on auditory trace quality: evidence from event-related potentials. *Dev. Psychol.* 35, 709–720. doi: 10.1037/0012-1649.35.3.709

Čeponienė, R., Yaguchi, K., Shestakova, A., Alku, P., Suominen, K., and Näätänen, R. (2002). Sound complexity and 'speechness' effects on pre-attentive auditory discrimination in children. *Int. J. Psychophysiol.* 43, 199–211. doi: 10.1016/S0167-8760(01)00172-6

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res.* 1128, 148–156. doi: 10.1016/j.brainres.2006.10.064

Cheour, M., Korpilahti, P., Martynova, O., and Lang, A.-H. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. *Audiol. Neuro Otol.* 6, 2–11. doi: 10.1159/000046804

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Clements, G. N. (1985). The geometry of phonological features. *Phonology* 2, 225–252. doi: 10.1017/s0952675700000440

Clements, G. N., and Hume, E. V. (1995). "The internal organization of speech sounds," in *The Handbook of Phonological Theory*, ed. J. A. Goldsmith, (Hoboken, NJ: Blackwell), 245–306.

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

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862

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

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Duanmu, S. (2007). *The Phonology of Standard Chinese*, 2nd Edn. Oxford: Oxford University Press.

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Fitzgerald, K., and Todd, J. (2020). Making sense of mismatch negativity. *Front. Psychiatry* 11:468. doi: 10.3389/fpsyt.2020.00468

Fu, M. (1956). Beijing hua de yinwei he pinyin zimu [phonemes and pinyin symbols in the beijing speech]. *Zhongguo Yuwen* 5, 3–12.

Gomes, H., Ritter, W., and Vaughan, H. G. (1995). The nature of preattentive storage in the auditory system. *J. Cogn. Neurosci.* 7, 81–94. doi: 10.1162/jocn.1995.7.1.81

Gow, D. W. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Percep. Psychophys.* 65, 575–590. doi: 10.3758/BF03194584

Greenberg, S. (1999). Speaking in shorthand – a syllable-centric perspective for understanding pronunciation variation. *Speech Commun.* 29, 159–176. doi: 10.1016/S0167-6393(99)00050-3

Haenschel, C. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *J. Neurosci.* 25, 10494–10501. doi: 10.1523/JNEUROSCI.1227-05.2005

Halle, M. (1983). On distinctive features and their articulatory implementation. *Nat. Lang. Linguistic Theory* 1, 91–105.

Halle, M. (2002). *From memory to speech and back: Papers on phonetics and phonology 1954-2002*. Berlin: Walter de Gruyter.
Halle, M., and Stevens, K. N. (1962). Speech recognition: a model and a program for research. *IRE Trans.n Inf. Theory* 8, 155–159. doi: 10.1109/tit.1962.1057686
Harris, J., and Lindsey, G. (1995). "The elements of phonological representation," in *Frontiers in Phonology*, eds J. Durand, and F. Katamba, (Harlow, Essex: Longman), 37–49.
Hestvik, A., and Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain Lang.* 152, 28–43. doi: 10.1016/j.bandl.2015.10.007
Hestvik, A., Shinohara, Y., Durvasula, K., Verdonschot, R. G., and Sakai, H. (2020). Abstractness of human speech sound representations. *Brain Res.* 2020:146664. doi: 10.1016/j.brainres.2020.146664
Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Lang. Cogn. Neurosci.* 29, 2–20. doi: 10.1080/01690965.2013.834370
Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
Hussain, Q., Proctor, M., Harvey, M., and Demuth, K. (2017). Acoustic characteristics of punjabi retroflex and dental stops. *J. Acoustical Soc. Am.* 141, 4522–4542. doi: 10.1121/1.4984595
Jacobsen, T., Schröger, E., Horenkamp, T., and Winkler, I. (2003). Mismatch negativity to pitch change: Varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials. *Neurosci. Lett.* 344, 79–82. doi: 10.1016/S0304-3940(03)00408-7
Jakobson, R., Fant, G., and Halle, M. (1961). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
Kazanina, N., Phillips, C., and Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proc. Natl. Acad. Sci.* 103, 11381–11386. doi: 10.1073/pnas.0604821103
Korpilahti, P., Lang, H., and Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography Clin. Neurophysiol.* 4:96.
Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., and Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *J. Cogn. Neurosci.* 7, 25–32. doi: 10.1162/jocn.1995.7.1.25
Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
Lahiri, A., and Reetz, H. (2002). Underspecified recognition. *Lab. Phonol.* 7, 637–675. doi: 10.1515/9783110197105.2.637
Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonetics* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002
Lee, W.-S. (1999). *An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin. Proceedings of the 14th International Congress of Phonetic Sciences*. Hong Kong: University of Hong Kong, 413–416.
Lee, W.-S., and Zee, E. (2003). Standard chinese (Beijing). *J. Int. Phonetic Assoc.* 33, 109–112. doi: 10.1017/S0025100303001208
Lee-Kim, S.-I. (2014). Revisiting mandarin 'apical vowels': an articulatory and acoustic study. *J. Int. Phonetic Assoc.* 44, 261–282. doi: 10.1017/S0025100314000267
Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.
Lombardi, L. (1995). Laryngeal features and privativity. *Linguistic Rev.* 12, 35–60. doi: 10.1515/tlir.1995.12.1.35
Mahajan, Y., Peter, V., and Sharma, M. (2017). Effect of EEG referencing methods on auditory mismatch negativity. *Front. Neurosci.* 11:560. doi: 10.3389/fnins.2017.00560
Martynova, O., Kirjavainen, J., and Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neurosci. Lett.* 340, 75–78. doi: 10.1016/S0304-3940(02)01401-5
McCarthy, J. J. (1988). Feature geometry and dependency: a review. *Phonetica* 45, 84–108. doi: 10.1159/000261820
McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
Mester, R. A., and Itô, J. (1989). Feature predictability and underspecification: palatal prosody in japanese mimetics. *Language* 65:258. doi: 10.2307/415333
Monahan, P. J. (2018). Phonological knowledge and speech comprehension. *Ann. Rev. Linguistics* 4, 21–47. doi: 10.1146/annurev-linguistics-011817-045537
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, 1–21. doi: 10.1111/1469-8986.3810001
Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9
Näätänen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42, 25–32. doi: 10.1111/j.1469-8986.2005.00256.x
Näätänen, R., and Kreegipuu, K. (2012). *The mismatch negativity (MMN). In The Oxford handbook of event-related potential components*. Oxford: Oxford University Press, 143–157.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0
Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026
Neergaard, K., Xu, H., and Huang, C.-R. (2016). *Database of Mandarin neighborhood statistics. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Hong Kong: Hong Kong Polytechnic University, 4032–4036.
Nenonen, S., Shestakova, A., Huotilainen, M., and Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cogn. Brain Res.* 16, 492–495. doi: 10.1016/S0926-6410(03)00055-7
Okada, K., and Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *NeuroReport* 17, 1293–1296. doi: 10.1097/01.wnr.0000233091.82536.b2
Oostenveld, R., and Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Hum. Brain Mapping* 17, 179–192. doi: 10.1002/hbm.10061
Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2012). *AMICA: An adaptive mixture of independent component analyzers with shared components. Swartz Center for Computational Neuroscience*. California: : University of California San Diego Technical Report, 15.
Paradis, C., and Prunet, J. (1991). *The special status of coronals: Internal and external evidence*. Cambridge, MA: Academic Press.
Pedroso, R. V., Fraga, F. J., Corazza, D. I., Andreatto, C. A. A., de Melo Coelho, F. G., Costa, J. L. R., et al. (2012). P300 latency and amplitude in Alzheimer's disease: a systematic review. *Brazilian J. Otorhinolaryngol.* 78, 126–132.
Peter, V., McArthur, G., and Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neurosci. Lett.* 482, 71–75. doi: 10.1016/j.neulet.2010.07.010
Peter, V., Mcarthur, G., and Thompson, W. F. (2012). Discrimination of stress in speech and music: a mismatch negativity (MMN) study. *Psychophysiology* 49, 1590–1600. doi: 10.1111/j.1469-8986.2012.01472.x
Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., et al. (2004). Automatic auditory processing of english words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear Hear.* 25, 284–301. doi: 10.1097/01.AUD.0000130800.88987.03
Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567
Poeppel, D., and Monahan, P. J. (2011). Feedforward and feedback in speech perception: revisiting analysis by synthesis. *Lang. Cogn. Proc.* 26, 935–951. doi: 10.1080/01690965.2010.493301
Politzer-Ahles, S., Schluter, K., Wu, K., and Almeida, D. (2016). Asymmetries in the perception of mandarin tones: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percep. Perform.* 42, 1547–1570. doi: 10.1037/xhp0000242

Polka, L., and Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *J. Acoustical Soc. Am.* 78, 1187–1197. doi: 10.1121/1.392887

Pulvermüller, F., and Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71. doi: 10.1016/j.pneurobio.2006.04.004

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Reiss, C. (2017). "Substance Free phonology," in *The Routledge Handbook of Phonological Theory*, 1st Edn, eds S. J. Hannahs, and A. R. K. Bosch, (Milton Park, UK: Routledge), 425–452. doi: 10.4324/9781315675428-15

Sagey, E. (1986). *The representation of features and relations in autosegmental phonology* [PhD Thesis]. Cambridge: Massachusetts Institute Of Technology.

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

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *J. Speech Lang. Hearing Res.* 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)

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

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: An ERP investigation of english fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behav. Res. Methods Instruments Comp.* 30, 131–145. doi: 10.3758/BF03209423

Shafer, V. L., Schwartz, R. G., and Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cogn. Brain Res.* 18, 242–254. doi: 10.1016/j.cogbrainres.2003.10.007

Sharma, A., and Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *J. Acoustical Soc. Am.* 107, 2697–2703. doi: 10.1121/1.428655

Steriade, D. (1995). "Underspecification and markedness," in *Handbook of phonological theory*, ed. J. A. Goldsmith, (Oxford, UK: Blackwell), 114–174.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoustical Soc. Am.* 111, 1872–1891. doi: 10.1121/1.1458026

Strotseva-Feinschmidt, A., Cunitz, K., Friederici, A. D., and Gunter, T. C. (2015). Auditory discrimination between function words in children and adults: a mismatch negativity study. *Front. Psychol.* 6:1930. doi: 10.3389/fpsyg.2015.01930

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., and Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magnetoencephalographic (MEG) study. *NeuroImage* 9, 330–336. doi: 10.1006/nimg.1999.0405

The Commercial Press. (2009). *Xinhua Zidian (Xinhua Dictionary)*, 10th Edn. Beijing, China: Commercial Press.

Trubetskoy, N. S. (1969). *Principles of Phonology*. California: University of California Press.

Vaden, K. I., Muftuler, L. T., and Hickok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *NeuroImage* 49, 1018–1023. doi: 10.1016/j.neuroimage.2009.07.063

van der Hulst, H. (2016). Monovalent 'Features' in phonology: monovalent 'features' in phonology. *Lang. Linguistics Compass* 10, 83–102. doi: 10.1111/lnc3.12158

Wang, L. (1979). Xiandai Hanyu yuyin fnxi zhong de jige wenti [some problems in modern chinese phonetic analysis]. *Zhongguo Yuwen* 4, 281–286.

Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percep. Psychophys.* 37, 35–44. doi: 10.3758/BF03207136

Winkler, I., Kujala, T., Alku, P., and Näätänen, R. (2003). Language context and phonetic change detection. *Cogn. Brain Res.* 17, 833–844. doi: 10.1016/S0926-6410(03)00205-2

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn. Brain Res.* 7, 357–369. doi: 10.1016/S0926-6410(98)00039-1

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

Yu, K., Li, L., Chen, Y., Zhou, Y., Wang, R., Zhang, Y., et al. (2019). Effects of native language experience on mandarin lexical tone processing in proficient second language learners. *Psychophysiology* 56:e13448. doi: 10.1111/psyp.13448

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Front. Neurosci.* 11:95. doi: 10.3389/fnins.2017.00095

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., et al. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *NeuroReport* 16, 2015–2019. doi: 10.1097/00001756-200512190-00009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fnhum.2021.629517

Neural Representation of the English Vowel Feature [High]: Evidence From /ɛ/ vs. /ɪ/

Yan H. Yu[1]* *and Valerie L. Shafer*[2]

[1] *Department of Communication Sciences and Disorders, St. John's University, Queens, NY, United States,* [2] *The Graduate Center, City University of New York, New York, NY, United States*

Many studies have observed modulation of the amplitude of the neural index mismatch negativity (MMN) related to which member of a phoneme contrast [phoneme A, phoneme B] serves as the frequent (standard) and which serves as the infrequent (deviant) stimulus (i.e., AAAB vs. BBBA) in an oddball paradigm. Explanations for this amplitude modulation range from acoustic to linguistic factors. We tested whether exchanging the role of the mid vowel /ɛ/ vs. high vowel /ɪ/ of English modulated MMN amplitude and whether the pattern of modulation was compatible with an underspecification account, in which the underspecified height values are [−high] and [−low]. MMN was larger for /ɛ/ as the deviant, but only when compared across conditions to itself as the standard. For the within-condition comparison, MMN was larger to /ɪ/ deviant minus /ɛ/ standard than to the reverse. A condition order effect was also observed. MMN amplitude was smaller to the deviant stimulus if it had previously served as the standard. In addition, the amplitudes of late discriminative negativity (LDN) showed similar asymmetry. LDN was larger for deviant /ɛ/ than deviant /ɪ/ when compared to themselves as the standard. These findings were compatible with an underspecification account, but also with other accounts, such as the Natural Referent Vowel model and a prototype model; we also suggest that non-linguistic factors need to be carefully considered as additional sources of speech processing asymmetries.

Keywords: mismatch negativity, late discriminative negativity, brain asymmetry, underspecification, vowel, event-related potential, prior, predictive modeling

OPEN ACCESS

Edited by:
István Czigler, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences (MTA), Hungary

Reviewed by:
Andreas Højlund, Aarhus University, Denmark
Ryan Rhodes, Rutgers University, Newark, United States

Correspondence:
Yan H. Yu
yanhyu@gmail.com;
yuy1@stjohns.edu

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *15 November 2020*
Accepted: *08 March 2021*
Published: *09 April 2021*

Citation:
Yu YH and Shafer VL (2021) Neural Representation of the English Vowel Feature [High]: Evidence From /ɛ/ vs. /ɪ/. Front. Hum. Neurosci. 15:629517. doi: 10.3389/fnhum.2021.629517

INTRODUCTION

The ability to discriminate and categorize speech sound contrasts is crucial for fast and efficient lexical access, but our understanding of how this process unfolds is still incomplete. An enduring question has been how speech sound information is represented in the human mind/brain to support the process of lexical access (Strange, 2011). Two general proposals have been offered. One proposal is that phonological representations precisely encode the physical world (in terms of sensory, motor, and statistical properties, such as lexical type and token frequency) (Bybee, 2002). The alternative is that phonological representations are abstract in nature and that the information that is part of the representation is determined by constraints that cannot be fully explained using sensory and motor factors or other information external to the phonological system. Models arguing for abstract representations have favored reducing speech sounds to a small set of features (often binary) that are assumed/hypothesized to reflect what is stored as part of the representation (Chomsky and Halle, 1968). These "abstract" models also favor representations that have the minimum necessary information, with predictable or redundant information being filled in by some process during production and perception. No consensus has yet been reached regarding which model is superior, perhaps because both types of models have found some support. Much of the evidence bearing on these questions has come from cross-linguistic studies that examined the patterning of phonological systems, or from behavioral psycholinguistic studies (Kazanina et al., 2018; Samuel, 2020).

Until fairly recently, few studies have used neurophysiological methods to directly address the nature of phonological representations. Initially, these methods were designed to test questions of phonetic representations, such as categorical perception or questions asking whether particular brain measures, such as mismatch negativity (MMN) or P3b indexed acoustic (auditory general), phonetic (language general), or phonemic (language specific) processes (Buchwald et al., 1994; Näätänen et al., 1997; for review, Näätänen et al., 2007).

Mismatch negativity has emerged as the primary method for testing questions of neural representation of speech. MMN is a neural discriminative response generated in the auditory cortex in response to a discriminable change in a repetitive auditory stimulus, which can be generated with or without attention (Alho, 1995; Näätänen et al., 2007, 2019). The current understanding of the processes underlying MMN is that the repeated stimulus or stimulus pattern (standard) leads to construction of a central sound representation. This standard representation is then used to predict subsequent stimuli, and on encountering a stimulus that diverges from the prediction (deviant), the MMN is generated (Näätänen et al., 2011). Both within block MMN (deviant stimulus minus standard stimulus from the same block) and identity MMN (iMMN) (the stimulus used as the deviant in one condition minus the same stimulus used as the standard in the other condition) have been widely used in the literature. The iMMN response isolates the contextual effects of sound discrimination because it eliminates the difference in the auditory evoked potential (AEP) that is attributable to low-level acoustic differences in spectral, temporal, and intensity information (Jacobsen and Schröger, 2001, 2003; Kujala et al., 2007; Möttönen et al., 2013).

The Eulitz and Lahiri (2004) paper was, perhaps, the first study using neurophysiology that was directly designed to test a phonological model of representation. They tested the Featurally Underspecified Lexicon (FUL) model and used evidence from an asymmetric discrimination pattern of the MMN to test predictions derived from the model. In FUL, all phonetic features in the surface form are extracted from the acoustic signal, but at the level of the mental lexicon, some speech sounds will be underspecified for certain features (Eulitz et al., 1995; Lahiri and Reetz, 2002; Eulitz and Lahiri, 2004). Unspecified features are then filled in by other means because they are predictable. Eulitz and Lahiri (2004) predicted that presenting the vowel with the underspecified feature (in this case, default [+coronal]) as the standard stimulus, and the stimulus with the specified feature (in this case explicitly specified as [−coronal]) as the deviant stimulus would result in a smaller MMN. In other words, no conflict would be observed because no feature value is specified in the neural representation constructed to the standard for the underspecified member of the pair. They observed a smaller MMN to the vowel contrast when the underspecified member of the pair served as the standard. Since this paper, several other studies have tested the viability of underspecification models and generally have found support (Cornell et al., 2011, 2013; Scharinger et al., 2012; Hestvik and Durvasula, 2016; Cummings et al., 2017; Scharinger and Samuels, 2017; Højlund et al., 2019).

A number of studies that pre-dated Eulitz and Lahiri (2004), and that used MMN have observed patterns that bear upon the question of phonological asymmetry, even though they were not designed to specifically test a specific phonological theory. The earliest studies focused on questions related to categorical perception (e.g., Aaltonen et al., 1987, 1997; Sams et al., 1990; Maiste et al., 1995; Näätänen et al., 1997; Sharma and Dorman, 1998) or to auditory physiology (e.g., Kraus et al., 1992). The design of these first speech experiments was informed by a fairly large number of studies using MMN to examine auditory processing of non-speech features such as frequency, duration, and pitch (e.g., Kaukoranta et al., 1989; Nyman et al., 1990; Ritter et al., 1992; see Näätänen et al., 2007 for review).

Several of these early studies noticed and commented on the asymmetrical results of the MMN related to which auditory sound served as the standard and which served as a deviant. The practice of "flipping" the role of the two stimuli of interest was initially undertaken so that the event-related potential (ERP) to the deviant stimulus could be compared to the ERP to the same stimulus when it served as the standard (e.g., Sharma et al., 2004). The initial motivation for this exchange (e.g., flip-flop) was to control for differences in AEPs that indexed difference in timing of acoustic features. For example, the N1 amplitudes were larger for the vowel onset following a long-lag voice onset time (VOT) consonant than that following a short-lag VOT (Toscano et al., 2010), and VOT has also been shown to influence the latency of the N1 (M100) (Frye et al., 2007). In addition, different spectral properties of a stimulus will engage different neural populations in the primary auditory cortex. The longer inter-deviant interval compared to the inter-standard interval would then allow for greater recovery from neural refractoriness, which can be seen as an increased negativity of the N1b (e.g., McGee et al., 1997; Sharma et al., 2004). The question of what should serve as the control condition in the MMN paradigm to minimize these acoustic effects has sporadically been addressed, but is not directly a concern of the current paper (e.g., Cowan et al., 1993; Phillips et al., 1995; McGee et al., 1997; Sharma et al., 2004; Scharinger et al., 2012, 2016; Hestvik and Durvasula, 2016).

What is relevant for the current paper is that these first studies often observed an asymmetry in the amplitude of the MMN dependent on which of two contrasts served as the standard and which served as the deviant (e.g., Maiste et al., 1995; Shafer et al., 2004). For example, using a continuum of nine stimuli, Maiste et al. (1995) observed a clear MMN to cross-category stop consonant place deviants [da] only when [ba] served as the standard. Flipping the standard and deviant so that [da] was the standard stimulus resulted in no MMN, even to the phonetically most different stimulus [ba]. Maiste et al. (1995) attributed this finding to acoustic factors rather than to phonetic/phonological properties. Specifically, [da] showed a broader spectrum than [ba], and [da] contained the frequencies of the [ba] onset. But their finding of asymmetry is also compatible with the FUL model claim, in which [coronal] is the underspecified feature.

Shafer et al. (2004) observed a similar asymmetry to the Maiste et al. (1995) finding, but they suggested that the asymmetry was related to the linguistic property called markedness. They observed a larger MMN to the bilabial [ba] as a standard than to the dental or retroflex stop [da] as a standard (from a continuum of 10 stimuli from bilabial [ba] to retroflex [da]). Cross-linguistic surveys have observed that the retroflex category is less common and have characterized this as more "marked" (Maddieson, 1984), but this was a *post hoc* explanation. As pointed out by Haspelmath (2006), invoking markedness is not a satisfactory explanation, in part because the term has many different meanings in linguistics, but also because it is only a relabeling of the observation that the retroflexed category is somehow difficult to discriminate/produce compared to the bilabial category. In addition, the Shafer et al. (2004) study found that language experience modulated the asymmetry effect in that Hindi compared to naïve American English listeners showed a larger and earlier MMN to the [ba] deviant in the context of the retroflex standard.

Predictions derived from categorical perception (the original motivation for the study) cannot account for the cross-linguistic difference observed in Shafer et al. (2004), since for both groups the bilabial and retroflexed stops are perceived as distinct categories, as shown in the identification behavior of the participants in the study. The findings are compatible with either the acoustic explanation of Maiste et al. (1995) or with the underspecification approach of Eulitz and Lahiri (2004). Specifically, [coronal] (referring to the tongue tip/blade articulator) is argued to be the underspecified feature in many languages (Kiparsky, 1985; Paradis and Prunet, 1991), and if we accept this analysis for both Hindi and English, then the bilabial [ba] must be explicitly marked (that is [−coronal]). It is important to note that arguments used to determine which feature is underspecified in a language or across languages often make use of the notion of markedness.

Despite these findings consistent with linguistic accounts for asymmetry, it is important to fully consider psychophysical explanations. A few studies have demonstrated larger MMN to frequency increments compared to decrements using non-speech tones (e.g., Peter et al., 2010; Shiramatsu and Takahashi, 2018). The asymmetrical pattern observed for /ba/ and /da/ are consistent with this finding because /da/ has a higher frequency F2 formant onset than /ba/.

Asymmetries associated with speech sound length features have also been observed (e.g., Kirmse et al., 2007; Hisagi et al., 2010, 2015b; Chládková et al., 2015), and in some studies, have been attributed to psychophysical properties. However, the findings for speech are somewhat mixed. Several studies have observed a larger MMN to length increases for vowels or consonants irrespective of the phonological status of [length] in the language (Kirmse et al., 2007; Hisagi et al., 2010). The explanation for this finding was that duration increment is easier to discriminate than decrement, and that this pattern is acoustic in nature because it has also been observed for non-speech auditory information (Takegata et al., 2008). Findings from investigations of languages where length is a secondary cue for phonemic contrast (e.g., English and Dutch), however, suggest that phonological factors better explain asymmetries (Chládková et al., 2015; Shafer et al., accepted). For example, the study by Chládková et al. (2015) found evidence of an asymmetry only for the longer vowel /a:/, but not for the shorter /ɑ/ and suggested that this was consistent with specification of the long duration (and the short vowel being unspecified).

Another explanation that has been offered for asymmetries in speech sound processing and which could be extended to the findings in the MMN literature is the Natural Referent Vowel (NRV) framework proposed by Polka and Bohn (2003, 2011). According to the NRV, discrimination is easier when the reference (standard) is more centralized in the articulatory space compared to the change (deviant). This model makes similar predictions to Kuhl's perceptual magnet model, in which the better exemplar (prototype), when serving as the reference leads to poorer discrimination of a less prototypical vowel of the same category (in this case /i/) (Kuhl et al., 1992). Thus, both the NRV and the Perceptual Magnet model predict that MMN would be smaller when a more peripheral (prototypical) vowel serves as the reference (standard) compared to a more central (less prototypical) vowel. Two studies using MMN show support for the Perceptual Magnet effect (Aaltonen et al., 1997; Sharma and Dorman, 1998), although the pattern was modulated by whether a listener was good or poor at categorizing the speech stimuli in the study by Aaltonen et al. (1997).

A different explanation for asymmetries in MMN amplitude is related to experimental design. In a series of experiments, MMN was shown to be attenuated to a deviant stimulus if it had previously served as a standard (e.g., Todd et al., 2014). In addition, brief blocks that alternate which stimulus is the standard in the first half and which is the deviant in the second half can result in even greater attenuation of the MMN (Todd et al., 2014). Thus, it is possible that asymmetries could be an artifact of the order effect for any study that did not carefully counterbalance condition order.

The Present Study

The current study was designed to further address whether neurophysiological evidence is compatible with a model of underspecification. As we noted above, many studies that observed asymmetries in MMN did not predict these patterns, and rather, offered *post hoc* explanations. Thus, in the current study, we first looked for independent evidence that would allow predictions to be made with regards to the vowel contrast /ɛ/ vs. /ɪ/. We chose this contrast because we have used it to test a range of questions related to infant development, child language disorders and second language learning (e.g., clinical population: Shafer et al., 2005; Datta et al., 2010; developmental populations: Shafer et al., 2010, 2011; Yu et al., 2019; second language: Hisagi et al., 2015a; Datta et al., 2020). In all of these papers, which focused on group differences, /ɛ/ served as the standard and /ɪ/ as the deviant (and this order was selected based on pilot data that indicated a larger MMN for this direction; see Morr, 2002). The stimuli consisted of a resynthesized naturally produced vowel in which the F1 and F2 formants were edited to create a nine-step continuum from /ɛ/ to /ɪ/. Previous studies with adults and children indicated that step 3 and step 9 were consistently identified as /ɛ/ and /ɪ/, respectively and that they were both the same distance from the perceptual boundary (at step 6 on the continuum) (Shafer et al., 2005; Datta et al., 2010; Hisagi et al., 2015a).

We chose to use this simple paradigm of one token per category rather than using multiple tokens (e.g., Eulitz and Lahiri, 2004; Hisagi et al., 2010; Hestvik and Durvasula, 2016; Yu et al., 2017, 2018) because we wanted to be able to directly relate the findings to the previous studies where we used these stimuli. Importantly, speech sounds, whether presented as a single token, or as one of a set of tokens are processed at a phonological level. Research suggests that listeners automatically extract native-language phonetic features to allow for phoneme categorization (if the auditory information is sufficiently speech-like) (Strange, 2011). In the case of the discrimination process indexed by MMN, listeners rely on automatic selective perception routines, which reflect the native language phonological categories (Hisagi et al., 2010; Yu et al., 2017; Shafer et al., accepted). MMN will also index discrimination on the basis of acoustic factors, which can complicate interpretation. However, our study comparing MMN to this vowel contrast between monolingual English speakers and bilingual Spanish–English speakers clearly indicated that these vowel stimuli engage phonological processing (Hisagi et al., 2015a).

The current study was designed to examine whether there is an asymmetry to this vowel contrast in native English-speaking adults. Several studies lead to the claim that /ɛ/ is the underspecified vowel of the pair. Stemberger (1991a,b) found evidence from speech errors supporting that English /ɛ/ is the underspecified vowel (with underspecification of [−high] [−low] [−back] [−round]). Stemberger (1992) compared the mispronunciation rates of vowel contrasts involving [high] and [back] features (e.g., /ɪ/-/ɛ/, /ɪ/-/ʌ/, /ɛ/-/ʌ/). He observed significantly more errors involving the non-high vowel /ɛ/ being replaced by the high vowel /ɪ/ than the other way around. He also examined whether type frequency (number of English words with a specific phoneme) would account for the error pattern. If so, vowels with higher type frequency should be more resistant to speech errors. His findings, however, were incompatible with this word frequency account. Another study using MMN also supports the claim that /ɛ/ is underspecified (Scharinger et al., 2012). They tested the proposal that [high] and [low] features of vowels are specified in the underlying representation, and thus /ɛ/, as a mid vowel, has no specification for these values. They observed a larger MMN when a low vowel /æ/ served as the standard and /ɛ/ as the deviant compared to the MMN to the contrast presented in the reverse direction. In addition, they observed no asymmetry in the MMN for /ɪ/ vs. /æ/, which both are marked for height in the model, but for different features ([high] for /ɪ/ and [low] for /æ/).

Taken together, these results lead to the hypothesis that the MMN will be larger when /ɪ/ serves as the standard and /ɛ/ as the deviant because /ɪ/ is marked as [+high] and /ɛ/ has no height specification. The NRV model predicts a larger MMN when the more peripheral vowel /ɪ/ serves as the standard. The acoustic account does not lead to a clear prediction that one order will show a larger MMN because vowels are both the same in duration; the spectrum of these two vowels differs minimally. In considering the direction of frequency change (in terms of increment or decrement), F2 is higher for /ɪ/ than /ɛ/, but the reverse pattern is found for F1. Thus, the two directions of change conflict with regards to predicting direction of asymmetry. We also examined whether the order of presentation of the conditions would modulate the MMN, predicting a smaller MMN to the vowel which was first presented as a standard. Finally, we performed analysis comparing the deviant stimulus to the standard within a condition (cross-block MMN, comparing ERPs to two acoustically-different stimuli) and comparing the deviant to the standard across conditions (iMMN from the identity stimuli, where the ERPs are to the same stimulus serving as the deviant in one condition and the standard in the other condition). We chose to undertake the analysis in both ways to allow an explicit comparison of how these different methods affect the MMN and to allow us to relate our findings here to our previous studies.

MATERIALS AND METHODS

Participants

Twenty young adult participants between the ages of 19 and 27 years old were recruited and provided written informed consent. Data from two participants were collected using the incorrect sampling rate, and data from one participant were too noisy. Data from the remaining 17 participants (Male = 5) were included in the statistical analyses. All participants passed a standard hearing screening in the laboratory and reported no history of hearing/speech-language/neurological/developmental impairment. All participants were native English speakers, seven of them were monolingual English speakers, and ten had some Spanish exposure from family and/or had taken regular school Spanish as second language classes, but all reported dominance in English. Each participant was paid $10 per hour for participation. The study was approved by the human subject research institutional review board at St. John's University, New York, and was conducted in compliance with the Declaration of Helsinki.

Stimuli

Two English vowels /ɪ/ (as in the word *bit*) and /ɛ/ (as in the word *bet*) were used in the study. To create the vowels, a natural token of a neutral vowel /ʌ/ was produced by a female with an F0 of approximately 190 Hz. This vowel was resynthesized and edited using target formant frequencies based on natural productions of /ɪ/ and /ɛ/ from the same speaker using Analysis by Synthesis Lab, version 3.2 (see Shafer et al., 2010, 2011 for details). A nine-equal-step continuum was created using equal steps for the first formant (F1) and second formant (F2). The two tokens for this study were Step 3 and Step 9 on the continuum and were selected to be equidistant from the boundary (determined in piloting). These stimuli had the following mean center frequencies: for Step 3, F1 = 500, F2 = 2160; for Step 9 F1 = 650 Hz, F2 = 1980 Hz. The two stimuli had an identical duration of 250 ms and identical third (F3 = 2174) and fourth (F4 = 3175) formants. Step 3 was identified as /ɪ/ and Step 9 as /ε/, respectively, by both monolingual English-speaking children and adults in studies from our laboratory (e.g., Datta et al., 2010; Hisagi et al., 2015a). **Figure 1** shows the waveforms and power spectrum of the two stimuli.

The ERP Paradigm

The stimuli were presented using over-the-ear headphones at a comfortable listening volume. Two blocks (one condition per block) of 1000 stimuli (20% deviant) were presented in a counter-balanced order across participants at the rate of 650 ms (interstimulus interval of 400 ms). In one condition Step 3 /ɪ/ was the standard, and Step 9 /ɪ/ was the deviant, and in the other condition the standard and deviant were switched. Thus, a total of 200 deviant and 800 standard trials were delivered for each stimulus. The data were recorded in an electrically shielded and sound-attenuated booth, and the participants were watching a muted movie with captions played over a handheld tablet during the data recording for the purpose of keeping the participants occupied and directing the participants' attention away from the auditory speech sounds. All the participants watched the same movie. This procedure is commonly used in MMN designs because it engages the participant's attention for the long period of time needed to obtain a sufficient number of trials for good signal to noise ratio.

The experiment was programmed with the E-prime 2 Professional software (version 2.0.10.356) (Psychology Software Tools) to deliver stimuli. The data were acquired and digitized via Netstation software version 5.4. The 65-channel HydroCel sensor nets from Electrical Geodesics, Inc., 400 system, using Ag/AgCl plated electrodes housed in electrolyte-soaked sponges were placed on the participant's scalp. The impedances of the electrodes were kept at or below 50 kΩ. The EEG was recorded using a bandpass filter of 0.1–100 Hz and sampling rate of 1000 Hz with Cz as the reference electrode. The continuous EEG waveforms were processed offline, using a bandpass filter of 0.3–30 Hz in Netstation version 5.4, and were then segmented into epochs of 200 ms pre-stimulus and 800 ms post-stimulus periods. BESA Research 6.0 (BESA GmbH, 2014) was used for further offline processing. Automatic artifact correction was applied to each participant's data using an HEOG threshold of 150 μV and a VEOG threshold of 250 μV for eye movement noise, and thresholds for bad channels were set at 120 μV for amplitude (gradient of 75). The individual averaged data were referenced to an average reference. There was no significant difference in the number of accepted trials by condition (/ɪ/ deviant: mean = 173 trials, SD = 18; /ε/ deviant: mean = 175 trials, SD = 21; /ɪ/ standard: mean = 505 trials, SD = 64; /ε/ standard: mean = 510 trials, SD = 46; p-values > 0.74).

ERP Analysis

The data were downsampled to 20 ms per data point (after filtering using the engineer's Nyquist). Permutation analyses were used to control the multiple comparison problems that commonly arise in parametric statistical procedures (e.g., multiple t-tests, analysis of variance) when these involve a large number of statistical comparisons (e.g., multiple correlated sensor sites and correlated time points) (Maris and Oostenveld, 2007). This approach reduces the rate of false positives (Type I error) in ERP data analyses (Lage-Castellanos et al., 2010). Permutation tests also have the advantage of making no

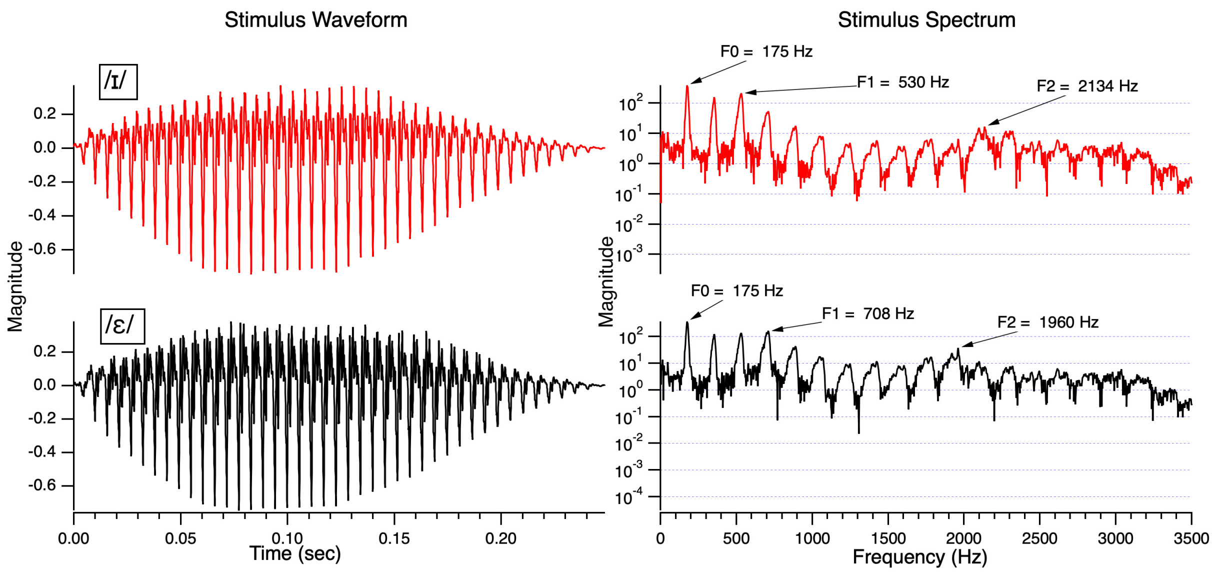


FIGURE 1 | The waveforms of /ɪ/ and /ε/ on the left and the corresponding spectra on the right. The stimuli were resynthesized and edited to retain the natural bandwidth of each formant and the natural change in F0 across the stimulus (higher at the onset and falling off at the offset).

assumptions about the distribution of the data. The test was performed in Rstudio using the RVAideMemoire package.

To examine the features of MMN, we utilized a two-step sequential temporo-spatial principal component analysis (PCA) to determine the time window of analysis and electrodes to include for analysis (Dien, 2010). We then used the time window and electrode sites obtained from the PCA to examine the MMN amplitude effects. The two-step sequential temporo-spatial PCA has the advantage of objectively identifying time windows and electrode regions for examining the target effects (Dien and Frishkoff, 2005; Dien, 2010, 2012). This is a mathematical way to isolate the underlying latent ERP components in the temporal and spatial domain so that there is no need to subjectively select the time window and electrode sites for further analysis (Luck and Gaspelin, 2017). We used the difference waves as the input for the PCA to better focus on the temporal and spatial features of the mismatch itself (Handy, 2005; Hestvik and Durvasula, 2016). Based on the PCA results, we averaged the amplitudes of the electrode sites for each temporal-spatial component, and adopted six time-bins of 20 ms, centered around the peak of each temporal-spatial component as the time of interest. Permutation ANOVAs using stimulus condition (/ɛ/ vs. /ɪ/) and time as the independent variables were performed to determine the main effects of stimulus condition and the interaction between the two variables. The iMMN was generated by subtracting the standard from the deviant of the same stimulus (deviant /ɪ/–standard /ɪ/; deviant /ɛ/–standard /ɛ/). Our main focus was the iMMN. We also examined the identity late discriminative negativity (LDN) based on the results of the PCA. Finally, we examined the MMN asymmetry within the same block and presentation order effect to allow our findings to be directly related to our prior studies and the NRV model; to do this, MMN was generated by subtracting the standard from the deviant in the same block (deviant /ɛ/–standard /ɪ/; deviant /ɪ/–standard /ɛ/).

The main effect of stimulus was followed up by a permutation Student's *t*-test. Permutation ANOVAs and permutation Student's *t*-tests were also used to examine whether there was any order effect between those who heard /ɛ/ as the standard first vs. those who heard /ɪ/ as the standard first using the iMMN. We chose to include this factor because a number of studies have shown attenuation of the MMN to the deviant if it previously had occurred as a standard (McGee et al., 2001; Todd et al., 2014). Nine participants heard /ɛ/ as the standard first, and eight participants heard /ɪ/ as the standard first.

RESULTS

Results of the PCA for Identity MMN/LDN (Deviant /ɛ/–Standard /ɪ/; Deviant /ɪ/–Standard /ɛ/)

The goal of the PCA was to identify a set of sites and the time range that contributed to the MMN and LDN that would not be biased for one analysis approach. These sites would then be used to construct unbiased measures to test the question of whether MMN and LDN amplitudes are different depending on which stimulus serves as the standard and which serves as the deviant.

The analyses were performed using the EP tools by Dien (2010). First, a temporal Promax rotation with a covariance relationship matrix and Kaiser weighting (kappa = 3) was performed followed by the Scree test in combination with the Parallel Test (Horn, 1965), which compares the Scree of the dataset to that generated from a fully random dataset. We retained components that had a total variance larger than 5%; only the first four temporal components met this criterion. For the temporal PCA, 25 temporal components were retained, which accounted for a total variance of 81%, with the first four components each having variance accounted for greater than 5%. The time windows and variances accounted for (in parenthesis) for these four components are TF1 at 520 ms (23%), TF2 at 781 ms (15.6%), TF3 at 297 ms (10.4%), and TF4 at 159 ms (6.5%). These temporal factors were submitted to the spatial infomax rotation, which resulted in five spatial factors for each temporal component.

The TF2 factors were discarded because they were outside the time window of MMN. Only TF1SF5, TF3SF3, and TF4SF3 showed a spatial distribution with maxima at the fronto-central regions. TF1SF5 showed the maximal negativity at site 51 (slightly anterior to the midpoint of Cz and C4) with inversion near site 55 (adjacent to the mastoid). The sites that had factor loadings >0.6 for TF1SF5 were only sites 51 and 41. TF3SF3 showed the maximal negativity at site 4 (FCz) and inversion at site 10 (Fp1). The sites that had factor loadings >0.6 for TF3SF3 were 4, 7, 16, 20, 41, 50, 51, 53, 54, and 65. TF4SF3 showed the maximal negativity at Cz, which is slightly posterior to what is expected for MMN, but it was the only factor that was in the temporal window of MMN and that also showed contribution from frontal sites and inversion at inferior sites (maximal inversion at site 1/F10). The sites that had factor loadings >0.6 for TF4SF3 were 4, 7, 15, 16, 41, 50, 51, 53, 54, and 65, which were highly similar to the sites for TF3SF3. Therefore, three PCA components (TF4SF3, TF3SF3, and TF1SF5) were selected, and sites with loadings >0.6 for each component were averaged to derive one measure corresponding to each of the three PCA components and then used in the subsequent statistical analyses. TF4SF3 peaked at 159 ms, which fell within the typical timeframe for MMN and TF1SF5 peaked at 520 which fell within the LDN time window reported in prior literature. The TF3SF3 peaked at 297 ms, falling in a later time frame than generally reported for MMN, and somewhat early for LDN. We selected the time window for each derived measure to correspond to the onset and offset latencies where the amplitude was 1/2 of the peak amplitude value observed for the PCA component. Because we downsampled by intervals of 20 ms, the onset and offset of the selected interval was the nearest value (for example, if the 1/2 amplitude onset was 109 ms, then the interval onset was 100 ms).

Identity MMN and LDN Results

TF4SF3-Derived Measure

The 100–220 ms interval was selected (see **Figure 2**). The permutation ANOVA using condition (/ɛ/ vs. /ɪ/) and time

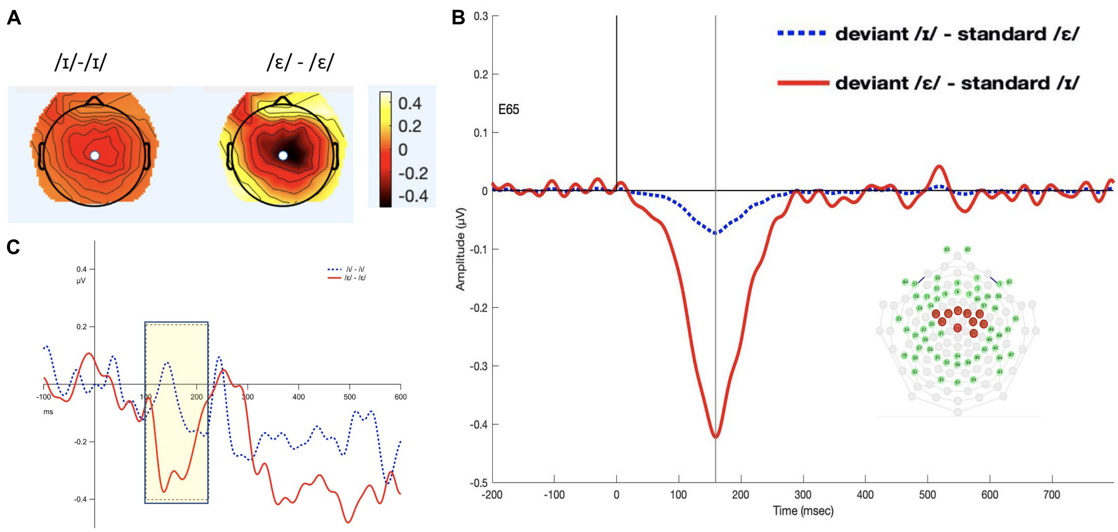


FIGURE 2 | Panel **(A)** shows the topomap of the principal component TF4SF3. Panel **(B)** shows the waveforms of the iMMNs from the principal component TF4SF3. The blue dotted line indicates the /ɪ/ condition, and the red solid line indicates the /ɛ/ condition. The PCA component peaked at 159 ms, and the sites with loading larger than 0.6 were 4, 7, 15, 16, 41, 50, 51, 53, 54, and 65 (electrodes marked red on the figure of the EGI 65 channel sensor net). Panel **(C)** shows the waveforms of the scalp event-related potentials for the iMMNs for TF4SF3. The blue dotted line indicates the /ɪ/ condition, and the red solid line indicates the /ɛ/ condition. Time window of statistical analysis is marked with a shaded rectangle.

(six, 20-ms time bins) as the independent variable revealed a significant main effect of condition ($F_{1,192} = 3.962$, $p < 0.05$). The time main effect ($F_{5,192} = 0.3676$, $p = 0.88$) and time by condition interaction ($F_{5,192} = 0.9577$, $p = 0.44$) were not significant. Follow-up permutation Student's t-test on the condition effect revealed that the amplitude of MMN for /ɛ/ was significantly more negative than for /ɪ/ ($t = 5.424$, $p < 0.001$).

TF3SF3-Derived Measure

The interval 260–380 ms was selected (see **Figure 3**). The permutation ANOVA using condition (/ɛ/ vs. /ɪ/) and time (six, 20-ms time bins) as the independent variable found that neither main effects nor the interaction was significant (condition: $F_{1,192} = 0.098$, $p = 0.75$; time: $F_{5,192} = 1.0626$, $p = 0.38$ condition × time: $F_{5,192} = 0.2820$, $p = 0.41$).

TF1SF5-Derived Measure

The interval 460–580 ms was selected (see **Figure 4**). The permutation ANOVA using condition (/ɛ/ vs. /ɪ/) and time (six, 20-ms time bins) as the independent variable revealed a significant main effect of condition ($F_{1,192} = 20.98$, $p < 0.001$). The time main effect ($F_{5,192} = 0.2630$, $p = 0.93$) and time by condition interaction ($F_{5,192} = 0.1225$, $p = 0.99$) were not significant. Follow-up permutation Student's t-test on the condition effect revealed that the amplitude of LDN for /ɛ/ was significantly more negative than for /ɪ/ ($t = 3.9447$, $p < 0.001$).

Within-Condition MMN (Deviant /ɛ/–Standard /ɪ/; Deviant /ɪ/–Standard /ɛ/)

The analysis approach was identical to the iMMN. Twenty-one temporal factors were retained, which accounted for a total variance of 83.3%. The temporal factors were then submitted for spatial decomposition by using the spatial infomax rotation method. The first three temporal components (TF1, TF2, and TF3) peaked at 650, 298, and 159 ms, and accounted for 36.5, 14.3, and 8.9% of the variance, respectively. No other components had variance accounting for larger than 5%. Only TF3 was in the expected time window of the MMN. Therefore, TF1 and TF2 were discarded from further analysis. Spatial PCA revealed that TF3SF4 showed a spatial pattern over fronto-central regions that was consistent with the MMN topography with the maximal negativity at site 54 (midpoint between Cz and F4) and negativity at inferior posterior regions. The sites with factor loadings >0.6 for TF3SF4 were 4, 41, 50, 51, 53, 54, and 65, and these sites were averaged to derive the measure corresponding to TF3SF4.

TF3SF4-Derived Measure

The interval 100–220 ms was selected (see **Figure 5**). The results from the permutation ANOVA revealed a significant condition effect ($F_{1,192} = 3.9618$, $p = 0.05$). Neither the main effect of time nor the interaction between time and condition was significant (time: $F_{5,192} = 0.3676$, $p = 0.87$, condition × time $F_{5,192} = 0.9577$, $p = 0.45$). Follow-up permutation Student's t-test on the condition effect revealed that the MMN amplitude was larger when the deviant /ɪ/ was subtracted from /ɛ/ than when the deviant /ɛ/ was subtracted from /ɪ/ ($t = -1.673$, $p = 0.04$).

Order Bias

The overall average waveforms from the two standard and deviant conditions were presented in **Figure 6**. The first time interval (100–220 ms) was selected to examine order bias effect. The results from permutation ANOVA on the subgroup of participants ($N = 8$) who received /ɛ/ as the standard condition

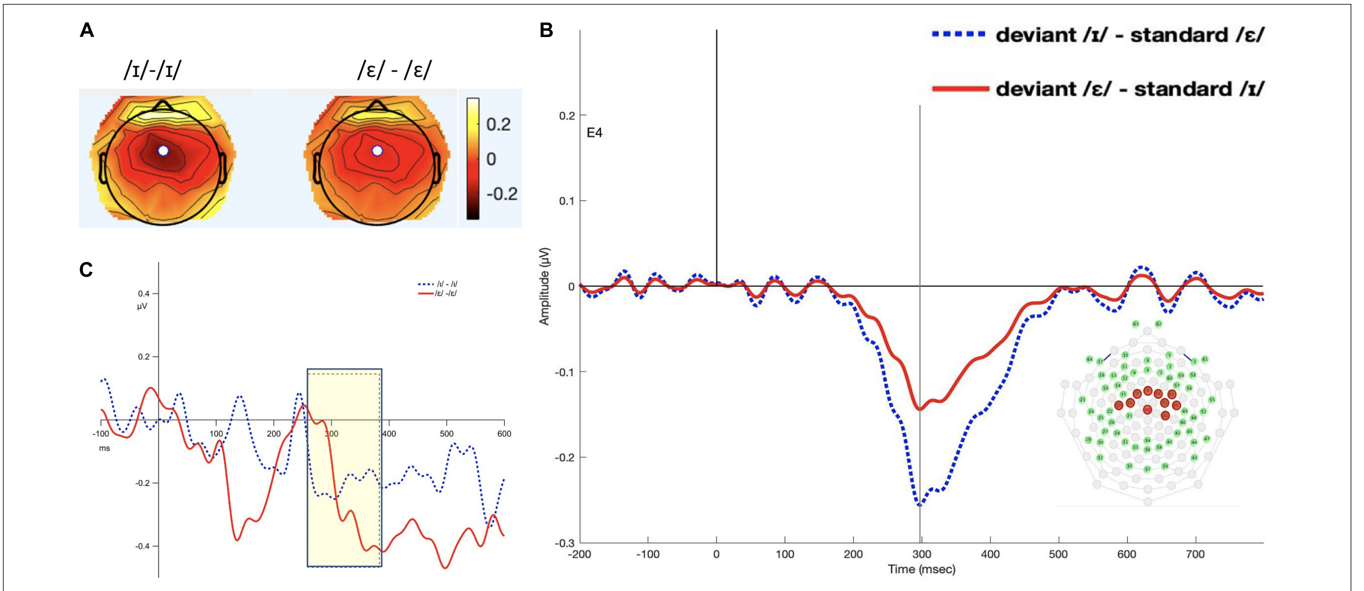


FIGURE 3 | Panel **(A)** shows the topomap of the principal component TF3SF3. Maximal negativity is at site 4 (white dot). Panel **(B)** shows the waveforms of the iMMNs from the principal component TF3SF3. The blue dotted line indicates the /ɪ/ condition, and the red solid line indicates the /ε/ condition. The PCA component peaked at 297 ms, and the sites with loading larger than 0.6 were 4, 7, 16, 20, 41, 50, 51, 53, 54, and 65 (electrodes marked red on the figure of the EGI 65 channel sensor net). Panel **(C)** shows the waveforms of the scalp event-related potentials for the iMMNs for TF3SF3. The blue dotted line indicates the /ɪ/ condition, and the red solid line indicates the /ε/ condition. Time window of statistical analysis is marked with a shaded rectangle.

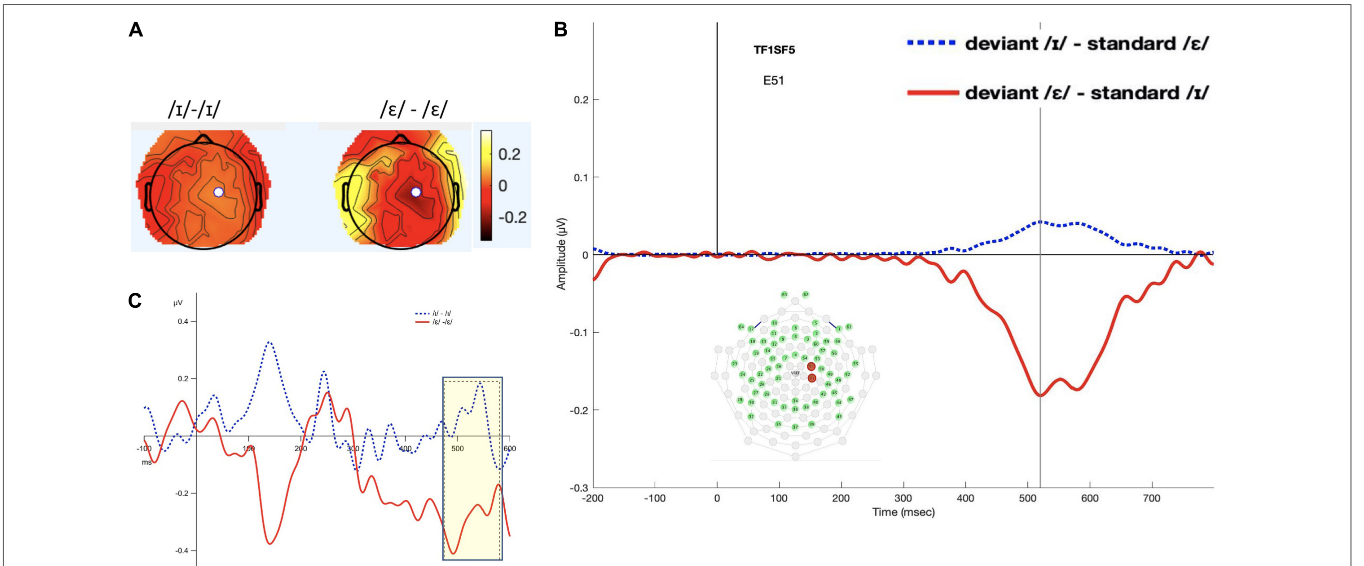


FIGURE 4 | Panel **(A)** shows the topomap of the principal component TF1SF5. Panel **(B)** shows the waveforms of the iMMNs from the principal component TF1SF5. The blue line indicates the /ɪ/ condition, and the red line indicates the /ε/ condition. The PCA component peaked at 520 ms, and the sites with loading larger than 0.6 were 41 and 51 (electrodes marked red on the figure of the EGI 65 channel sensor net). Panel **(C)** shows the waveforms of the scalp event-related potentials for the iMMNs for TF1SF5. The blue dotted line indicates the /ɪ/ condition, and the red solid line indicates the /ε/ condition. Time window of statistical analysis is marked with a shaded rectangle.

first showed that deviant /ɪ/ generated larger MMN than deviant /ε/ ($p < 0.001$). However, for the subgroup ($N = 9$) who received /ɪ/ as standard first, deviant /ɪ/ generated smaller MMN than deviant /ε/ ($p < 0.001$). That is, the deviant stimulus that was used as the standard first led to reduced MMN (**Figure 7**).

DISCUSSION

Main Findings

In the present study, we investigated the modulation of the amplitude of MMN to the English vowel contrast /ε/ and /ɪ/ in

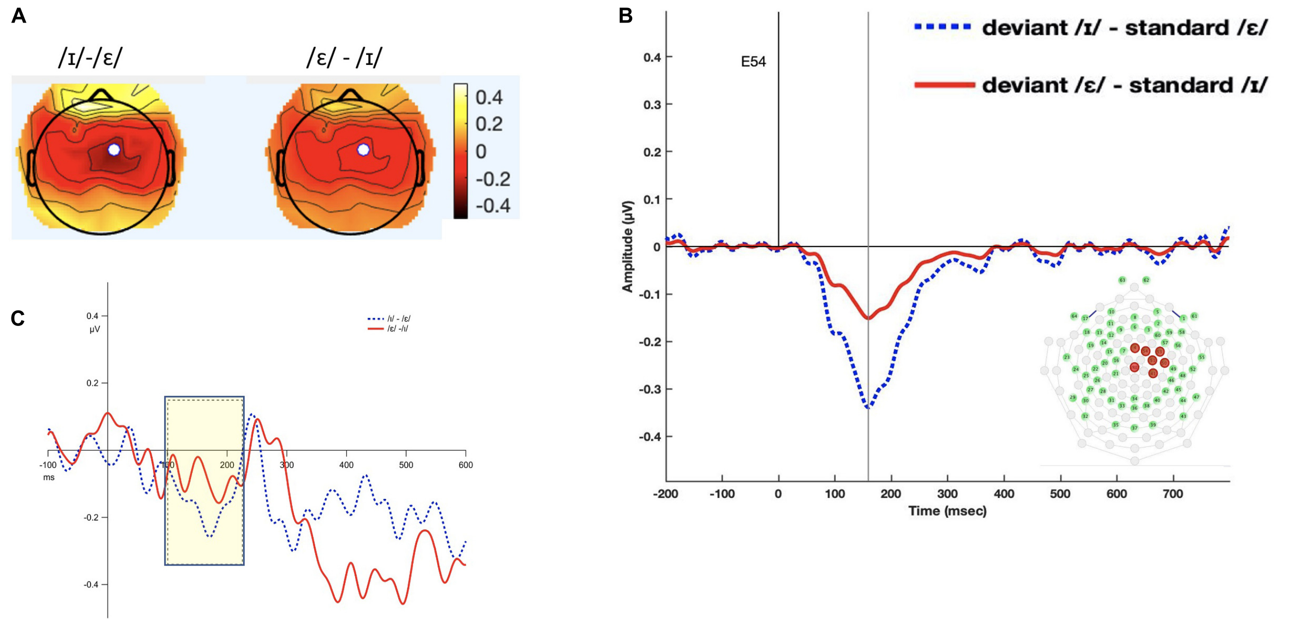


FIGURE 5 | Panel **(A)** shows the topomap of the principal component TF3SF4 for the within-block MMN. Panel **(B)** shows the waveforms of the within-block MMNs from the principal component TF3SF4. The blue line indicates the /ɪ/ condition, and the red line indicates the /ɛ/ condition. The PCA component peaked at 159 ms, and the sites with loading larger than 0.6 were 4, 41, 50, 51, 53, 54, and 65 (electrodes marked red on the figure of the EGI 65 channel sensor net). Panel **(C)** shows the waveforms of the scalp event-related potentials for the within-block MMNs for TF3SF4. The blue line indicates the /ɪ/ condition, and the red line indicates the /ɛ/ condition. Time window of statistical analysis is marked with a dashed rectangle.

adult native English speakers using a passive auditory oddball paradigm. We had predicted that the presentation of the less specified vowel /ɛ/ as the frequent stimulus and the more specified vowel /ɛ/ as the infrequent stimulus would lead to a smaller MMN than when flipping the stimulus probability for the two vowels (Stemberger, 1991a,b; Scharinger et al., 2012). Our results were consistent with this claim, in the cross-condition comparison, in that both the iMMN and LDN were larger to deviant /ɛ/ than to deviant /ɪ/ when compared to themselves serving as the standards (deviant /ɛ/–standard /ɛ/; deviant /ɪ/–standard /ɪ/). However, when comparing the deviant to the within-condition standard (which was a different stimulus), the opposite pattern was observed; specifically, a larger MMN was observed to /ɪ/ as the deviant (minus /ɛ/ as the standard) than for the reverse of /ɛ/ as the deviant (minus /ɪ/ as the standard). This within-condition finding is consistent with the NRV model, in which perceptual discrimination is easier when the reference (standard) is a more central vowel. This finding of consistency with both models is not really contradictory because in both calculations the MMN was subtracted from the same standard /ɛ/. In addition, as predicted, we found an order effect, showing a smaller MMN to the deviant stimulus, if it had previously served as the standard. This order effect, however, did not influence our main findings, because we had roughly equal participants in each order (8 and 9 per order).

Support for Underspecification

Models of underspecification predict asymmetry of processing and perception. However, one drawback with these models is that the decision with regards to which features are underspecified is often circular. More specifically, underspecified features are proposed to be those that are the least "marked" (e.g., more frequent in the languages of the world, less difficult to learn in child language, perceptually more salient, articulatorily easier, etc.) Often, a feature is declared underspecified because of one or more of these patterns. What is often lacking in these accounts is a unified and consistent proposal for how markedness should be determined (e.g., should it be derived from physical constraints of perception and production or from the language-specific system?) These questions have been grappled with by many linguists (e.g., Haspelmath, 2006; Scharinger and Samuels, 2017), and the current experiment does not provide a solution. Rather, the findings of our study add another piece of evidence showing asymmetries of speech processing that are not easily explained on the basis of acoustic properties.

The stimuli used in the current study were closely matched (via resynthesis and editing) so that the only cues available were the F1 and F2 formants. On a physical scale, both /ɪ/ and /ɛ/ are centralized compared to the closest tense vowels /i/ and /e/. Thus, on the basis of acoustic-phonetic details, one would predict that these vowels should be equally well-represented at the cortical level, and as a result, we should see no directional difference in the MMN amplitudes for the two vowels. We did, however, see an asymmetry.

Thus, our findings are consistent with a model of underspecification, such as the FUL model (Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002, 2010). A few fairly recent studies using the MMN measure have also supported the model of underspecification. Hestvik and Durvasula (2016)

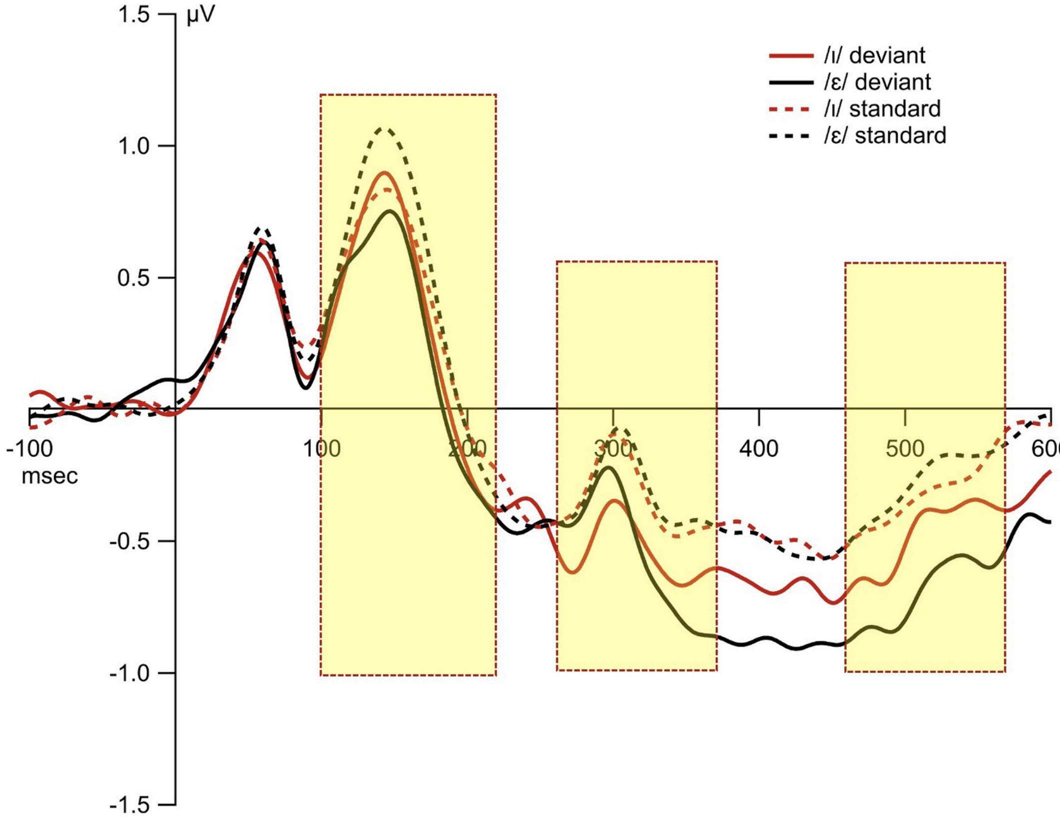


FIGURE 6 | The waveforms from the standard and deviant conditions. The 10 fronto-central sites were chosen based on principal component analysis, and were averaged for display purposes. The three shaded rectangles indicate the time window of statistical analyses based on the principal component analysis.

replicated Phillips et al.'s (2000) asymmetry findings on the consonant voicing contrast /da/ and /ta/. They observed a clear asymmetry with a significant MMN for the /da/ deviant, but not for /ta/ deviant when multiple exemplars of stimuli for each category were used. Similarly, for the LDN, they observed a larger effect for the /da/ deviant compared to the /ta/ deviant. They had predicted this pattern from the claim that English voiceless stops are phonologically specified for spread glottis ([+spread]), and voiced stops are underspecified. Cummings et al. (2017) also found asymmetry of MMN responses for [coronal] underspecification in the English /da/-/ba/ contrast (also Shafer et al., 2004). But this pattern is consistent with an acoustic explanation (Maiste et al., 1995), as well as the underspecification account. However, lending increased support to a phonological account was the finding that Japanese listeners revealed an MMN amplitude asymmetry in the opposite direction of English speakers, using the same stimuli as Hestvik and Durvasula (2016) (Hestvik et al., 2020). This latter finding cannot easily be explained in terms of the acoustic properties of the stimuli. Hestvik et al. (2020) argue that their findings support an underspecification account, but it will also be important to consider the finding in relation to a prototypicality effect.

Our finding that the amplitude of MMN is larger for /ɛ/ than /ɪ/ was consistent with Scharinger et al. (2012). They observed that MMN was larger to /æ/ as the standard and /ɛ/ as the deviant and had proposed that /ɛ/ was underspecified for vowel height, whereas /æ/ and /ɪ/ were specified. Consistency in the pattern for these two studies, however, does not preclude acoustic-phonetic factors underlying this pattern. The finding related to speech production from Stemberger's (1991b) study provides somewhat stronger support for a linguistic-internal explanation because there is no reason that the asymmetrical pattern should necessarily be consistent across the two different modalities, unless they are linked in some way via phonological representations.

Support for the NRV Model

Interestingly, our results also can be taken as support for the NRV model (Polka and Bohn, 2003, 2011). In behavioral perception studies, the data are accuracy or response times to a target stimulus. In an oddball paradigm, the target stimulus is the infrequent stimulus. The studies that have observed an asymmetry in perception generally presented stimuli in a habituation paradigm (for infants) or match-to-sample task (stimulus pairs where the participant determines whether the second stimulus is the same or different from the first). Thus, the asymmetry is observed as having higher accuracy scores (or detection of change in the infant studies) when the referent was the more central vowel. The within-condition subtraction (e.g., /ɪ/ minus /ɛ/) method more directly matches this behavioral

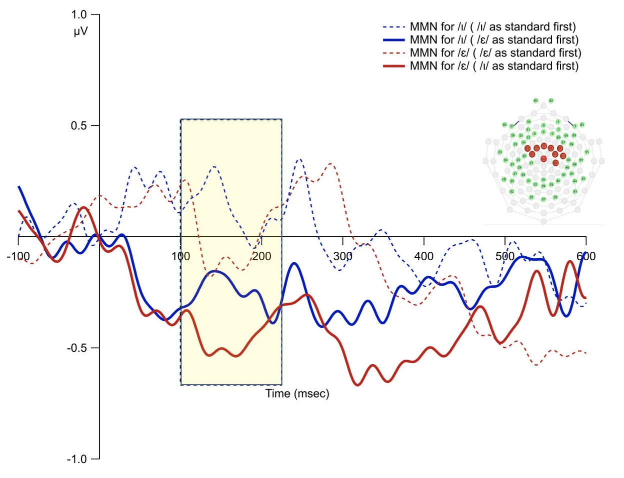


FIGURE 7 | The MMN waveforms from the averages of the 10 fronto-central sites for the subgroup who heard the /ɛ/ as standard first vs. those who heard the /ɪ/ as standard first. The shaded rectangle indicates time window of analysis based on the principal component analysis.

paradigm. This method is sometimes used in MMN designs when there is a time limitation (for example infant studies). For the within-condition subtraction in the current study, the MMN amplitude was larger when the deviant was /ɪ/ (/ɪ/ minus /ɛ/) than when the deviant was /ɛ/ (/ɛ/ minus /ɪ/). The NRV posited that vowels with more extreme articulatory-acoustic properties acts as NRVs (e.g., high front vowel /ɪ/ in this experiment), and a vowel change from a more central to a more peripheral position is easier to discriminate than the reversed direction of change. This claim is not necessarily inconsistent with an underspecification account. Specifically, the NRV is consistent with the claim that the peripheral vowels are less marked and could, thus, predict which features are underspecified. Note, however, that the NRV framework has primarily been supported by infant data. Polka and Bohn (2011) emphasized that asymmetrical perceptual bias changes as infants' language experience increases. Attunement to the first language leads to an attenuation of the default bias favoring peripheral vowels. But given that the perceptual bias favoring vowels in the peripheral spaces is grounded in the acoustic patterns that have an "easy, privileged fit with human auditory/articulatory abilities," this bias may re-emerge under degraded listening conditions to a native phonemic contrast (Polka and Bohn, 2011).

Category Goodness

Another explanation that could account for asymmetrical patterns is related to the within-category speech sound structure. Phonemic categories include phonetic variation. Kuhl et al. (1992) proposed that the more prototypical vowel of a speech sound category will be less discriminable from other category members that are less prototypical (Kuhl et al., 1992; Iverson and Kuhl, 1995). In their studies, pairs of stimuli closer to the prototype were more difficult to discriminate than pairs further from the prototype. Other behavioral studies of speech perception have also observed asymmetries in vowel perception. For example, Cowan and Morse (1986) observed better discrimination (in an AX task) of the vowel pair /ɪ/ vs. /i/ when the second in the pair was the more peripheral /i/ in a short interstimulus interval. The effect was greater for longer interstimulus intervals (2-s) between vowels in a pair. They argued that the memory for a vowel was represented as a small, bounded area within the vowel space, and when memory decayed to the first stimulus (A), the representation of the boundary for the vowel space would expand over time, which leads to a shift toward a more centralized vowel. Note that these findings are consistent with the predictions of the NRV model, but that the explanation for the asymmetry is different.

The Kuhl et al. (1992) study also observed a language experience effect in that 6-month-old infants showed a stronger magnet effect for the language-specific prototype of the ambient language (/i/ for Engish and /y/ for Swedish). Thus, language experience influences the internal structure of phoneme categories. It is well-established that early language-specific experience has shaped adult speech perception and neural processing (Näätänen et al., 2007; Shafer et al., accepted). For an underspecification model to show viability, it needs to explain cross-linguistic differences as well as universal patterns (e.g., Hestvik et al., 2020). It is not clear that it can do this better than a prototype model. For example, the pattern of MMN asymmetry observed in Shafer et al. (accepted) showed modulation by language-specific experience (English vs. Spanish first language) where the duration decrement from longer /ɑ:/ as the standard to the shorter /ʌ/ as the deviant showed a larger MMN than the reverse for Spanish listeners, with no amplitude difference observed for American English participants. These results cannot be easily explained in terms of underspecification. Berent et al. (2007) suggested that universal patterns of markedness are visible in non-native listener's perception. More marked forms are less common across languages. Following this logic, /ʌ/ is more likely to be "marked" than /ɑ:/ because it is the less common vowel across languages and, thus, /ʌ/, as the "specified" form would result in a larger MMN for non-native listeners (all else being equal). The finding with the Spanish listeners, however, did not support this prediction. Rather, the Spanish result is more consistent with a prototype model in which the long interval between repetitions of the American English /ɑ:/ or /ʌ/ (about 1300 ms) resulted in decay in short term memory and the "filling in" of the standard representation with a prototypical Spanish /a/. In consequence, the phonetically more-similar English /ɑ/, as the deviant was not discriminable. In contrast, the American English /ʌ/ as the deviant was sufficiently different from this Spanish vowel representation to allow for discrimination and elicitation of the MMN.

Our finding of an asymmetry in processing could possibly be related to one of the two vowels being a better match to the category prototype. We did not evaluate this in the current study, although identification data from Hisagi et al. (2015a) suggest that the /ɪ/ stimulus used in this MMN study was a better exemplar than the /ɛ/ stimulus. In this case, the prototype model would predict better discrimination when /ɛ/ was the standard and /ɪ/ was the deviant. The trace-decay model of Cowan and Morse (1986), in which the English vowel representations shift centrally would also predict better discrimination for this direction.

Acoustic Explanation

It was less clear that our findings can be taken as consistent with an acoustic-phonetic explanation. However, we cannot completely dismiss the possibility that the higher F2 of /ɪ/ somehow is acoustically more important than the F1. To test this, we would need to use non-speech counterparts with complex tones to examine how the pattern for the higher harmonics modulates MMN. It is clear that psychophysical properties of auditory information do modulate processing (Kirmse et al., 2007; Hisagi et al., 2010), so it is important to fully consider these alternative explanations. Furthermore, some phonetic contrasts are physically more different than others, and given that MMN is larger and earlier to greater physical differences, it is possible that asymmetry is minimized. Thus, the absence of an asymmetry in MMN to /ɪ/ vs. /æ/ in the study by Scharinger et al. (2012) might simply be due to the greater physical difference between this vowel pair. It may also be that psychophysical properties and linguistic experience differentially affect the amplitude vs. latency of the MMN (Näätänen et al., 1997; Shafer et al., accepted). For example, Shafer et al. (accepted) observed an earlier latency of the MMN for an American English vowel duration increment from standard /ʌ/ to deviant /ɑː/ than for a duration decrement from standard /ɑː/ to deviant /ʌ/ for English, Japanese and Russian listeners. Only a Spanish group of listeners did not show this latency difference because they had no MMN to the duration increment.

Other Factors

Group differences in the MMN can be driven by the differences in the standard or deviant alone or both combined. Cummings et al. (2017) found no amplitude differences between the deviant /ba/ and /da/ in the MMN time window, but they did observe a more positive response to the standard /da/ than the standard /ba/. The larger responses were interpreted to indicate a larger neuronal population firing to the coronal place of articulation. Studies of repetition suppression show that repeating a stimulus results in increased positivity of the ERP, which is claimed to be related to a memory trace encoding process (Garrido et al., 2009a,b). It is not clear in what way this notion should be related to underspecification. In the current study, we did not find significant differences between the two standards or between the two deviant stimuli, even though we did see an asymmetry.

The finding of an order effect highlights the importance of examining a range of factors that might influence the study outcome. Specifically, we observed that the amplitude of the MMN was smaller to deviant stimulus if it appeared as the standard in the first block, similar to the finding of Hestvik and Durvasula (2016) and Hestvik et al. (2020). This pattern suggests a lingering memory trace of the deviant when previously presented as the standard. Todd et al. (2014) proposed that the MMN reflects "precision weighted prediction" coding. The MMN responses were primarily decided by the post-synaptic sensitivity of superficial pyramidal cells that encode prediction error. In the flip-flop MMN paradigm, when the standard in the first block was reversed to be the deviant in the second block, the high probability in the first block suppressed the ERP to the deviant in the second block. The reduction of spiking rate to the standard stimuli has been called stimulus-specific adaption. This adaption effect is found at the single-neuron level in both the primary auditory cortex (Ulanovsky et al., 2003, 2004) and in the subcortical structures in animal studies (e.g., Anderson et al., 2009). In a human study, Costa-Faidella et al. (2011) showed that the ERP to a standard after 36 repetitions is more suppressed than that after 24 repetitions. At the same time, there is a concurrent increment in the negativity of the ERP to the deviant (e.g., more negative ERP to the deviant after 36 repetitions than after 24 repetitions). As a result, the MMN can be significantly affected by both short- and long-term stimulus history. What is relevant to the current study is that, if the repetition effect and switch effect are not sensitive to stimulus features, then we would not expect a directional asymmetry.

In addition, McGee et al. (2001) found that MMN amplitudes decline after 10–15 min due to habituation. Our experiment lasted 14 min in each block. It is possible that the habituation effect impacted the MMN amplitudes more generally across the experiment. Irrespective of whether our finding was a primacy or habituation effect, consideration of these order effects is important in this type of design. Future studies are needed better understand whether the long-term repetition factors observed for non-speech stimuli show the same effect for speech. Considering that speech is highly overlearned (Strange, 2011), there is no reason to assume that it will show the same adaptation timecourse as non-speech.

Limitations

Our study was not designed to directly discriminate among the several competing theories/frameworks discussed above. Our findings provided support for the underspecification, NRV and prototype models. Our study, however, did not test whether listeners exhibited perceptual (behavioral) asymmetries, in discrimination, identification or prototype goodness judgments. Our previous study that tested behavior suggested that /ɪ/ might be closer to the English prototypes than /ɛ/) (Hisagi et al., 2015a), but to verify this claim it would be necessary to obtain category goodness judgments. It also would have been useful to test a language group for whom these speech sounds would be perceived differently. For example, Hisagi et al. (2015a) found that Spanish listeners were more consistent in labeling /ɛ/ than /ɪ/ in the identification task, and thus, a reverse asymmetry to that of the English participants would be predicted.

CONCLUSION

This study provided additional evidence that asymmetric patterns in speech processing, related to which stimulus can be considered the standard (or referent), are robust. Specifically, the asymmetry observed for the /ɪ/ vs. /ɛ/ contrast is consistent with other studies. Our findings were consistent with a model of underspecification; however, they were also consistent with other explanations, such as the NRV model or a prototype model. There was little evidence indicating that a purely psychophysical explanation could support the findings. In addition, the finding of an order effect revealed that non-linguistic factors can contribute to asymmetries. Future research needs to examine whether these asymmetries are present early in life and to track these patterns developmentally and in relation to language experience. In addition, to fully test these various models, future studies need to examine a wider range of languages, which have different inventories, such as those with less dense acoustic space vowel inventories (e.g., Mandarin and Japanese).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) of St. John's University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VS designed the study and created the stimuli. YY helped design and implement the experiment, and collected and analyzed the data supported by research assistants. Both authors wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank M. DiCosta, J. Gonzalez, B. Perdomo, A. Skliras, and S. Sorensen for their help with data collection, and E. Adodo for preliminary data processing.

REFERENCES

Aaltonen, O., Eerola, O., Hellström, Å, Uusipaikka, E., and Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *J. Acoust. Soc. Am.* 101, 1090–1105. doi: 10.1121/1.418031

Aaltonen, O., Niemi, P., Nyrke, T., and Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biol. Psychol.* 24, 197–207. doi: 10.1016/0301-0511(87)90002-0

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear.* 16, 38–51. doi: 10.1097/00003446-199502000-00004

Anderson, L. A., Christianson, G. B., and Linden, J. F. (2009). Stimulus-specific adaptation occurs in the auditory thalamus. *J. Neurosci.* 29, 7359–7363. doi: 10.1523/JNEUROSCI.0793-09.2009

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104, 591–630. doi: 10.1016/j.cognition.2006.05.015

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., and Vanlancker, D. (1994). Influence of language structure on brain-behavior development. *Brain Lang.* 46, 607–619. doi: 10.1006/brln.1994.1033

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Lang. Var. Change* 14, 261–290. doi: 10.1017/S0954394502143018

Chládková, K., Escudero, P., and Lipski, S. C. (2015). When "AA" is long but "A" is not short: speakers who distinguish short and long vowels in production do not necessarily encode a short–long contrast in their phonological lexicon. *Front. Psychol.* 6:438. doi:10.3389/fpsyg.2015.00438

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper & Row.

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

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862

Costa-Faidella, J., Grimm, S., Slabu, L., Díaz-Santaella, F., and Escera, C. (2011). Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials: adaptation in the human auditory system. *Psychophysiology* 48, 774–783. doi: 10.1111/j.1469-8986.2010.01144.x

Cowan, N., and Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *J. Acoust. Soc. Am.* 79, 500–507. doi: 10.1121/1.393537

Cowan, N., Winkler, I., Teder, W., and Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 909–921. doi: 10.1037/0278-7393.19.4.909

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

Datta, H., Hestvik, A., Vidal, N., Tessel, C., Hisagi, M., Wróbleski, M., et al. (2020). Automaticity of speech processing in early bilingual adults and children. *Biling. Lang. Cogn.* 23, 429–445. doi: 10.1017/S1366728919000099

Datta, H., Shafer, V. L., Morr, M. L., Kurtzberg, D., and Schwartz, R. G. (2010). Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development. *J. Speech Lang. Hear. Res. JSLHR* 53, 757–777. doi: 10.1044/1092-4388(2009/08-0123)

Dien, J. (2010). The ERP PCA toolkit: an open source program for advanced statistical analysis of event-related potential data. *J. Neurosci. Methods* 187, 138–145. doi: 10.1016/j.jneumeth.2009.12.009

Dien, J. (2012). Applying principal components analysis to event-related potentials: a tutorial. *Dev. Neuropsychol.* 37, 497–517. doi: 10.1080/87565641.2012.697503

Dien, J., and Frishkoff, G. A. (2005). "Principal components analysis of event-related potential datasets," in *Event-Related Potentials: A Methods Handbook*, ed. T. C. Hardy (Cambridge, MA: MIT Press), 189–208.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., and Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *J. Neurosci.* 15, 2748–2755. doi: 10.1523/JNEUROSCI.15-04-02748.1995

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Frye, R. E., Fisher, J. M., Coty, A., Zarella, M., Liederman, J., and Halgren, E. (2007). Linear coding of voice onset time. *J. Cogn. Neurosci.* 19, 1476–1487. doi: 10.1162/jocn.2007.19.9.1476

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Friston, K. J. (2009a). Repetition suppression and plasticity in the human brain. *Neuroimage* 48, 269–279. doi: 10.1016/j.neuroimage.2009.06.034

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009b). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Handy, T. C. (2005). *Event-Related Potentials: A Methods Handbook*. Cambridge, MA: MIT Press.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *J. Linguist.* 42, 25–70. doi: 10.1017/S0022226705003683

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

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

Hisagi, M., Garrido-Nag, K., Datta, H., and Shafer, V. L. (2015a). ERP indices of vowel processing in Spanish–English bilinguals. *Biling. Lang. Cogn.* 18, 271–289.

Hisagi, M., Shafer, V. L., Strange, W., and Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Res.* 1360, 89–105. doi: 10.1016/j.brainres.2010.08.092

Hisagi, M., Shafer, V. L., Strange, W., and Sussman, E. S. (2015b). Neural measures of a Japanese consonant length discrimination by Japanese and American English listeners: effects of attention. *Brain Res.* 1626, 218–231. doi: 10.1016/j.brainres.2015.06.001

Højlund, A., Gebauer, L., McGregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. *Lang. Cogn. Neurosci.* 34, 545–560. doi: 10.1080/23273798.2019.1572204

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447

Iverson, P., and Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* 97, 553–562. doi: 10.1121/1.412280

Jacobsen, T., and Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* 38, 723–727. doi: 10.1111/1469-8986.3840723

Jacobsen, T., and Schröger E. (2003). Measuring duration mismatch negativity. *Clin Neurophysiol.* 114, 1133–1143. doi:10.1016/s1388-2457(03)00043-9

Kaukoranta, E., Sams, M., Hari, R., Hämäläinen, M., and Näätänen, R. (1989). Reactions of human auditory cortex to a change in tone duration. *Hear. Res.* 41, 15–21. doi: 10.1016/0378-5955(89)90174-3

Kazanina, N., Bowers, J. S., and Idsardi, W. (2018). Phonemes: lexical access and beyond. *Psychon. Bull. Rev.* 25, 560–585. doi: 10.3758/s13423-017-1362-0

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonol. Yearb.* 2, 85–138. doi: 10.1017/S0952675700000397

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schroger, E., and 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. *Int. J. Psychophysiol.* 67, 131–143. doi: 10.1016/j.ijpsycho.2007.10.012

Kraus, N., McGee, T., Sharma, A., Carrell, T., and Nicol, T. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear Hear.* 13, 158–164. doi: 10.1097/00003446-199206000-00004

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364

Kujala, T., Tervaniemi, M., and Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol.* 74, 1–19. doi:10.1016/j.biopsycho.2006.06.001

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., and Galán, L. (2010). False discovery rate and permutation test: an evaluation in ERP data analysis. *Stat. Med.* 29, 63–74. doi: 10.1002/sim.3784

Lahiri, A., and Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245–294. doi: 10.1016/0010-0277(91)90008-r

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Laboratory Phonology VII*, eds C. Gussenhoven, and N. Warner (Berlin: Mouton de Gruyter), 637–677.

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phon.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Luck, S. J., and Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* 54, 146–157. doi: 10.1111/psyp.12639

Maddieson. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511753459

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., and Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear Hear.* 16, 68–89. doi: 10.1097/00003446-199502000-00006

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

McGee, T., King, C., Tremblay, K., Nicol, T., Cunningham, J., and Kraus, N. (2001). Long−term habituation of the speech−elicited mismatch negativity. *Psychophysiology* 38, 653–658. doi: 10.1111/1469-8986.3840653

McGee, T., Kraus, N., and Nicol, T. (1997). Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects. *Electroencephalogr. Clin. Neurophysiol.* 104, 359–368. doi: 10.1016/s0168-5597(97)00024-5

Morr, M. L. (2002). *Mismatch Negativity Elicited to Vowels in Children and Adults and Its Association to Behavioral Performance*. Doctoral dissertation, City University of New York, New York, NY.

Möttönen, R., Dutton, R., and Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cereb. Cortex* 23, 1190–1197. doi: 10.1093/cercor/bhs110

Näätänen, R., Kujala, T., and Light, G. (2019). *The Mismatch Negativity (MMN): A Window to the Brain*, 1st Edn. Oxford: Oxford University Press.

Näätänen, R., Kujala, T., and Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses: auditory processing that leads to conscious perception. *Psychophysiology* 48, 4–22. doi: 10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nyman, G., Alho, K., Laurinen, P., Paavilainen, P., Radil, T., Reinikainen, K., et al. (1990). Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. *Electroencephalogr. Clin. Neurophysiol.* 77, 436–444. doi: 10.1016/0168-5597(90)90004-w

Paradis, C., and Prunet, J.-F. (1991). *The Special Status of Coronals: Internal and External Evidence*. San Diego, CA: Academic Press.

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

Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., and Yellin, E. (1995). "Brain mechanisms of speech perception: a preliminary report," in *MIT Working Papers in Linguistics*, Vol. 26, eds C. T. Schutze, J. B. Ganger, and K. Broihier (Cambridge, MA: MIT Press), 125–163.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *J. Phon.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, M., et al. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalogr. Clin. Neurophysiol.* 83, 306–321. doi: 10.1016/0013-4694(92)90090-5

Sams, M., Aulanko, R., Aaltonen, O., and Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *J. Cogn. Neurosci.* 2, 344–357. doi: 10.1162/jocn.1990.2.4.344

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *J. Mem. Lang.* 111:104070. doi: 10.1016/j.jml.2019.104070

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., and Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PLoS One* 7:e40953. doi: 10.1371/journal.pone.0040953

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

Scharinger, M., and Samuels, B. (2017). "Are there brain bases for phonological markedness?," in *Beyond Markedness in Formal Phonology*, ed. B. Samuels (Amsterdam: John Benjamins Publishing Company).

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., and Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *J. Cogn. Neurosci.* 17, 1168–1180. doi: 10.1162/0898929054475217

Shafer, V. L., Schwartz, R. G., and Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cogn. Brain Res.* 18, 242–254. doi: 10.1016/j.cogbrainres.2003.10.007

Shafer, V. L., Yu, Y. H., and Datta, H. (2010). Maturation of speech discrimination in 4- to 7-yr-old children as indexed by event-related potential mismatch responses. *Ear Hear.* 31, 735–745. doi: 10.1097/AUD.0b013e3181e5d1a7

Shafer, V. L., Yu, Y. H., and Datta, H. (2011). The development of English vowel perception in monolingual and bilingual infants: neurophysiological correlates. *J. Phon.* 39, 527–545. doi: 10.1016/j.wocn.2010.11.010

Sharma, A., and Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *J. Acoust. Soc. Am.* 104, 511–517. doi: 10.1121/1.423252

Sharma, M., Purdy, S. C., Newall, P., Wheldall, K., Beaman, R., and Dillon, H. (2004). Effects of identification technique, extraction method, and stimulus type on mismatch negativity in adults and children. *J. Am. Acad. Audiol.* 15, 616–632. doi: 10.3766/jaaa.15.9.3

Shiramatsu, T. I., and Takahashi, H. (2018). Mismatch negativity in rat auditory cortex represents the empirical salience of sounds. *Front. Neurosci.* 12:924. doi: 10.3389/fnins.2018.00924

Stemberger, J. P. (1991a). Apparent anti-frequency effects in language production: the addition bias and phonological underspecification. *J. Mem. Lang.* 30, 161–185. doi: 10.1016/0749-596X(91)90002-2

Stemberger, J. P. (1991b). Radical underspecification in language production. *Phonology* 8, 73–112.

Stemberger, J. P. (1992). A performance constraint on compensatory lengthening in child phonology. *Lang. Speech* 35, 207–218. doi: 10.1177/002383099203500216

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phon.* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001

Takegata, R., Tervaniemi, M., Alku, P., Ylinen, S., and Näätänen, R. (2008). Parameter-specific modulation of the mismatch negativity to duration decrement and increment: evidence for asymmetric processes. *Clin. Neurophysiol.* 119, 1515–1523. doi: 10.1016/j.clinph.2008.03.025

Todd, J., Heathcote, A., Mullens, D., Whitson, L. R., Provost, A., and Winkler, I. (2014). What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topogr.* 27, 578–589. doi: 10.1007/s10548-013-0344-4

Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* 21, 1532–1540. doi: 10.1177/0956797610384142

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J. Neurosci.* 24, 10440–10453. doi: 10.1523/JNEUROSCI.1905-04.2004

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

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Front. Neurosci.* 11:95. doi: 10.3389/fnins.2017.00095

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2018). The duration of auditory sensory memory for vowel processing: neurophysiological and behavioral measures. *Front. Psychol.* 9:335. doi: 10.3389/fpsyg.2018.00335

Yu, Y. H., Tessel, C., Han, H., Campanelli, L., Vidal, N., Gerometta, J., et al. (2019). Neural indices of vowel discrimination in monolingual and bilingual infants and children. *Ear Hear.* 40, 1376–1390. doi:10.1097/AUD.0000000000000726

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 22 April 2021
doi: 10.3389/fnhum.2021.617318

Asymmetric Influence of Vocalic Context on Mandarin Sibilants: Evidence From ERP Studies

Yaxuan Meng[1]*, Sandra Kotzor[1,2], Chenzi Xu[1], Hilary S. Z. Wynne[1] and Aditi Lahiri[1]

[1] Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, United Kingdom, [2] School of Education, Oxford Brookes University, Oxford, United Kingdom

In the present study, we examine the interactive effect of vowels on Mandarin fricative sibilants using a passive oddball paradigm to determine whether the HEIGHT features of vowels can spread on the surface and influence preceding consonants with unspecified features. The stimuli are two pairs of Mandarin words ([sa] ∼ [ʂa] and [su] ∼ [ʂu]) contrasting in vowel HEIGHT ([LOW] vs. [HIGH]). Each word in the same pair was presented both as standard and deviant, resulting in four conditions (/standard/$_{\text{[deviant]}}$: /sa/$_{\text{[ʂa]}}$ ∼ /ʂa/$_{\text{[sa]}}$ and /su/$_{\text{[ʂu]}}$ ∼ /ʂu/$_{\text{[su]}}$). In line with the Featurally Underspecified Lexicon (FUL) model, asymmetric patterns of processing were found in the [su] ∼ [ʂu] word pair where both the MMN (mismatch negativity) and LDN (late discriminative negativity) components were more negative in /su/$_{\text{[ʂu]}}$ (mismatch) than in /ʂu/$_{\text{[su]}}$ (no mismatch), suggesting the spreading of the feature [HIGH] from the vowel [u] to [ʂ] on the surface. In the [sa] ∼ [ʂa] pair, however, symmetric negativities (for both MMN and LDN) were observed as there is no conflict between the surface feature [LOW] from [a] to [ʂ] and the underlying specified feature [LOW] of [s]. These results confirm that not all features are fully specified in the mental lexicon: features of vowels can spread on the surface and influence surrounding unspecified segments.

Keywords: LDN, MMN, tongue height, vowel, mandarin sibilant

Edited by:
Yang Zhang,
University of Minnesota Health Twin Cities, United States

Reviewed by:
Stephen Politzer-Ahles,
Hong Kong Polytechnic University, Hong Kong
Yan H. Yu,
St. John's University, United States

***Correspondence:**
Yaxuan Meng
yaxuan.meng@ling-phil.ox.ac.uk

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *21 October 2020*
Accepted: *23 February 2021*
Published: *22 April 2021*

Citation:
Meng Y, Kotzor S, Xu C, Wynne HSZ and Lahiri A (2021) Asymmetric Influence of Vocalic Context on Mandarin Sibilants: Evidence From ERP Studies. Front. Hum. Neurosci. 15:617318. doi: 10.3389/fnhum.2021.617318

INTRODUCTION

To comprehend spoken language, listeners need to decode the incoming speech stream and segment it into units which map onto the phonological representations of words. However, the incoming acoustic cues for consonants and vowels can vary quite substantially due to factors such as context, speaking rate, and speaker characteristics. Nevertheless, mature listeners rarely experience any difficulty in recognizing spoken words and inferring the intended message (Marslen-Wilson, 1984; Norris et al., 1995; Lahiri and Reetz, 2002, 2010).

The speech signal varies in different contexts where the realization of a particular sound can differ within and across individual words (cf. Holt and Kluender, 2000). Furthermore, contextual modifications (contiguous sounds affecting each other such as vowels affecting consonants, consonants affecting other consonants, etc.) can alter the pronunciation of a sound quite drastically. A familiar example is that of place assimilation where the underlined medial sequences [ng] in *greengage* or [np] *gunpoint* are habitually articulated as [ŋg] and [mp] respectively. Here, the place of articulation of the [CORONAL] nasal [n] is affected by that of the following consonant, transforming it into a [DORSAL] [ŋ] or [LABIAL] [m] nasal. Vowels can also affect consonants as is seen in word pairs such as *face∼ facial* or *commerce∼ commercial,* where the final sound [s] of the first word of each pair becomes [ʃ] in the context of the vowel [i] when suffixed with *-ial* [iəl]. Here the [i] is no longer pronounced; however, in other contexts, such as in *dictator∼ dictatorial*, the vowel [i] does not change. In this paper, we investigate brain responses to variability in sound sequences where vowels alter neighboring consonants.

The effect of one sound on another tends not to be symmetric. For example, in the example given above (*greengage* and *gunpoint*), the assimilation of the place of articulation is asymmetric. Although [CORONAL] [n] can change to [m] and [ŋ], the reverse is usually not the case: a [DORSAL] nasal, as in the sequence [ad] *kingdom* does not become *[nd] nor does the [LABIAL] nasal in *sometime* change to *[nt]. Thus, [CORONAL] consonants such as [n] can assimilate easily to the place of articulation of the following [LABIAL] (e.g., [p], [b]) or [DORSAL] consonants (e.g., [k], [g]) but not vice versa (cf. Cornell et al., 2013). One approach to capture this asymmetry is to assume that not all features or properties of consonants and vowels are fully specified in the lexicon (cf. the Featurally Underspecified Lexicon (FUL) model; Lahiri and Reetz, 2010; Scharinger et al., 2012; de Jonge and Boersma, 2015; Schluter et al., 2016; Højlund et al., 2019; Kotzor et al., 2020). In this model, consonants and vowels are defined by PLACE OF ARTICULATION which include ARTICULATOR features such as [CORONAL], [DORSAL], [LABIAL], and HEIGHT features [HIGH] and [LOW]. Of these, [CORONAL] is assumed to be universally underspecified (see **Figure 1**).

Since each word has a unique phonological representation, the features extracted from the acoustic signal are used to map speech onto underlying representations. Listeners process the variable speech signal and parse it into features which are then directly mapped onto the lexicon (Lahiri and Reetz, 2010; Kotzor et al., 2020). This mapping from the features in the signal to the lexicon is based on a ternary logic*: match, mismatch*, *and no-mismatch*. The first two options are transparent: *match* equates to the feature from the signal matching the lexicon completely while *mismatch* occurs when there is a conflict. Thus, the feature [CORONAL] from the acoustic signal of [n], for instance, will *mismatch* with the lexically represented feature [LABIAL] of [m]. The *no-mismatch* condition suggests a level of tolerance and is particularly important for underspecified features such as [CORONAL]. Consequently, [LABIAL] extracted from the signal of [m] will be in a no-mismatch relationship with [n] since its place feature [CORONAL] is not specified. Thus, during speech processing, all words in the lexicon with matching and no mismatching features are activated, but when mismatching features are encountered, words are deactivated.

There has been considerable evidence from both behavioral and neurophysiological studies for the underspecification of [CORONAL] place of articulation (Lahiri and Reetz, 2002, 2010; Eulitz and Lahiri, 2004; Cornell et al., 2011). For instance, the mismatch negativity (MMN) component has been used widely as a robust measure to examine [CORONAL] underspecification (Eulitz and Lahiri, 2004; Cornell et al., 2011). The MMN component, which usually peaks at 100–250 ms after the onset of a stimulus, signals the automatic or pre-attentive detection of an infrequent change in regular auditory stimulations (Näätänen et al., 2007). The MMN can be elicited by the deviant that violates the representation of repetitive standards before the occurrence of that deviant, suggesting that the sensory memory trace of preceding stimuli is compared against incoming sounds (Näätänen and Winkler, 1999; Horváth et al., 2008). The amplitude or latency of the MMN component depends on the magnitude of the stimulus deviation, with larger deviance resulting in an increase in amplitude and shorter latencies (Näätänen et al., 2007). In MMN studies examining coronal underspecification (e.g., Eulitz and Lahiri, 2004; Roberts et al., 2014), [CORONAL] deviants elicit larger MMN amplitudes in the context of, for instance, [LABIAL] standards as this creates a mismatch while deviants which are fully specified (e.g., [LABIAL]) and occur in the context of a [CORONAL] standard result in an attenuated MMN amplitude (no-mismatch).

Similar arguments have been made for ARTICULATOR features for vowels. Asymmetric MMN contrasts also support the concept of underspecified representations for vowels. Cornell et al. (2011) compared the phonological representations of vowels [ø] and [o] in the mental lexicon by means of MMN. The vowel [ø] is [CORONAL] and thus underspecified for its place of articulation while the vowel [o] is specified as [DORSAL]. In the context of a series of standard [ø], a fully specified phonetic [o] is a less different stimulus (i.e., no mismatch) than a deviant [ø] in the context of a series of fully specified standard [o] (i.e., a mismatch). Asymmetry occurs such that a deviant [o] in a standard [ø] context ($[\text{o}]_{/\text{ø}/}$) elicits a smaller MMN than a deviant [ø] in an [o] context ($[\text{ø}]_{/\text{o}/}$). Here, the representation activated by the repeated processing of standard stimuli is from a long-term memory trace, and associated to the underlying representation in the mental lexicon. In contrast, the sound percept elicited by the deviant stimulus corresponds to the surface representation, which is formed by the phonological features extracted from the acoustic signal (Eulitz and Lahiri, 2004; Cornell et al., 2011). The change detection response reflects the contrast between the underlying and surface representations.

Comparing both ARTICULATOR and TONGUE HEIGHT features, Kotzor et al. (2020) examined asymmetric ARTICULATOR features as well as symmetric HEIGHT features in vowels in words and non-words (**Table 1**). They contrasted the ARTICULATOR asymmetry in the vowels [ɛ] [CORONAL] and [ɔ] [DORSAL] in the verbs *get* [gɛt] ∼ *got* [gɔt] and the pseudowords **gef* ∼ **gof.* In the same study, conflicting fully specified, and hence symmetric, HEIGHT features which mismatch in both directions were also compared while keeping the ARTICULATOR feature [CORONAL] constant: *sit* [sɪt] ∼ *sat* [sæt] and **sif* ∼ **saf,* where [ɪ] [HIGH] and [æ] [LOW] conflict and hence mismatch. While the place features [CORONAL] and [DORSAL] were predicted to elicit asymmetric MMNs in both words and pseudowords, the height features, which are both fully specified, mismatch and should thus elicit high MMNs of comparable amplitude regardless of which vowel occurs as standard or deviant. The results confirmed their hypotheses: due to [CORONAL] underspecification, [CORONAL] and [DORSAL]

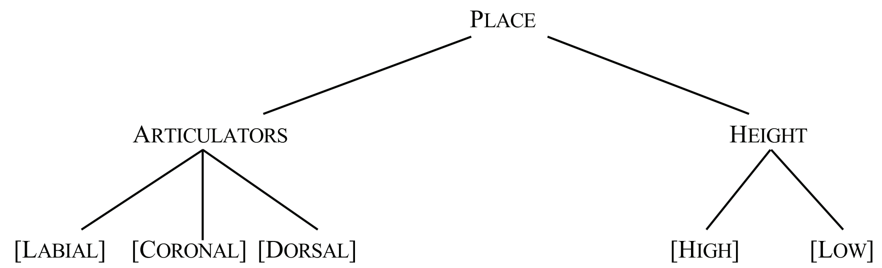


FIGURE 1 | Feature organization of PLACE OF ARTICULATION in FUL.

place features elicited asymmetric MMNs, while conflicting height features [HIGH] and [LOW] mismatched and the MMNs did not differ.

So far, we have discussed pairs of individual features on which the influence from surrounding segments have been kept constant. However, in normal speech, contiguous consonants, and vowels always lead to coarticulation or spreading of features, some more than others. Thus, in a VOWEL + NASAL sequence such as the English word *an,* the [NASAL] feature of [n] can spread to the vowel [ae] leading to [æ̃n]. This assimilation also has processing consequences (Lahiri and Marslen-Wilson, 1991). In English, this is purely allophonic, which means that the nasalization is entirely predictable and there is no real phonemic contrast between oral and nasal vowels; e.g., *cad* [kæ̃aed] vs. *can* [kʰæ̃n]. Nevertheless, on perceiving nasality on the vowel, English listeners can anticipate a following nasal consonant. e.g., [n] can be anticipated after hearing the sequence [kʰæ̃] in *can*. In this paper, we address the consequence of similarities and differences of more complex CV units where the feature ARTICULATOR is kept constant, but HEIGHT spreads from vowels to consonants. Using an MMN paradigm, we examine the contrastive [CORONAL, STRIDENT] consonants [s] and [ʂ] which differ in HEIGHT in the context of both [HIGH] and [LOW] vowels. The relevant vowels are [u] and [a] which also differ in HEIGHT: [sa]∼[ʂa] and [su]∼[ʂu] (as shown in **Table 2**). At first glance, the pairs appear to be straightforward; however, the underlying phonological representation of the features for these pairs depends not only on the phonemes but on the general phonological inventory of Mandarin.

The phonological feature specifications within a language are determined by the number of contrastive segments. In Mandarin, there are two sets of [CORONAL] obstruents: *dental* [t, tʰ, ʦ, ʦʰ, s] and *retroflex* [tʂ, tʂʰ, ʂ]. There are fewer retroflex consonants than dentals in Mandarin: Duanmu (2007) states that the retroflex series is a "major characteristic of Standard Chinese (SC) speakers from Beijing" (p. 24) and that speakers of other Chinese dialects replace the retroflex with the dental; e.g., there would be no distinctions between [sa] "sprinkle" and [ʂa] "stupid." To distinguish between the two types of [CORONAL] obstruents, our feature system uses the HEIGHT features [HIGH] and [LOW] (cf. Lahiri, 2018). Based on their acoustic characteristics, the retroflex consonants would be characterized as [HIGH] and the dentals as [LOW]: dental sibilants have more energy in the higher frequencies compared to retroflexes and palatals (Stevens and Blumstein, 1975; Lahiri and Reetz, 1999; Lahiri and Kennard, 2019; Kennard and Lahiri, 2020). However, Mandarin only has a two-way contrast in the voiceless sibilant fricatives [s] and [ʂ]; thus, it is only necessary to lexically specify one of these phonemes; the other remains *unspecified*. Since there are more dental consonants than retroflexes and since the dentals are less likely to vary in comparison to the retroflexes, the dentals would be more likely to be specified for HEIGHT in the lexicon.

Further evidence of the specification of the HEIGHT feature is provided by the co-occurrence restriction that certain adjacent identical elements are prohibited in consonant-glide sequences (Yip, 1988; Wiese, 1997; Duanmu, 2007). As for vowels, descriptively Mandarin allows five basic vowels where [i u y] are high vowels, [ə] is a mid vowel and [a] is usually characterized as a low vowel. In terms of features, the mid vowel is underspecified while the high and the low vowels are specified. Mandarin syllables can only have single consonants as onsets and codas and no clusters are permitted. Thus, since all initial consonants followed by high vowels /i u y/ attain a secondary articulation described as glide spreading, /CuC/ becomes [CʷuC] where the [Cᶢ] holds a single position in the onset. As we can see, [s] can occur with both high and non-high vowels, such as /suuŋ4/ [sʷuŋ] 送 "send" or /suən1/ [sʷən] 孙 "grandchild," where the glide formation rule turns the vowel [u] into a glide leading to a secondary articulation [sʷ]. Since [s] is specified as [LOW] and the glides are high, the secondary articulation is allowed. Had/ ʂ/ been specified for [HIGH], the sequence /ʂuu/ > [ʂʷu] "to lose" would not have been permitted because of identical height features (**Table 2**). Crucially, the feature [HIGH] is not specified in the language for any of the consonants, thus allowing them to take on the secondary articulations triggered by high vowels. We will examine the two sibilants [s] and [ʂ], which are differentiated in HEIGHT, in combination with two vowels also differing in HEIGHT: [u] and [a].

As we mentioned above, not only do ARTICULATOR features such as [LABIAL] or [DORSAL] spread leading to assimilation in words such as *greengage*, but TONGUE HEIGHT features such as [LOW] or [HIGH] can also spread to preceding unspecified segments (Kunisaki and Fujisaki, 1977; Mann and Repp, 1980; Lahiri and Reetz, 2002). In a study by Kunisaki and Fujisaki (1977), a continuum of synthetic fricative sounds varying from [ʃ] to [s] was combined with different vowels. The category boundary was found to shift to [s] when followed by vowels [u] or [o], while it shifted toward [ʃ] in the context of [a]

TABLE 1 | The mismatching and matching relationships in the study by Kotzor et al. (2020).

	Articulators		Height	
Features in the lexical Representation	[CORONAL] underspecified	[DORSAL]	[HIGH]	[LOW]
Matching relationship	↑ *nomismatch*	⇞ *mismatch*	⇞ *mismatch*	⇞ *mismatch*
Features from the signal	[DORSAL]	[CORONAL]	[LOW]	[HIGH]

The symbol ↑ represents nomismatch and ⇞ represents mismatch.

TABLE 2 | Critical features for the Mandarin CV sequences.

	Place							
	Articulator				Height			
	[s]	[ʂ]	[a]	[u]	[s]	[ʂ]	[a]	[u]
Features: Lexicon	COR	COR	[DOR]	[DOR]	[LOW]	—	[LOW]	[HIGH]
Features: Signal	[COR]	[COR]	[DOR]	[DOR]	[LOW]	—	[LOW]	[HIGH]

The feature [CORONAL] in the representation is underspecified and is shown in gray.

or [e], suggesting an effect of vocalic context on fricative consonant perception. Thus, it appears that simple coarticulation in contiguous segments can influence perception. Similarly, Mann and Repp (1980) also found that listeners were more likely to perceive a synthetic fricative consonant from a [ʃ]~[s] continuum as a [s] in the context of [u] compared to the context of [a]. The authors attributed the influence of vowel on consonant to an assimilatory change where the vowel rounding and consonant place of articulation coarticulated (Mann and Soli, 1991). If this is the case, symmetric MMNs between the phonological contrasts [sa] ~ [ʂa] and [su] ~ [ʂu] would be expected independent of the direction of presentation of the standard and deviant, as no feature is unspecified.

In contrast, asymmetric MMNs would be expected in the reversal of phonological contrasts if the influence of vocalic context is due to the spreading of features on the surface. Given that certain features are underspecified, the influence of vocalic context will be greater if the contiguous segment lacks a feature. The fricatives [s] and [ʂ] share the same ARTICULATOR feature [CORONAL] as well as the MANNER feature [STRIDENT], but differ with respect to their HEIGHT features (see **Table 3**). As only [LOW] is specified, the HEIGHT features of following vowels could spread to preceding [ʂ] but not [s]. In other words, the surface height feature of [ʂ] is determined by the following vowels. In their underlying representations, dental [s] is assumed to be [LOW] with [CORONAL] underspecified, while retroflex [ʂ] is assumed to be underspecified for ARTICULATOR and unspecified for the HEIGHT feature. Since [ʂ] lacks specification of TONGUE HEIGHT features, it can take on the HEIGHT features of the following vowel [u] and [a] (**Table 3** a, b). In contrast, [s] is specified for [LOW], and thus, phonologically, it is not affected by the features of the following vowel and retains its own HEIGHT feature (**Table 3** c, d). If this feature spreading account holds, then we would assume that, although the vowels are identical and both sibilant fricatives are [CORONAL], [sa] vs. [ʂa] and [su] vs. [ʂu] should elicit different activation patterns: specifically, we predict that [sa]~[ʂa] will lead to symmetric activation while [su]~[ʂu] will not.

Along with MMN, an additional negativity, the late discriminative negativity (LDN), was observed in our study. The LDN is a recently established component found in oddball paradigms and serves as an index of phonological discriminative abilities (Hill et al., 2004; Horváth et al., 2009; Jakoby et al., 2011; David et al., 2020). Similar to the MMN, the LDN is also an automatic response associated with higher cognitive processes and may represent the recruitment of additional cortical resources needed to extract the phonological differences between the standard and deviant stimulus and form phonological representations (Shestakova et al., 2003; Hill et al., 2004; Zachau et al., 2005; Barry et al., 2009). The LDN can be elicited by both speech and non-speech sounds, and its amplitude was found to be related to the difficulty in discriminating the stimuli (Korpilahti et al., 1995, 2001; Schulte-Körne et al., 1998). For example, Yu et al. (2017) compared the processing of Mandarin disyllabic non-words with different inter-stimulus intervals (ISIs) between Mandarin- and English-speaking groups. For both groups, robust MMNs to contrasts with either similar or contrastive lexical tones at shorter ISIs were observed. Compared to the English group, a larger LDN was only found for the Mandarin group when processing contrasts at longer ISIs, especially those with similar lexical tone. These results suggest that it is easier to discriminate the acoustic correlates of lexical tone at shorter ISIs. To discriminate words at longer ISIs, language-specific experience is necessary. Following the FUL model, Hestvik and Durvasula (2016) examined the underspecification-driven asymmetry in the processing of the English contrast between /d/, which is underspecified for [VOICE], and /t/, which is specified for the feature [SPREAD], using the oddball paradigm. The LDN component exhibits the same asymmetry as the MMN with a

TABLE 3 | The spreading of HEIGHT features from vowel to consonant.

	Height features							
	(a)		**(b)**		**(c)**		**(d)**	
UR	**/ʂ/**	**/u/**	**/ʂ/**	**/a/**	**/s/**	**/u/**	**/s/**	**/a/**
	TH	TH	TH	TH	TH	TH	TH	TH
	[ ]	[HIGH]	[ ]	[LOW]	[LOW]	[HIGH]	[LOW]	[LOW]
SR	[HIGH]	[HIGH]	[LOW]	[LOW]	[LOW]	[HIGH]	[LOW]	[LOW]

mismatch for /t/$_{[d]}$ but not for /d/$_{[t]}$. Based on these previous studies, we anticipate that, consistent with the activation patterns of MMN, [sa]∼[ʂa] will lead to symmetric LDNs for the subtle difference between standard and deviant, while [su]∼[ʂu] will not.

Methodology

The presented study examines the interactive effect of vowels on fricative sibilants to determine whether the TONGUE HEIGHT features of vowels can spread on the surface and influence unspecified preceding consonants. Coarticulation, which leads to feature spreading, would suggest symmetric MMNs between phonological contrasts, independent of the direction of presentation of the standard and deviant if the features are fully specified in both standards and deviants. In contrast, asymmetric MMNs would be expected between the two directions of presentation (i.e., standard vs. deviant) of phonological contrasts where the HEIGHT feature [HIGH] is unspecified in one of the stimuli.

Since Mandarin also has a tonal contrast, it was necessary to keep the tones consistent across the stimuli. Two monosyllabic word pairs with Tone 1, [sa]∼[ʂa] and [su]∼[ʂu], were used as the standard and deviant stimuli. Mandarin is a language where one syllable corresponds to one morpheme in most cases, with each syllable being comprised of an optional initial consonant, optional glide, a vowel, and an optional final consonant [*n* (n) *or ng* (ŋ)]. We already described the two voiceless sibilant fricatives in Standard Chinese (SC, or Mandarin), represented as the dental/alveolar [s] and retroflex [ʂ]. Here, the retroflex [ʂ] in Mandarin is different from the palatoalveolar [ʃ] in English in terms of the consonant position and air flow through the mouth. The palatoalveolar is pronounced with the air flow through the tongue blade and even a portion of the front part of the tongue. For the retroflex, the air flow is more limited to the tongue tip/blade region (Lin, 2007). Here, we follow Duanmu's (2007) position and treat the voiceless fricative sibilants in Mandarin as a two-way contrast: the dental/alveolar [s] and retroflex [ʂ]. Unlike the two-way contrast between fricative sibilants, the vowels in Mandarin are categorized into a three-way height distinction, including three high vowels [i y u], one mid vowel [ə], and one low vowel [a] (Duanmu, 2007). As mentioned above, both [HIGH] and [LOW] should be specified when there is a three-way height difference (Lahiri and Reetz, 2010). Therefore, the surface and underlying representations of [a] and [u] are consistent in that [a] is assumed as [LOW][DORSAL] and [u] is assumed as [HIGH][DORSAL], respectively (refer to **Table 2**).

We predict that: (1) for the [sa]∼[ʂa] word pair, no conflict will occur between /sa/$_{[\text{ʂa}]}$ and /ʂa/$_{[\text{sa}]}$ (/Standard/$_{[\text{Deviant}]}$). As discussed above, features spread on the surface with the HEIGHT feature of the vowel affecting the preceding unspecified sibilant. Therefore, the HEIGHT feature [LOW] of the vowel [a] spreads to [ʂ] when [ʂa] occurs as a deviant. As a result, the surface HEIGHT features of [ʂa] in the /sa/$_{[\text{ʂa}]}$ condition become [LOW] + [LOW] with the [CORONAL] PLACE feature being the only no-mismatching feature. For the /ʂa/$_{[\text{sa}]}$ condition, both the PLACE and HEIGHT features are in a no-mismatch relationship with the underlying representations and hence no-mismatching patterns are found. Therefore, symmetric MMNs and LDNs are predicted for these two conditions (as shown in **Figure 2**).

Our second prediction (2) is that, for the [su]∼[ʂu] word pair, a phonological conflict occurs between /su/$_{[\text{ʂu}]}$ and /ʂu/$_{[\text{su}]}$, causing the deviant [ʂu] in a standard [su] context to elicit a larger MMN and LDN than a deviant [su] in a [ʂu] context. For the /su/$_{[\text{ʂu}]}$ condition, the HEIGHT feature spreads from [u] to [ʂ], resulting in the surface HEIGHT features [HIGH] + [HIGH] of [ʂu]. The HEIGHT feature of [s], however, is specified as [LOW] resulting in the underlying HEIGHT features [LOW] + [HIGH] for [su]. Thus, a mismatch occurs in the /su/$_{[\text{ʂu}]}$ condition. Compared to /su/$_{[\text{ʂu}]}$, there is no surface feature spread in the /ʂu/$_{[\text{su}]}$ condition so that the HEIGHT and PLACE features of deviant [su] no mismatch the underlying representation of standard [ʂu]. Consequently, asymmetric MMNs and LDNs are predicted with a larger amplitude for /su/$_{[\text{ʂu}]}$ than /ʂu/$_{[\text{su}]}$.

If, on the other hand, we assume a phonemic representation with every feature fully specified in all sounds, all variants should mismatch to the same degree, as the spreading from [u] would not alter the specification of [LOW] in [ʂ]. In such a case, we would expect to see symmetric MMN and LDN responses for both pairs of words, regardless of the direction of presentation (i.e., which is the standard and which the deviant).

MATERIALS AND METHODS

Participants

Twenty-one students (11F/10M, mean age = 23.86 years), recruited at the University of Oxford, participated in the study. They were all native Mandarin speakers who lived in China until adulthood and were residing in Oxford at the time of

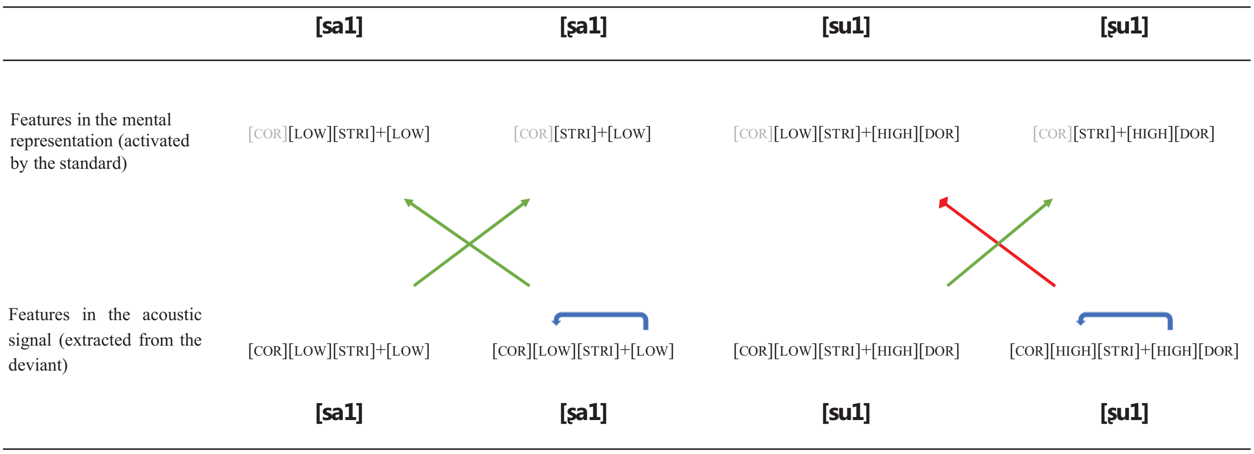


FIGURE 2 | Predictions made about the feature conflict in the four experimental conditions. The arrows illustrate the main statistical model. The blue arrow reflects the feature spread on the surface from vowel to consonant. The green arrows indicate combinations of standard and deviant stimuli in no-mismatch conditions and the red arrow represents the mismatch condition.

testing. All participants had normal or corrected-to-normal vision and self-reported as right-handed (a modified version of the Edinburgh Handedness Inventory was also used to assess handedness, Oldfield, 1971). No history of neurological disorders or hearing deficits was reported. The study was approved by the Central University Research Ethics Committee (CUREC) and written informed consent was acquired from subjects prior to the experiment. They were compensated for their participation.

Stimuli

Two pairs of Mandarin monosyllabic words that differ only in initial fricative consonants were used in the experiments ([sa] ∼ [ʂa]; [su] ∼ [ʂu]). Monosyllabic words are plentiful in Chinese and thus are polysemous. Each permissible syllable, with any one of the four lexical tones, could represent various meanings. As there are only two fricative sibilants in Mandarin, it is difficult to construct pseudowords with a combination of vowels differing in TONGUE HEIGHT. Thus, our four stimuli are all words, each of which predictably has several meanings. The most obvious meanings of the four syllables are as follow: [sa] 撒 "let go"∼ [ʂa] 沙 "sand"; [su] 苏 "place name"∼ [ʂu] 输 "to lose." According to the SUBTLEX-CH, the frequencies for these syllables are 3.46 ([sa]^1), 3.34 ([ʂa]^1), 3.17 ([su]^1), and 3.35 ([ʂu]^1) (Cai and Brysbaert, 2010).

As expected, the spectrogram of the same fricative varies depending on the following vowel (**Figure 3**). The coarticulation was maintained in order to preserve the naturalness of the stimuli. Each pair contained contrasting coronal fricatives [s] and [ʂ] embedded in respective vowel contexts: the [a] with feature [LOW], and the [u] with features [HIGH] and [DORSAL]. Since Mandarin has a lexical contrast in tones, it was important to control for this as well. All syllables were chosen to have lexical Tone 1, which is usually described as a high-level tone (Duanmu, 2007). Thus, the pitch is held at a constant level.

Multiple repetitions of four syllables were recorded by a female native speaker of Mandarin in a sound-attenuated recording room using a professional quality USB microphone (Røde NT-USB) at a sampling rate of 44.1 kHz. From these syllables, we generated four naturally sounding stimuli recordings. A representative utterance of each syllable with similar duration was selected. The recordings were extracted and segmented using the speech analysis program Praat (Boersma and Weenink, 2018). The [a] and [u] vowels in [sa] and [su] were cross-spliced to the corresponding [ʂ] consonant in each pair such that the acoustic differences between the stimuli in each pair were minimized to the contrasting consonants. As shown in **Figure 3**, the vowel portions in each pair were identical.

Across pairs, stimuli were also controlled for duration (**Figure 3**). In the recordings, the vowel [u] was slightly longer than [a], so some trailing pulses at the end of [u] were removed. Likewise, some initial pulses of noise were removed in [su] and [ʂu] because their frication duration was slightly longer than those of [sa] and [ʂa]. Such manipulations avoided the parts of formant transitions in order to minimize the distortions of F0 and spectral features. Therefore, all initial consonants were approximately 182 ms, the vowels 328 ms, and the overall duration of a syllable was about 510 ms (as shown in **Figure 3**). The intensities of all stimuli were equalized in Praat.

Experimental Procedure

Two pairs of words with Tone 1 were presented to participants during the experiment. Each word pair was presented in two conditions; one with a [s] consonant as the deviant and a [ʂ] consonant as the standard, and one with the direction of presentation reversed (see **Table 4**). The word pairs will be described respectively as /sa/$_{[\text{ʂa}]}$ ∼ /ʂa/$_{[\text{sa}]}$ and /su/$_{[\text{ʂu}]}$ ∼ /ʂu/$_{[\text{su}]}$ in the paragraphs below (/standard/$_{[\text{deviant}]}$).

As a result of this reversed design, four oddball blocks were presented to each participant with the sequence of blocks counterbalanced among the participants. Within each block, the deviant occurred pseudo-randomly among the standards with a probability of 15%. Any two adjacent deviants were separated

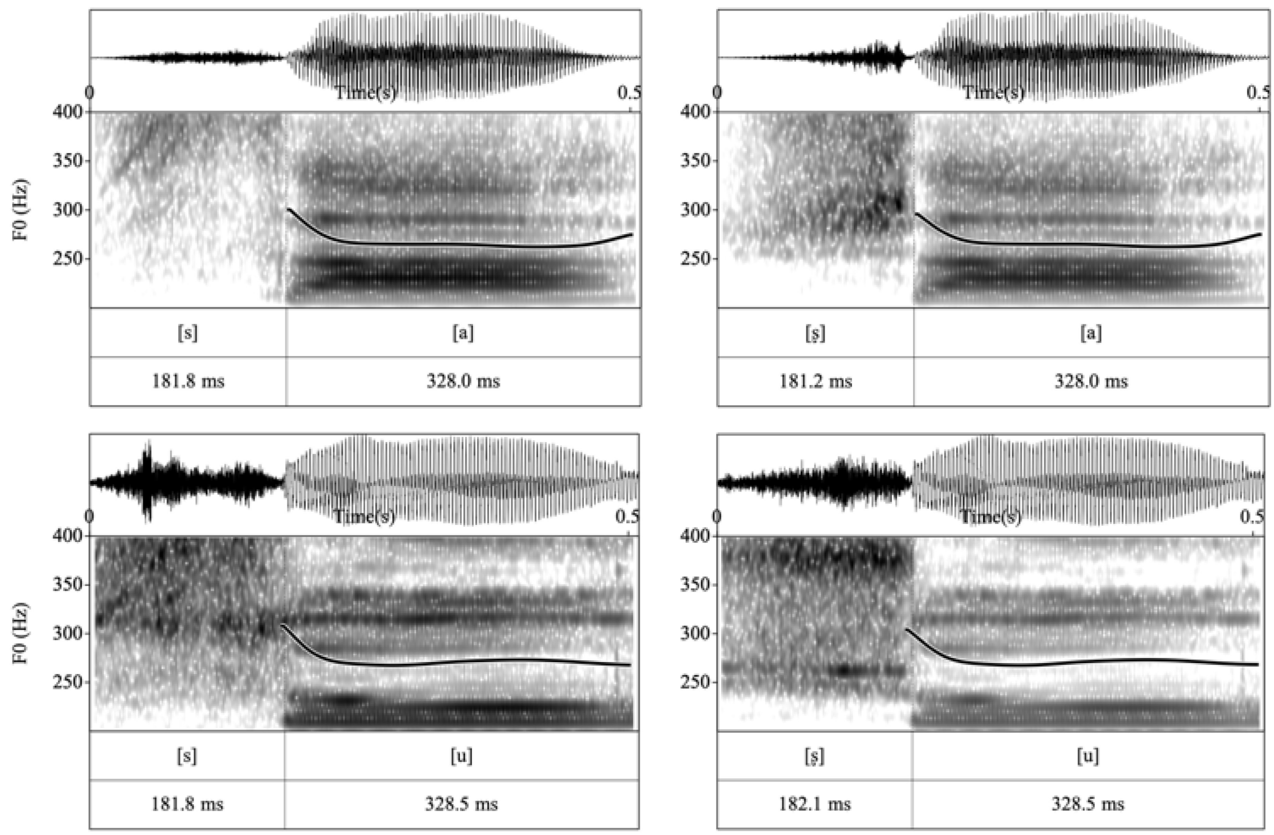


FIGURE 3 | Oscillograms (above), spectrograms (below, 0–7,000 Hz), and F0 tracks of the four stimuli. All the stimuli are Tone 1 syllables.

TABLE 4 | Task design in MMN tasks.

	Experiment 1		Experiment 2	
	Standard	**Deviant**	**Standard**	**Deviant**
Block 1	[sa]	[ʂa]	[su]	[ʂu]
Block 2	[ʂa]	[sa]	[ʂu]	[su]

by at least two standards. A total of 610 stimuli, with ten continuous standard stimuli occurring at the beginning, were presented in each block. To eliminate the influence of a rhythmic pattern established by temporal characteristics of the acoustic stimuli, the ISI between standard and deviant varied randomly between 350 and 650 ms.

EEG Recordings

EEG recordings were made using a Biosemi ActiveTwo amplifier with 64 sintered Ag/AgCl pin electrodes placed in a 10–20 montage, online referenced to the mastoids. EOG activity was measured using four facial electrodes (IO1, IO2, LO1, and LO2). All electrode offsets (in an active-electrode system this is comparable to impedance) were kept below 30 mV and signals were sampled at 2,048 Hz. The audio stimuli were presented through headphones and participants watched a self-selected silent documentary during the experiment. All subjects participated in all four blocks and the order of the four blocks was counterbalanced across subjects. The total duration of the experiment was about 90 min and subjects had a short break between blocks.

Data Analysis

EEG data were analyzed offline using EEGLAB 14.1.2b. All continuous data were digitally-filtered offline in 0.3–30 Hz range using a finite impulse response filter (FIR filter). Bad channels and artifacts were detected and removed automatically by the artifact subspace reconstruction (ASR) method as implemented in the Clean Raw Data plug-in. EEG data were re-referenced to the linked mastoids for all analyses except for mastoid amplitudes. Using an independent components analysis (ICA, Delorme and Makeig, 2004), ICA components that may represent eye blinking, lateral eye movement, muscle activity, or channel noise were detected and excluded from further analysis. Furthermore, epochs were created from −100 to 800 ms with the time windows from −100 to 0 ms used as a baseline. An additional artificial detection was carried out so that trials were rejected if they exceeded an amplitude of 100 μV. In addition, any participant with an acceptance rate lower than 70% was excluded, which led to the exclusion of three participants from further analysis. Finally, the first ten responses of each block and two standards after each deviant were rejected in the grand average. For the difference waves, a deviant-minus-standard calculation was carried out for each participant and condition; namely, the difference was generated by subtracting the waveform of the stimuli when it was presented as standard in one block from that of the same stimuli when it was presented as deviant in another block.

RESULTS

Based on visual inspection of the grand-average waveform, the amplitudes of MMN and LDN were determined for each participant and condition as the mean amplitude within 140–180 ms and 320–360 ms after the onset of stimuli at *Fz.* According to previous studies, both the MMN and LDN are typically maximal over fronto-central electrode sites (Näätänen et al., 1992; Jakoby et al., 2011). Thus, the analyses were restricted to twelve frontocentral electrodes (*AF3, AFz, AF4, F3, Fz, F4, FC3, FCz, F4, C3, Cz,* and *C4*). For each experiment, repeated ANOVAs with *Condition*, *Vowel*, *Laterality* (left, middle, and right), and *Gradient* (AF-, F-, FC-, and C- line) as within-subject variables were carried out for mean amplitude and peak latency, respectively. For all analyses, degrees of freedom were adjusted according to the method of Greenhouse–Geisser.

Mismatch Negativity

Repeated ANOVAs were conducted and significant main effects of *Condition* and *Vowel* were found, $F_1(1, 17) = 6.99$, $p_1 = 0.017$, $\eta_p^2 = 0.29$; $F_2(1, 17) = 5.62$, $p_2 = 0.030$, $\eta_p^2 = 0.25$. However, the interaction between *Vowel* and *Condition* was also significant, $F(1, 17) = 21.39$, $p < 0.001$, $\eta_p^2 = 0.56$. *Post hoc* analyses were conducted and the results showed that for vowel [a], there was no significant difference between the mean amplitude of /ʂa/$_{[sa]}$ and /sa/$_{[ʂa]}$, $F(1, 17) = 2.69$, $p = 1.12$, $\eta_p^2 = 0.14$, indicating non-significant difference in MMN amplitudes between the features of the /ʂa/$_{[sa]}$ and /sa/$_{[ʂa]}$ word pairs as in both pairs the feature [CORONAL] of the deviant generates a no-mismatch with the underspecified [CORONAL] feature of the standard (as shown in **Figure 4**). Compared to the /sa/$_{[ʂa]}$ pair, the surface feature [LOW] of the consonant [s] in /ʂa/$_{[sa]}$ is also in a no-mismatch relationship with the underlying unspecified TONGUE FEATURE of [ʂ] and therefore, the mean amplitude of the /ʂa/$_{[sa]}$ pair was more negative than that for /sa/$_{[ʂa]}$ (**Figure 4**). However, this difference did not reach statistical significance. For word pairs with vowel [u], the amplitude of the MMN response triggered by /su/$_{[ʂu]}$ was significantly more negative than that for /ʂu/$_{[su]}$, $F(1, 17) = 33.84$, $p < 0.001$, $\eta_p^2 = 0.67$. As predicted by the FUL model, the asymmetric HEIGHT pair shows a larger MMN in the mismatch condition, when the HEIGHT feature [HIGH] of the deviant [ʂ] maps onto the pre-activated HEIGHT feature [LOW] of the standard [s]. A reduced MMN amplitude was found in the reversed condition /ʂu/$_{[su]}$, where the features [LOW] [CORONAL] of deviant [s] are in a no-mismatch relationship with the underlying features of standard [ʂu]. Furthermore, the mean amplitudes of conditions where the initial consonant of deviants was [s], were more negative when combining with vowel [a] than with [u], $F(1, 17) = 5.71$, $p = 0.029$, $\eta_p^2 = 0.25$. For conditions where the initial consonant of the deviants was [ʂ], the amplitude was more negative when followed by vowel [u] than vowel [a], $F(1, 17) = 23.90$, $p < 0.001$, $\eta_p^2 = 0.58$.

To further investigate these patterns of activation in both directions when followed by different vowels, the wave difference between /su/$_{[ʂu]}$ and /ʂu/$_{[su]}$ was compared to that between /ʂa/$_{[sa]}$ and /sa/$_{[ʂa]}$ within the 140–180 ms time window. The results showed significant differences across all gradients, *ps* < 0.001, suggesting asymmetric pattern of activation (see **Figure 5**).

Late Discriminative Negativity

Repeated ANOVAs were conducted for the LDN component and a three-way interaction between *Vowel*, *Condition*, and *Gradient* was also significant [$F(3, 51) = 7.32$, $p = 0.003$, $\eta_p^2 = 0.30$]. *Post hoc* analyses were conducted, and the results showed that no significant difference was found between the /ʂa/$_{[sa]}$ and /sa/$_{[ʂa]}$ conditions across gradients. Non-significant LDNs were also observed when surface features no mismatch the underlying underspecified [CORONAL] or unspecified [HIGH]. For words with the vowel [u], a significant difference was found between /ʂu/$_{[su]}$ and /su/$_{[ʂu]}$ where the mean amplitude of /su/$_{[ʂu]}$ was more negative than that of /ʂu/$_{[su]}$ at *AF-*, *F-*, *FC-* and *C-* [t_1 (17) = −6.24, $p_1 < 0.001$, *hedge's* $g_1 = 2.02$; t_2 (17) = −6.69, $p_2 < 0.001$, *hedge's* $g_2 = 2.15$; t_3 (17) = −6.34, $p_3 < 0.001$, *hedge's* $g_3 = 2.21$; t_4 (17) = −5.75, $p_4 < 0.001$, *hedge's* $g_4 = 1.89$]. Therefore, the subtle difference between the [sa]∼[ʂa] word pair may suggest symmetric LDNs while the TONGUE HEIGHT difference in the [su]∼[ʂu] word pair elicits an asymmetric late negativity. In addition, the mean amplitudes of conditions where the initial consonant of the deviants was [s] were more negative when combined with the vowel [a] than with [u] at *AF-*, *F-*, *FC-* and *C-* [t_1 (17) = −6.80, $p_1 < 0.001$, *hedge's* $g_1 = 1.94$; t_2 (17) = −6.32, $p_2 < 0.001$, *hedge's* $g_2 = 1.96$; t_3 (17) = −6.33, $p_3 < 0.001$, *hedge's* $g_3 = 2.01$; t_4 (17) = −5.40, $p_4 < 0.001$, *hedge's* $g_4 = 1.79$]. The wave difference between /su/$_{[ʂu]}$ and /ʂu/$_{[su]}$ was further compared to that between /ʂa/$_{[sa]}$ and /sa/$_{[ʂa]}$ in the 320–360 ms time window. The results showed that the difference was significant across all gradients, *ps* < 0.01, suggesting asymmetric pattern of activation (see **Figure 5**).

DISCUSSION

The present study was designed to examine the interactive effect of different vowels on fricative sibilants. We compared both the MMN and LDN responses to two pairs of Mandarin words ([sa]∼[ʂa] and [su]∼[ʂu]). The two consonants [s] and [ʂ] share the same place of articulation [CORONAL] but differ in TONGUE HEIGHT. As only the feature [LOW] is specified, the underlying

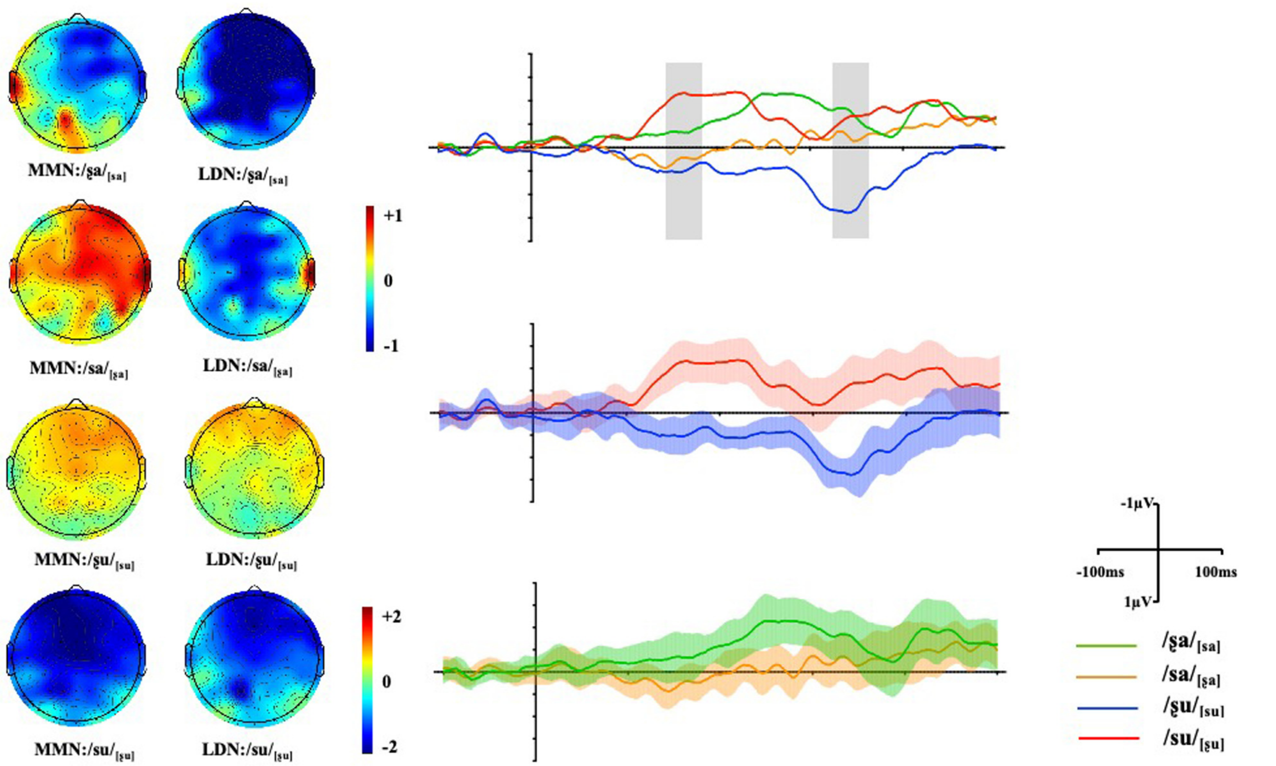


FIGURE 4 | Maps display the topographic distribution of the mean amplitude in both the MMN and LDN analysis windows from 140–180 ms and 320–360 ms respectively. Grand-average difference waveforms of all four conditions at *Fz* (see **Supplementary Material** for the waveforms at all selected electrodes). Shade areas show 95% confidence intervals.

representation of the consonant [ʂ] is unspecified for TONGUE HEIGHT. The vowels [a] and [u] mismatch in HEIGHT with [a] specified as [LOW], while [u] is specified as [HIGH]. As features can spread on the surface, the HEIGHT feature of the unspecified consonant [ʂ] changes when combined with different vowels.

Our results support the predictions of the FUL model (Lahiri and Reetz, 2002, 2010), which proposes that phonological contrasts can either match, mismatch or stand in a no-mismatch relation depending on whether the individual phonological features are fully specified or underspecified in the underlying representation. Previous studies have argued that the influence of vocalic context on fricative sibilants is due to the coarticulation of vowel rounding and consonant place of articulation (Mann and Soli, 1991). However, phonemic coarticulation would predict symmetric MMNs between phonological contrasts, independent of the direction of presentation of the standard and deviant. Thus, only an underspecification account can explain the asymmetry found in our results, as the features of vowels spread on the surface and the unspecification of TONGUE HEIGHT in the consonant [ʂ] leads to an asymmetric pattern depending on which stimulus is presented as standard and which as deviant (Lahiri and Reetz, 2002, 2010). Symmetric MMNs and LDNs were found between the no-mismatched contrasts (/ʂa/$_{[sa]}$ ∼ /sa/$_{[ʂa]}$). The feature [LOW] of vowel [a] spreads to the consonant [ʂ] when the [ʂa] is presented as deviant, resulting in the only no-mismatched feature being the underspecified [CORONAL] in both cases. When [ʂa] played the role as the standard, both the features [LOW] and [CORONAL] of consonant [s] resulted in a no-mismatch with the underlying representation of [ʂ]. In contrast, an asymmetric pattern was observed in the [su]∼[ʂu] word pair with the [HIGH] vowel [u]. When combined with the unspecified consonant [ʂ] as the deviant, the feature [HIGH] of [ʂ] conflicted with the underlying specified feature [LOW] of [s] and resulted in larger amplitudes of both MMN and LDN. No conflict was found when the feature [LOW] of the deviant [s] was in a no-mismatch relationship with the underlying unspecified [ʂ]. Consequently, the MMN and LDN amplitudes were significantly greater for the /su/$_{[ʂu]}$ pair than for /ʂu/$_{[su]}$.

Similar results for both symmetric and asymmetric MMN patterns were also reported by previous studies when considering both PLACE and MANNER features of consonants. Cornell et al. (2013) compared the phonological representations of four consonants [g], [d], [n], and [z], the first two being [PLOSIVE] and the latter two [NASAL] and [STRIDENT] respectively. Furthermore, the place feature of the first consonant is [DORSAL], while the remaining three are all [CORONAL]. The consonants were embedded in a non-word VCV structure, resulting in the sequences [egi], [edi], [eni], and [ezi]. Based on the FUL model, the features [PLOSIVE] and [CORONAL] are underspecified,

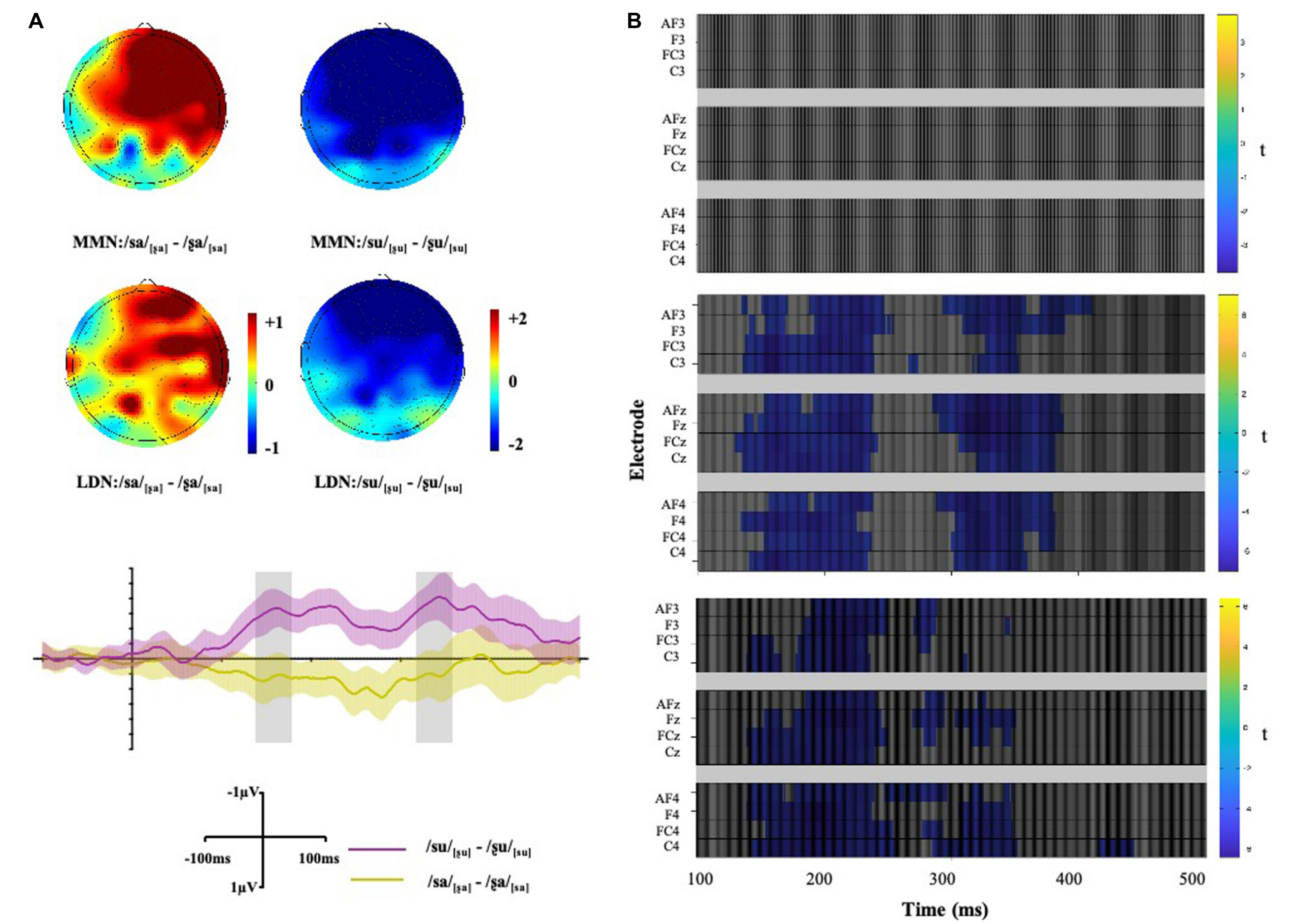


FIGURE 5 | (A) Maps display the topographic distribution of the mean amplitude difference between conditions in both the MMN (140–180 ms) and LDN (320–360 ms) analysis windows. Grand-average difference waveforms between conditions at *Fz*. Shade areas show 95% confidence intervals. **(B)** Within-subject *t*-tests between conditions at all selected electrodes after multiple comparison corrections using mass univariate ERP toolbox (Groppe et al., 2011). The difference is represented at each time point from 100 to 500 ms relative to the stimulus onset. Difference between /sa/$_{[ʂa]}$ ∼ /ʂa/$_{[sa]}$ (top); difference between /su/$_{[ʂu]}$ ∼ /ʂu/$_{[su]}$ (middle); difference between (/su/$_{[ʂu]}$ - /ʂu/$_{[su]}$) and (/sa/$_{[ʂa]}$ - /ʂa/$_{[sa]}$) (bottom).

while the others are specified in the mental representation. Asymmetric MMNs were observed in the /g/$_{[d]}$ condition as the [CORONAL] extracted from the deviant [d] conflicts with the specified feature [DORSAL] which has been activated by the standard /g/. In the reversed condition /d/$_{[g]}$, a non-conflicting situation occurs as the feature [DORSAL] extracted from the deviant [g] is tolerated (no mismatch) due to the underspecified [CORONAL] of the standard /d/. Similarly, the feature [PLOSIVE] extracted from the deviant [d] conflicts with the underlying specified [NASAL] of the standard [n] in the /n/$_{[d]}$ condition, while no conflict occurs in the reversed condition /d/$_{[n]}$ as [d] is underspecified for manner of articulation ([PLOSIVE]). In contrast, symmetric MMNs were found between [n] ∼ [z] as the features [NASAL] and [STRIDENT] are both fully specified and thus conflict equally in both directions. The results support our findings: both unspecified TONGUE HEIGHT and underspecified MANNER features can trigger asymmetric MMNs in different directions when the PLACE feature of the two consonants is kept constant. The difference is that the underspecified MANNER feature itself can trigger asymmetry while unspecified TONGUE HEIGHT feature needs to absorb additional features from surrounding segments. Therefore, different patterns of activation were found when followed by different vowels.

However, unlike the underspecification of [CORONAL], the lack of specification of [HEIGHT] is not universally applicable to all languages. It is central to the FUL model that the phonological representation of each segment is feature-based and constrained by universal properties, as well as language specific requirements (Lahiri and Reetz, 2002, 2010). Among these features, some are opposing binary pairs, such as consonantal ∼ vocalic and sonorant ∼ obstruent. The members of each pair are conflicting: a consonantal segment, for instance, cannot be vocalic and vice versa. Other features, such as [HIGH] and [LOW], are mutually exclusive but not binary. In other words, a segment cannot be both [HIGH] and [LOW] but can be neither. As discussed earlier, the number of contrastive segments in a certain language determines the specification of phonological features. In Mandarin, there is only a two-way contrast of voiceless fricative sibilants: the dental/alveolar [s] and retroflex [ʂ]. Here, the HEIGHT feature [HIGH]∼[LOW] is the only distinction between the two consonants as both of them are [CORONAL]. However, it is not necessary to specify both consonants as the phoneme that is not [HIGH] can be automatically categorized as [LOW] (Lahiri and Kennard, 2019; Kennard and Lahiri, 2020). This rule cannot be applied to segments with a three-way contrast, for instance, the Mandarin vowels. Different from two-way contrasts, both features [HIGH] and [LOW] are specified for a three-way contrast. Thus, the feature [MID] does not need to be stored and can be determined as the consequence of a binary distinction between high vs. non-high and low vs. non-low (Scharinger and Lahiri, 2010). Therefore, the results found in our study might not hold in investigations of the spreading of TONGUE HEIGHT features in other languages with a different number of contrastive segments.

Since the initial logic of the experiment was built into the framework of FUL's feature model and assumptions regarding the matching algorithm, we discussed the results in that context. However, aside from the FUL model, there are other models focusing on perception asymmetry, such as the Natural Referent Vowel (NRV) framework (Polka and Bohn, 2003, 2011) and the Native Language Magnet (NLM) theory (Kuhl, 1991, 1992, 1993). In the NRV model, Polka and Bohn suggested that vowel perception is asymmetric with respect to the location of each vowel within a traditional articulatory or F1/F2 acoustic vowel space; namely, a change from a central vowel to a peripheral vowel (e.g., from [y] to [u]) would be much easier to discriminate than the same change in the reverse direction (e.g., from [u] to [y]). Here, the peripheral vowels serve as perceptual reference for listeners to discriminate vowels and the listeners show a bias in favoring a "focal" vowel, resulting in asymmetric processing of the vowel pair in different directions. Directional asymmetry was also reported by Kuhl (1991, 1992): listeners' discrimination from a prototypical to a non-prototypical vowel within a given category is more difficult than the same change in the reverse direction. For instance, listeners were presented with a range of synthesized [i] vowels which varied in F1/F2 and asked to rate the perceived goodness of the vowels. They consistently attached the highest goodness values to vowels within a particular vowel space (Kuhl, 1991). Variants with changes to F1/F2 were synthesized on the basis of the prototype and non-prototype exemplars selected according to the ratings. Compared to a non-prototype exemplar, it is more difficult to discriminate the prototype from its variants (Kuhl, 1992). Therefore, the NLM theory argues that early linguistic experience influences perceptual patterns, such that listeners become biased toward native prototypes. These prototypes in turn function as perceptual magnets for other members within category while stretching the distance between categories (Kuhl, 1992, 1993).

However, neither of the two models are applicable to our study as the difference wave was obtained by subtracting the waveform of the stimulus when presented as standard in one block from that of the same stimulus when presented as deviant in another block. In other words, there is no difference in vowel space or phonetic category between standard and deviant. The MMN component is automatically generated by change-detection and the neurons activated by standards are separate from those activated by deviants (Jacobsen et al., 2003; Näätänen et al., 2005, 2007). The repetition of stimuli, though, might lead to a refractory effect on neurons that are either activated by the standard or the deviant, but not both. Compared to the deviant, the neural response to standards is more likely to be suppressed due to its high probability of occurrence, resulting in a misestimate MMN (Jacobsen and Schröger, 2001; Jacobsen et al., 2003). Adopting physically identical stimuli allows for the generation of genuine MMN responses without contamination by physical differences of the stimuli (Jacobsen and Schröger, 2001; Jacobsen et al., 2003). Note that subtracting the waveform of standard stimuli from that of the deviant one may not completely eliminate the potential influence of N1 on MMN, as the amplitude of N1 elicited by different stimuli varies. Previous studies also found that distinct acoustic properties of segments in a syllable or consonant-vowel transition can lead to potential P1-N1-P2, which may have an effect on the asymmetric activation of MMN and LDN (Martin and Boothroyd, 1999; Miller and Zhang, 2014). Indeed, N1 has been noted as a component which extracts phonological features (cf. Obleser et al., 2004). Future studies could use alternative measurements to separate the effects of MMN and N1, and investigate the influence of the transition within stimuli or vocalic cue on the ERP components (Schröger and Wolff, 1996; Miller and Zhang, 2014).

To sum up, our results provide neurophysiological evidence for the interactive effect of vowels on fricative sibilants in Mandarin. Features such as TONGUE HEIGHT spread on the surface so that unspecified sibilants are influenced by following vowels. When followed by a [HIGH] vowel such as [u], the unspecified sibilant [ʂ] takes on the HEIGHT feature from [u] while the specified [s] retains its own HEIGHT feature [LOW]. Therefore, asymmetries were triggered by the same phonological contrast [su] ∼ [ʂu] in two directions where the surface [HIGH] of the deviant [ʂ] conflicts with the underlying specified [LOW] of the standard [s], while the surface [LOW] activated by [s] does not mismatch with the unspecified [ʂ]. When followed by a [LOW] vowel such as [a], no such asymmetry was observed as there is no conflict between the surface [LOW] from [a] and the underlying specified [LOW]. In addition, the LDN component has demonstrated its reliability in linguistic processing among adults and its deflection pattern is roughly consistent with that of the MMN. Future studies should consider taking this component into consideration when investigating the underspecification of segments in the mental lexicon. In conclusion, not all features are fully specified in the mental lexicon and the specification of a feature such as TONGUE HEIGHT is determined by the number of contrastive segments in a certain language.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Central University Research Ethics Committee, University of Oxford. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YM: data collection, formal analysis, methodology, investigation, conceptualization, writing–original draft, and writing–review and editing. SK: methodology, conceptualization, and writing–review and editing. CX: data collection and writing–original draft. HW: writing–review and editing; AL: conceptualization, funding acquisition, methodology, project administration, supervision, writing–original draft, and writing–review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement number: 695481 awarded to AL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.617318/full#supplementary-material

REFERENCES

Barry, J. G., Hardiman, M. J., and Bishop, D. V. (2009). Mismatch response to polysyllabic nonwords: a neurophysiological signature of language learning capacity. *Plos One* 4:e6270. doi: 10.1371/journal.pone.0006270

Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer (Version 6.0. 37)[Computer software]*. Amsterdam: Institute of Phonetic Sciences.

Cai, Q., and Brysbaert, M. (2010). SUBTLEX-CH: chinese word and character frequencies based on film subtitles. *Plos One* 5:e10729. doi: 10.1371/journal.pone.0010729

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

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39:757. doi: 10.1037/a0030862

David, C., Roux, S., Bonnet-Brilhault, F., Ferré, S., and Gomot, M. (2020). Brain responses to change in phonological structures of varying complexity in children and adults. *Psychophysiology* 57:e13621.

de Jonge, M. J. I., and Boersma, P. (2015). *French high-mid vowels are underspecified for height. In Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: OUP Oxford.

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

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

Hill, P. R., McArthur, G. M., and Bishop, D. V. (2004). Phonological categorization of vowels: a mismatch negativity study. *Neuroreport* 15, 2195–2199. doi: 10.1097/00001756-200410050-00010

Højlund, A., Gebauer, L., Mc Gregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. *Lang. Cogn. Neurosci.* 34, 545–560. doi: 10.1080/23273798.2019.1572204

Holt, L. L., and Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica* 57, 170–180. doi: 10.1159/000028470

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., and Winkler, I. (2008). MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. *Psychophysiology* 45, 60–69.

Horváth, J., Roeber, U., and Schröger, E. (2009). The utility of brief, spectrally rich, dynamic sounds in the passive oddball paradigm. *Neurosci. Lett.* 461, 262–265. doi: 10.1016/j.neulet.2009.06.035

Jacobsen, T., Horenkamp, T., and Schröger, E. (2003). Preattentive memory-based comparison of sound intensity. *Audiol. Neurotol.* 8, 338–346. doi: 10.1159/000073518

Jacobsen, T., and Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* 38, 723–727. doi: 10.1111/1469-8986.3840723

Jakoby, H., Goldstein, A., and Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology* 48, 1517–1531. doi: 10.1111/j.1469-8986.2011.01227.x

Kennard, H. J., and Lahiri, A. (2020). Nonesuch phonemes in loanwords. *Linguistics* 58, 83–108. doi: 10.1515/ling-2019-0033

Korpilahti, P., Krause, C. M., Holopainen, I., and Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain Lang.* 76, 332–339. doi: 10.1006/brln.2000.2426

Korpilahti, P., Lang, H., and Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography Clin. Neurophysiol.* 4:96.

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

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percep. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Kuhl, P. K. (1992). *Infants' perception and representation of speech: Development of a new theory. in Proceedings of the International Conference on Spoken Language Processing*. Edmonton: Alberta, 449–456.

Kuhl, P. K. (1993). *Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In Developmental neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Springer, 259–274.

Kunisaki, O., and Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Ann. Bull.* 11, 85–91.

Lahiri, A. (2018). Predicting universal phonological contrasts. *Phonology and Phonetics* 229. doi: 10.1515/9783110451931-007

Lahiri, A., and Kennard, H. (2019). *Pertinacity in loanwords. In Historical Linguistics 2015: Selected papers from the 22nd International Conference on Historical Linguistics*, Vol. 348. Naples: John Benjamins Publishing Company, 57.

Lahiri, A., and Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38, 245–294. doi: 10.1016/0010-0277(91)90008-r

Lahiri, A., and Reetz, H. (1999). The FUL speech recognition system. *J. Acoust. Soc. Am.* 105, 1091–1091.

Lahiri, A., and Reetz, H. (2002). Underspecified recognition. *Lab. Phonol.* 7, 637–675. doi: 10.1515/9783110197105.2.637

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonetics* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lin, Y. H. (2007). *The Sounds of Chinese*, Vol. 1. Cambridge: Cambridge University Press.

Mann, V., and Soli, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Percep. Psychophys.* 49, 399–411. doi: 10.3758/bf03212174

Mann, V. A., and Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Percep. Psychophys.* 28, 213–228. doi: 10.3758/bf03204377

Marslen-Wilson, W. D. (1984). *Function and process in spoken word recognition: A tutorial review. In Attention and Performance: Control of Language Processes*. Mahwah, NJ: Erlbaum, 125–150.

Martin, B. A., and Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear hearing* 20, 33–44. doi: 10.1097/00003446-199902000-00004

Miller, S., and Zhang, Y. (2014). Neural coding of phonemic fricative contrast with and without hearing aid. *Ear Hearing* 35, e122–e133.

Näätänen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42, 25–32. doi: 10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Näätänen, R., Teder, W., Alho, K., and Lavikainen, J. (1992). Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport* 3, 493–496. doi: 10.1097/00001756-199206000-00009

Näätänen, R., and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* 125:826. doi: 10.1037/0033-2909.125.6.826

Norris, D., McQueen, J. M., and Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 21:1209. doi: 10.1037/0278-7393.21.5.1209

Obleser, J., Lahiri, A., and Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.* 16, 31–39. doi: 10.1162/089892904322755539

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Polka, L., and Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/s0167-6393(02)00105-x

Polka, L., and Bohn, O. S. (2011). Natural referent vowel (NRV) framework: an emerging view of early phonetic development. *J. Phonetics* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Roberts, A. C., Kotzor, S., Wetterlin, A., and Lahiri, A. (2014). Asymmetric processing of durational differencesâĂŞelectrophysiological investigations in Bengali. *Neuropsychologia* 58, 88–98. doi: 10.1016/j.neuropsychologia.2014.03.015

Scharinger, M., and Lahiri, A. (2010). Height differences in english dialects: consequences for processing and representation. *Lang. Speech* 53, 245–272. doi: 10.1177/0023830909357154

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *J. Speech Lang. Hearing Res.* 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of english fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Schröger, E., and Wolff, C. (1996). Mismatch response of the human brain to changes in sound location. *Neuroreport* 7, 3005–3008. doi: 10.1097/00001756-199611250-00041

Schulte-Körne, G., Deimel, W., Bartling, J., and Remschmidt, H. (1998). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport* 9, 337–340. doi: 10.1097/00001756-199801260-00029

Shestakova, A., Huotilainen, M., Čeponien, R., and Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clin. Neurophysiol.* 114, 1507–1512. doi: 10.1016/s1388-2457(03)00134-2

Stevens, K. N., and Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *J. Phon.* 3, 215–233. doi: 10.1016/S0095-4470(19)31431-7

Wiese, R. (1997). "Underspecification and the description of Chinese vowels," in *Studies in Chinese phonology*, eds W. Jialing and N. Smith (Berlin: Mouton de Gruyter), 219–249.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguist Inq.* 19, 65–100.

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Front. Neurosci.* 11:95. doi: 10.3389/fnins.2017.00095

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., et al. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *Neuroreport* 16, 2015–2019. doi: 10.1097/00001756-200512190-00009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 27 April 2021
doi: 10.3389/fnhum.2021.622904

Phonological Variations Are Compensated at the Lexical Level: Evidence From Auditory Neural Activity

Hatice Zora[1]*, Tomas Riad[2], Sari Ylinen[3,4] and Valéria Csépe[5]

[1] Department of Linguistics, Stockholm University, Stockholm, Sweden, [2] Department of Swedish Language and Multilingualism, Stockholm University, Stockholm, Sweden, [3] CICERO Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [4] Cognitive Brain Research Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [5] Brain Imaging Centre, Research Centre for Natural Sciences, Budapest, Hungary

OPEN ACCESS

Edited by:
Valerie L. Shafer, The City University of New York, United States

Reviewed by:
Suzanne V. H. van der Feest, The City University of New York, United States
Meera Alkaabi, United Arab Emirates University, United Arab Emirates

***Correspondence:**
Hatice Zora
hatice@ling.su.se

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *29 October 2020*
Accepted: *26 March 2021*
Published: *27 April 2021*

Citation:
Zora H, Riad T, Ylinen S and Csépe V (2021) Phonological Variations Are Compensated at the Lexical Level: Evidence From Auditory Neural Activity. Front. Hum. Neurosci. 15:622904. doi: 10.3389/fnhum.2021.622904

Dealing with phonological variations is important for speech processing. This article addresses whether phonological variations introduced by assimilatory processes are compensated for at the pre-lexical or lexical level, and whether the nature of variation and the phonological context influence this process. To this end, Swedish nasal regressive place assimilation was investigated using the mismatch negativity (MMN) component. In nasal regressive assimilation, the coronal nasal assimilates to the place of articulation of a following segment, most clearly with a velar or labial place of articulation, as in *utan mej* "without me" > [ʉːtam mɛjː]. In a passive auditory oddball paradigm, 15 Swedish speakers were presented with Swedish phrases with attested and unattested phonological variations and contexts for nasal assimilation. Attested variations – a coronal-to-labial change as in *utan* "without" > [ʉːtam] – were contrasted with unattested variations – a labial-to-coronal change as in *utom* "except" > *[ʉːtɔn] – in appropriate and inappropriate contexts created by *mej* "me" [mɛjː] and *dej* "you" [dɛjː]. Given that the MMN amplitude depends on the degree of variation between two stimuli, the MMN responses were expected to indicate to what extent the distance between variants was tolerated by the perceptual system. Since the MMN response reflects not only low-level acoustic processing but also higher-level linguistic processes, the results were predicted to indicate whether listeners process assimilation at the pre-lexical and lexical levels. The results indicated no significant interactions across variations, suggesting that variations in phonological forms do not incur any cost in lexical retrieval; hence such variation is compensated for at the lexical level. However, since the MMN response reached significance only for a labial-to-coronal change in a labial context and for a coronal-to-labial change in a coronal context, the compensation might have been influenced by the nature of variation and the phonological context. It is therefore concluded that while assimilation is compensated for at the lexical level, there is also some influence from pre-lexical processing. The present results reveal not only signal-based perception of phonological units, but also higher-level lexical processing, and are thus able to reconcile the bottom-up and top-down models of speech processing.

Keywords: phonology, assimilation, lexical access, MMN, Swedish

INTRODUCTION

Lexical access, the matching of auditory input onto lexical representations in the brain, is an essential component of speech perception. Although seemingly simple and effortless, it is a complex process given that speech is inherently highly variable. Changes in phonological shapes due to various factors, such as speech rate, dialect, coarticulation, and assimilation, make each pronunciation unique. Assimilation, which is the focus of the present paper, changes the surface forms of spoken words. It occurs when a sound is influenced by a neighboring segment and accommodates some aspect of it, as in the following Swedish example: *en båt* "a boat" > [ɛm boːt] where assimilation concerns the place of articulation of /n/. Although making articulation easier, this process introduces variability that the perceptual system has to deal with as phonological contrasts are neutralized and the lexical form of an item becomes less directly reflected. Several theories have been suggested to explain the processing of assimilatory variations, however, neither the nature of these variations and their consequences in lexical access nor the neural correlates of auditory matching mechanisms in this process are fully understood.

The aim of the present research is to investigate how listeners deal with attested phonological assimilations and unattested phonological variations during lexical access, and to elaborate on the findings with regard to previous theoretical accounts ranging from (i) *simple lexical compensation* accounts to (ii) *feature underspecification,* (iii) *regressive inference*, and (iv) *feature parsing* accounts (for an overview, see Gow, 2001; Jusczyk and Luce, 2002; Ranbom and Connine, 2007; Gaskell and Snoeren, 2008; Darcy et al., 2009; Lahiri and Reetz, 2010). Deriving from the different assumptions of these accounts, the objectives are to assess the auditory assimilatory processes operating at the pre-lexical or lexical level, the role of contextual information justifying the assimilation, and the nature of information required for the auditory matching, being either discrete phonological features, or gradient phonetic details for the perception of assimilation. These objectives are achieved by scrutinizing the neural correlates of this near-instantaneous perceptual process using the mismatch negativity (MMN) component of auditory event-related potentials (ERPs) in a well-balanced paradigm enabling the comparison of different theoretical assumptions. The MMN is considered an optimal tool to investigate the attested and unattested phonological variations given that it not only reflects the auditory variations but also the linguistic relevance of these variations in early speech comprehension processes. In the following, we first give an overview of four different theoretical accounts for phonological assimilation, present MMN studies investigating assimilatory processes, and then formulate MMN predictions based on these theoretical accounts.

The so-called *lexical compensation* account relies on stored lexical information to explain the spontaneous lexical access despite the changes in the surface forms. The auditory input is matched with words stored in the mental lexicon and the best-matching lexical item is retrieved among multiple candidates. In this account, candidates may be activated by incomplete input. A minimal mismatch between the features that are extracted from the signal and the features that comprise the lexical representation is compensated for by the listeners, and thus *gree*[m] might successfully activate *green* and any similar sounding words, since there is only one contradictory feature in the input (Marslen-Wilson and Welsh, 1978; Connine et al., 1993; Bölte and Coenen, 2002). Changes in the phonological shape and phonetic variations are tolerated based on higher-order top-down information such as semantic and syntactic contexts (Marslen-Wilson and Welsh, 1978; Samuel, 2001; Bölte and Coenen, 2002; Darcy et al., 2009; see also the TRACE model of McClelland and Elman, 1986).

Some researchers argue that the tolerance for phonological variation depends on the specification of features in the mental lexicon, such that only variations that do not mismatch the features specified in the lexical entry are allowed without obstruction of lexical access. According to the *featurally underspecified lexicon account* (FUL; Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002, 2010), for instance, the acoustic features extracted from the speech signal are matched with the phonological features in the lexicon, which stores only specified features constrained by language-specific properties. Features that exhibit variation based on the segmental or prosodic context and that can therefore be assigned by rule are not retained in the lexicon. Rather, such features are considered predictable and underspecified. The features *labial* and *dorsal*, for instance, are represented in the mental lexicon, while the feature *coronal* is considered underspecified. The coronal feature, as the universal place feature, can be activated not only by coronal but also by non-coronal information. Coronal phonemes are thus more likely to assimilate to non-coronal phonemes than the other way around. For example, labial [m] activates lexical representations of both /m/, which is specified for a labial place of articulation, and coronal /n/, which is underspecified for the place of articulation. Accordingly, *gree*[m] in a labial context as in *bean* will activate the word *green*. However, *sa*[n]*e* in a coronal context as in *duck* will not activate the word *same*. Based on these feature specifications, the FUL account suggested a ternary matching condition for the activation of word candidates: *match*, *mismatch*, and *no-mismatch*. Depending on the same or contradictory features between the signal and the representations in the lexicon, a *match* or a *mismatch* occurs, respectively. While a *match* accelerates the activation of potential candidates, a *mismatch* eliminates words as candidates. Full match and mismatch of features are not, however, the only alternatives since the lexicon would tolerate surface variations if features do not conflict. A *no-mismatch* reflects instances where (i) no feature, which is part of the mental lexicon, is extracted from the signal, or (ii) a feature, which is underspecified for the place of articulation, is extracted from the signal. A no-mismatch condition neither excludes candidates nor precludes lexical access, but receives less activation than a perfect match. Word candidates are activated according to the number of matching features as specified in the mental lexicon and the number of features extracted from the signal, along with the higher-order information (Lahiri and Reetz, 2002, 2010; Felder, 2009).

According to the FUL account, phonological underspecification is insensitive to assimilatory processes and phonological context. The *gree*[m] example above will thus activate the word *green* regardless of the following context. Experimental evidence for the indifference to the assimilatory processes and the phonological context has been indicated in a number of priming studies (Lahiri, 1995; Lahiri and van Coillie, 1999; Lahiri and Reetz, 2002; Wheeldon and Waksler, 2004). For instance, Lahiri and van Coillie (1999; cited in Lahiri and Reetz, 2002) indicated that *Bah*[m] (*Bahn* "railway" in a labial context) presented in isolation primed the semantically related word *Zug* "train" as much as the word *Bahn* did, in comparison to the unrelated word *Maus* "mouse," whereas *Lär*[n] (*Lärm* "noise" in a coronal context) did not prime the semantically related word *Krach* "bang" as much as the word *Lärm* did. There is, however, some research providing evidence for the role of phonological context in assimilatory processes (Gaskell and Marslen-Wilson, 1996, 2001; Coenen et al., 2001; Mitterer and Blomert, 2003; Mitterer et al., 2006; Gaskell and Snoeren, 2008). Using cross-modal priming, Gaskell and Marslen-Wilson (1996) for instance suggested that the perceptual system is more tolerant of assimilatory changes where the place of articulation of the following context matches with the place of articulation of the assimilated segment. According to this so-called *regressive inference account*, the perceptual system is faster and more accurate in processing assimilatory changes in phonologically appropriate contexts as in the Swedish example *en båt* "a boat" [ɛm boːt], above, than in inappropriate contexts as in *[ɛm toː] for *en tå* "a toe."

Another account that argues for context sensitivity is called the *feature parsing account* (Gow, 2001, 2003). The feature parsing account is based on acoustic processes that hold across languages, and covers both coarticulation and phonological assimilation. This account bases assimilatory operations on a pre-lexical level through basic perceptual grouping principles, and argues that the assimilated segment carries information not only about the original place of articulation present in the signal but also about the following segment (Nolan, 1992; Gow, 2000). Given the Swedish example above, [m] of the altered [ɛm] carries not only the properties of labiality of [b] of *båt*, but also the original properties of /n/ of *en*. If no trace of coronality is left in [ɛm], the original *en* cannot be parsed. Similarly, the bilabial cues cannot be parsed in the absence of a following bilabial consonant as in *[ɛm toː]. In this account, some mismatches between the extracted and expected features are tolerated, and this tolerance does not rely on the phonological nature of the variation causing the mismatch (cf. the FUL account above). Gow (2001), for instance, compared a phonological assimilation such as *gree*[m] *boat* with an example of an unattested phonological variation as in **glu*[n] *day* in a priming study. According to the FUL account, while *gree*[m] will prime *green* (a no-mismatch condition), the unattested variation **glu*[n] should not prime *glum* (a mismatch condition). The findings in Gow (2001), however, indicated no difference in priming for the two conditions. In another study, Gow (2003) investigated how listeners process assimilations by examining ambigious segments covering the acoustic properties of both coronals and labials (e.g., *cone* pronounced as [kon/m]). The results indicated, among other things, that the listeners accessed the labial alternative *comb* when the next segment was coronal as in *dents*. This was explained through perceptual grouping principles, which predict that the coronality of the /n/ in *cone* should group with the coronality of the /d/ in *dents,* and thus the coronality would be removed from the assimilated segment, which in turn would leave only the labial property to be associated with the final segment in [kon/m]. Coarticulated features in the assimilated segment are associated with the following assimilation context, and residuals of coarticulation are used to predict an upcoming segment.

Given that the matching of auditory input onto representations in the brain, be they lexical, phonological or acoustic, is near-instantaneous, the distinct neurophysiological patterns of these theoretical accounts can ideally be examined with the MMN component of ERPs, which can reflect the brain's automatic auditory information processing as early as 150–250 ms from stimulus onset. The MMN response is typically investigated using a passive oddball paradigm, where a rare stimulus (deviant) is interspersed among frequent stimuli (standard), and is elicited even when attention is directed elsewhere (Näätänen et al., 1978, 2007; Paavilainen et al., 2007; Winkler, 2007). The MMN response is optimal for investigating assimilatory processes at the pre-lexical and lexical levels, as it reflects not only low-level acoustic processing but also higher-level cognitive and linguistic processes such as activation and formation of long-term memory representations and predictive processes (Pulvermüller et al., 2001; Ylinen et al., 2010, 2016; Zora et al., 2015, 2016a,b, 2019; Garami et al., 2017). Given that the amplitude of MMN depends on the degree of variance between the stimuli (Sams et al., 1985; Pakarinen et al., 2007), several studies used the MMN component to investigate the variations introduced by assimilatory processes and their consequences for the auditory neural activity (Mitterer and Blomert, 2003; Mitterer et al., 2006; Tavabi et al., 2009). Some of these studies are reviewed in detail below.

Mitterer and Blomert (2003) investigated the phonological context dependency of assimilation in Dutch, and examined an attested change (from coronal /n/ to labial /m/) in appropriate and inappropriate phonological contexts as in *tuinbank* "garden bench" [tŒynbɑŋk] and *tuinstoel* "garden chair" [tŒynstuˑl], respectively. The authors hypothesized that if there is a regressive inference mechanism, the perceptual distance between *tuinbank* [tŒynbɑŋk] and [tŒymbɑŋk] (i.e., appropriate context) should be smaller than the distance between *tuinstoel* [tŒynstuˑl] and *[tŒymstuˑl] (i.e., inappropriate context), and accordingly, the MMN should be smaller in the appropriate context than in the inappropriate context. In line with these expectations, the results indicated smaller MMN to the assimilation of /n/ to [m] in the appropriate context [tŒymbɑŋk] than in the inappropriate one *[tŒymstuˑl]. The authors thus argued that an [m] that is followed by a [b] is perceived as a version of /n/ in automatic auditory processing, and concluded that phonological assimilations are coped with by early pre-lexical mechanisms rather than by lexical top-down mechanisms.

The MMN component has also been used to investigate whether the processing of phonological assimilations is affected by language experience and phonetic details of assimilated segments. In a series of experiments, Mitterer et al. (2006) examined Dutch listeners' perception of Hungarian liquid assimilation (from /l/ to [r]) as compared to that of native Hungarian participants. In the first experiment, MMN responses to Dutch words, where Hungarian liquid assimilation was applied as in [knɑlroːt] "vivid red" > [knɑrroːt], were recorded. As a control, an unattested variation as in [knɑlblɑu] "vivid blue" > *[knɑrblɑu] was used. Similar to the findings of Mitterer and Blomert (2003), the MMN elicited by [knɑrroːt] was smaller than the MMN elicited by *[knɑrblau]. The authors argued that Dutch listeners handle Hungarian liquid assimilation similarly to Dutch nasal place assimilations, and claimed therefore that processing of assimilations does not rely on language experience.

In the second experiment, Mitterer et al. (2006), tested whether Hungarian listeners process their native liquid assimilation in a context-dependent way like the Dutch listeners. To this end, MMN responses to Hungarian words with Hungarian liquid assimilation as in [bɔlroːl] > [bɔrroːl] and [bɔlnaːl] > *[bɔrnaːl] were investigated. Since, according to the liquid assimilation rule in Hungarian, the change from /l/ to /r/ is expected before the delative suffix [roːl] but not before the adessive suffix [naːl][1], the authors predicted the MMN to [bɔrroːl] to be smaller than the MMN to *[bɔrnaːl]. However, in contrast to the context sensitivity documented with Dutch listeners in the first experiment, the MMNs did not differ between these conditions. To see if the language background of the listeners might explain this difference, in the third experiment, the authors presented the stimuli used in the second experiment to Dutch listeners. These results were, however, similar to the ones obtained with Hungarian listeners, and the authors concluded that the difference in language background could not explain the different results between the first and second experiments.

Mitterer et al. (2006), in the fourth experiment, examined whether the acoustic quality of the stimuli were responsible for the difference in results. To this end, the authors presented altered versions of the Hungarian words (with a comparably weak /r/) used in the second and third experiments to Dutch listeners. Similar to the first experiment, the results indicated significant MMN for the inappropriate context, but not for the appropriate context, and accordingly the authors argued that context-sensitive MMN elicitation depends on the acoustic details of the stimuli. In the fifth experiment, the authors repeated the fourth experiment with Hungarian participants, and the results replicated the findings in the first and fourth experiment. The authors concluded that assimilatory processes do not rely on previous experience with a given assimilation rule. However, the phonetic details of the assimilated segment affect this process; assimilations are tolerated only when the assimilated phoneme is a weak example of the category. The authors argued that assimilatory processes take place at a pre-lexical level, independently of specific language experience in a similar fashion to coarticulatory compensation. The authors further claimed that as articulatory simplifications, assimilations are constrained by perception, and general perceptual preferences have an impact on the kind of assimilation rules applied.

This pre-lexical processing mechanism for assimilations has also been indicated using pseudowords in Tavabi et al. (2009). The authors investigated both the frequency of variation and the contextual appropriateness in nasal regressive place assimilation in items with no lexical representation. Frequent changes (from /n/ to /m/) were contrasted with rare changes (from/m/to/n/) in appropriate and inappropriate contexts (/b/or/d/). The MMN responses indicated an asymmetry in neural activity between the frequent and rare changes only in the appropriate context condition. While the rare changes elicited a much larger MMN response than the frequent changes in the appropriate contexts, the frequency of change had no significant effect in the inappropriate contexts. The authors argued that since the results were obtained using pseudowords, the lexical level is not essential for assimilatory processes in line with previous findings (Mitterer et al., 2006). The authors argued that although the results on the frequency of change provide some evidence for the FUL account, given the observed interaction between the frequency of change and the context appropriateness, their results are better understood with the feature parsing and inference accounts, which also argue for assimilatory processes operating on a pre-lexical level.

As noted earlier, the present paper aims to investigate the consequences of attested phonological assimilation and unattested phonological variation in lexical access, and to parse out neural correlates of their potential effects at the pre-lexical or lexical level using the MMN component. The processing strategies (if any) will also be elaborated in light of previous accounts as presented above. To this end, phonological variation introduced by Swedish nasal regressive place assimilation was compared to an instance of unattested phonological variation that does not appear naturally in the language. In Swedish, as in many other languages, the coronal nasal assimilates to the place of articulation of a following segment as in *en morgon* "a morning" [ɛm mɔrːgɔn], whereas the labial nasal stays unaffected

[1]Delative and adessive suffixes are Hungarian case marking suffixes, representing "from" and "at," respectively.

(e.g., *fem nålar* "five needles" [fɛm noːlar] > *[fɛn noːlar])[2] (Riad, 2014). In a well-balanced paradigm, we investigated an attested variation introduced by nasal regressive place assimilation (i.e., from coronal/n/to labial [m]) in an appropriate context as in [ʉːtan mɛjː] "without me" > [ʉːtam mɛjː] as well as in an inappropriate context [ʉːtan dɛjː] "without you" > *[ʉːtam dɛjː]. In addition and for comparison, the context sensitive interpretation of an unattested change (i.e., from labial/m/to coronal [n]) as in [ʉːtɔm mɛjː] "except me" > *[ʉːtɔn mɛjː] and [ʉːtɔm dɛjː] "except you" > *[ʉːtɔn dɛjː] was investigated. These phrases were presented in four oddball blocks; unaltered canonical versions of the phrases always served as standards ([ʉːtan mɛjː], [ʉːtan dɛjː], [ʉːtɔm mɛjː], [ʉːtɔm dɛjː]), and altered versions, whether with expected assimilation or not, served as deviants ([ʉːtam mɛjː], *[ʉːtam dɛjː], *[ʉːtɔn mɛjː], and *[ʉːtɔn dɛjː]). By examining the nature of variation and the phonological context in assimilatory processes using real words, the present paper introduces an improvement on the methodology of earlier MMN studies on the topic (Mitterer and Blomert, 2003; Mitterer et al., 2006; Tavabi et al., 2009). Although using real words, Mitterer and Blomert (2003) and Mitterer et al. (2006) looked at contextual appropriateness but did not fully investigate the nature of the change and lacked a control condition for an unattested phonological variation. Tavabi et al. (2009), although investigating both contextual appropriateness and the nature of the change, used only pseudowords and, therefore, focused only on the pre-lexical level. The current experimental paradigm admits the evaluation and comparison of different theoretical accounts. Given the sensitivity of MMN responses to any auditory differences (be it sensory or cognitive), the current experimental stimuli, which consist of phonetically and functionally identical words, critically allow the comparison of attested and unattested variations on equal grounds to a large extent in a diagonal design.

The theoretical accounts presented above are not fully exclusive. However, they differ in ascribing different roles to (i) the processing stage (pre-lexical vs. lexical), (ii) the relevance of contextual information, and (iii) the nature of information required for auditory matching (discrete phonological features vs. gradient phonetic details) for the perception of assimilation. The simple lexical compensation and FUL accounts both implement at the lexical level, and are insensitive to phonological context. But they differ in the representation of features in the mental lexicon. The former holds that all features of a word are fully specified and represented in the mental lexicon. According to this account, the perceptual system treats variations in the speech input as random noise and the higher-order information is employed to recover the signal from noise. The FUL account, on the other hand, argues that only specified features are stored in the mental lexicon, and words are activated depending on the number of matching features as specified in the lexical entries and the number of features extracted from the signal, along with the higher-order information. Both the regressive inference and feature parsing accounts claim for a pre-lexical processing stage for assimilations, and assert that the contextual appropriateness is crucial for the assimilatory processes, in contrast to the claims of simple lexical compensation and FUL accounts. However, while the regressive inference account relies on phonological rules and constraints, the feature parsing account builds on the gradient phonetic details in the signal and the language independent acoustic processes. In the feature parsing account, the auditory matching procedure does not rely on the specification of features in the mental lexicon as argued in the FUL account.

Depending on these differences across the major theoretical accounts, different patterns of MMN responses are predicted across experimental blocks (see **Table 2**). According to the simple lexical compensation account, there should not be any difference across these blocks since the correct forms would be retrieved, irrespective of the nature of the variation and the following phonological context, using semantic context. Thus, only acoustic MMN responses would be predicted in any of these blocks since assimilations would be compensated for at the lexical level. According to the FUL account, on the other hand, MMN responses are predicted to differ across [ʉːtan] and [ʉːtɔm], yet irrespective of the following phonological contexts, [mɛjː] and [dɛjː]. Accordingly, in both Block I and II, the deviants should be tolerated given the assimilation of [n] in [ʉːtan] to [m] due to the underspecification of coronal /n/ (no-mismatch condition), and consequently a smaller MMN response is predicted to the deviants. In Blocks III and IV, on the other hand, the deviants should not be tolerated by assimilation of the [m] in [ʉːtɔm] to [n], since nasal assimilation only applies to the coronal nasal (mismatch condition), and consequently a clear MMN response is predicted to the deviants. In contrast to the FUL account, according to the regressive inferential account, MMN responses are predicted to differ across [ʉːtan] and [ʉːtɔm] depending on the following context, [mɛjː] and [dɛjː]. In Block I, the deviant should be tolerated by assimilation of the [n] in [ʉːtan] to [m] due to the following phonological context

TABLE 1 | Results of one-sample *t* tests where the amplitudes of deviant-minus-standard subtractions were tested against zero.

			M	*SD*
Time window 120–180 ms	Block I	$t(14) = -0.122$, $p = 0.904$	–0.03	0.97
	Block II	$t(14) = -0.760$, $p = 0.460$	–0.21	1.07
	Block III	$t(14) = -2.414$, $p = 0.030$*	–0.70	1.12
	Block IV	$t(14) = -1.879$, $p = 0.081$	–0.44	0.90
Time window 250–300 ms	Block I	$t(14) = -1.108$, $p = 0.286$	–0.25	0.90
	Block II	$t(14) = -0.594$, $p = 0.562$	–0.09	0.65
	Block III	$t(14) = -1.709$, $p = 0.109$	–0.65	1.47
	Block IV	$t(14) = -0.682$, $p = 0.506$	–0.20	1.18
Time window 400–450 ms	Block I	$t(14) = -0.590$, $p = 0.565$	–0.14	0.92
	Block II	$t(14) = -3.447$, $p = 0.004$*	–0.53	0.60
	Block III	$t(14) = -1.503$, $p = 0.155$	–0.56	1.46
	Block IV	$t(14) = -1.137$, $p = 0.274$	–0.37	1.27

Mean amplitude values (M) are reported with standard deviations (SD).
$^*p < 0.05$.

[2] It should, however, be noted that the labial nasal assimilates to a labiodental before a labiodental fricative as in *dom farorna* "those dangers" [dɔɱ fɑːrʊŋa].

[mɛjː], and accordingly no MMN response is predicted to the deviant. In Block II, on the other hand, the deviant should not arise by assimilation of the /n/ in [ʉːtan] to [m], due to the lack of an appropriate following context for assimilation, and an MMN response is predicted to be elicited to the deviant. No direct MMN responses can be predicted according to the feature parsing account, given that the assimilated segments in the present paper consist of unmodified coronals and labials rather than segments which are phonetically ambiguous and show acoustic characteristics intermediate between underlying and surface forms, as tested in the feature parsing account. However, given that the labiality of the/m/[ʉːtam] should group with the labiality of the /m/ in [mɛjː], leaving only the coronal property to be associated with the final segment of the preceding word as in [ʉːtan], and the coronality of the /n/ in [ʉːtɔn] should group with the coronality of the /d/ in [dɛjː], thus leaving only the labial property to be associated with the final segment[3] in [ʉːtɔm], an attenuated MMN response is predicted to the deviants in Blocks I and IV compared to the deviants in Blocks II and III.

MATERIALS AND METHODS

Participants

The participants were 15 native speakers of Swedish (8 females, 7 males; age range 19–37 years, $M = 28.06$, and $SD = 5.13$). All participants were strongly right-handed as assessed with the Edinburgh Handedness Inventory (Oldfield, 1971) and reported normal development and hearing.

Ethics Statement

Written informed consent was obtained from all participants before testing. The study complied with the ethical guidelines and the experimental procedure was approved by the Stockholm Regional Ethics Committee (2019/05501).

Stimuli and Experimental Procedure

The standard stimuli were a set of Swedish phrases with attested and unattested phonological variations in various phonological contexts: (i) coronal nasal /n/, [n] followed by labial /m/, [m] as in *utan mej* [ʉːtan mɛjː] "without me"; (ii) coronal nasal /n/, [n] followed by coronal /d/, [d] as in *utan dej* [ʉːtan dɛjː] "without you"; (iii) labial /m/, [m] followed by labial /m/, [m] as in *utom mej* [ʉːtɔm mɛjː] "except me"; and (iv) labial /m/, [m] followed by coronal [d] as in *utom dej* [ʉːtɔm dɛjː] "except you." The deviant stimuli consisted of either attested phonological assimilations or unattested variations, created through changes from /n/ to [m] as in [ʉːtam mɛjː] and *[ʉːtam dɛjː], and through changes from /m/ to [n] as in *[ʉːtɔn mɛjː] and *[ʉːtɔn dɛjː][4].

A speech and language pathologist (female from Stockholm, 60 years old) produced all the stimuli in an anechoic chamber. The recordings were performed using the REAPER digital audio workstation (version 5.93; 44.1 kHz/16 bits). Acoustic analysis and manipulations were carried out on exemplars selected among several repetitions of each stimulus using Praat (version 6.0.33; Boersma and Weenink, 2014). Selected exemplars were segmented, where boundaries were determined by visual inspection of waveforms and Gaussian window broadband spectrograms (bandwidth = 260 Hz). Extracted segments were then matched for duration using the Vocal Toolkit plugin (Corretge, 2012), while preserving the other acoustic characteristics. In order to keep the deviants and standards identical and get an equal ground for the comparison, the stimuli differed from each other only in the variable segments /a/, /o/, /n/ and /m/ as well as /dej/ and /mej/. The deviant stimulus [ʉːtam mɛjː], for instance, was created out of the standard stimulus [ʉːtan mɛjː] by a splicing technique; the critical segment [n] was extracted from the relevant context and replaced with [m]. The critical segments were extracted and spliced at zero-crossings in order to avoid spurious clicks in the spliced signal, and pulses were added and deleted when necessary. To eliminate spurious clicks at the beginning and end of the stimuli, 2 ms ramps were added to the onset and offset. The length of each stimulus was 800 ms, and the divergence point between standards and deviants was at 400 ms. The acoustic quality of the stimuli was validated by independent judgment of five listeners, including the authors themselves.

The stimuli were presented in a passive auditory oddball paradigm using E-Prime (version 2.0). The stimuli were delivered via loudspeakers at a comfortable listening level while a silent movie was used to direct participants' attention away from the auditory stimuli. The experiment had four blocks and each block consisted of 600 stimuli – 480 standards (80%) and 120 deviants (20%), following the typical probabilities of the oddball paradigm. The order of the blocks was counterbalanced across participants. The deviants were semi-randomly placed among the standards (with at least two intervening standards between the two consecutive deviants) and a random interstimulus interval (ISI) was used to avoid rhythmicity. The ISI was centered around 500 ms, with a range between 450 and 550 ms.

EEG Data Collection and Analysis

The electroencephalography (EEG) data were collected with the BioSemi ActiveTwo system and the ActiView acquisition software (BioSemi, Netherlands) in an electrically insulated and sound-attenuated recording booth. Recordings were made from sixteen cap-mounted active electrodes (Fp1, Fp2, F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, O1, Oz, and O2) positioned according to the International 10–20 system. A common mode sense active electrode and a driven right leg passive electrode replaced the ground electrode. Electrooculogram and nose data (used for offline referencing) were collected through external electrodes.

The EEG data analysis was performed in Matlab (version 9.4; The Math Works Inc., Natick, MA, United States) using the EEGLAB toolbox (Delorme and Makeig, 2004). The continuous EEG data were filtered using a finite impulse response band-pass filter of 0.5 to 30 Hz. The channels were re-referenced to the nose channel. The EEG data were decomposed using independent component analysis (Jung et al., 2000) and eye

[3]The coronal and labial properties are associated with the final segments of the preceding words depending on the semantic context.

[4]The normative spellings of these pronouns are <mig> and <dig>.

artifacts, which were in leading positions in the component array, were then removed from the data. On average, two components were removed. The EEG data were segmented into epochs of 1,200 ms and baseline corrected using the 100 ms pre-divergence interval. Additional artifact rejection was carried out automatically, removing any epochs containing EEG fluctuation exceeding ±100 μV (4.3% excluded trials in total).

A time window of 50 ms, centered at the peak latency, was used for the quantification of the MMN amplitude. The statistical analysis of these data was performed in SPSS (International Business Machines Corp., Armonk, NY, United States). Mean amplitude values from frontal electrodes (F3, Fz, and F4) at three time windows (120–180, 250–300, and 400–450 ms from the divergence point) were selected for the analysis. The time windows were chosen to optimally capture ERP modulations related to target phonemes or syllables in grand-average waveforms. In order to test whether MMN responses significantly differed from zero, deviant-minus-standard difference amplitudes were tested against zero with one-sample *t*-tests. To evaluate the overall effect of deviations on the ERP responses, two-way repeated-measures ANOVAs with factors Block (I–IV) and Stimuli (Standard and Deviant) were subsequently carried out in each time window. Effect sizes are given in partial η2 measures, and mean values are reported with standard deviations.

RESULTS

The grand-average ERPs for the standard and deviant stimuli from Fz are displayed for each block in **Figure 1**. The results from *t*-tests and the ANOVAs are presented in detail below and elaborated on with regard to MMN predictions.

Results From *t*-Tests

Deviant-minus-standard difference amplitudes were tested against zero with one-sample *t*-tests. The results of *t*-tests are presented in **Table 1**, and mean and the standard error of the

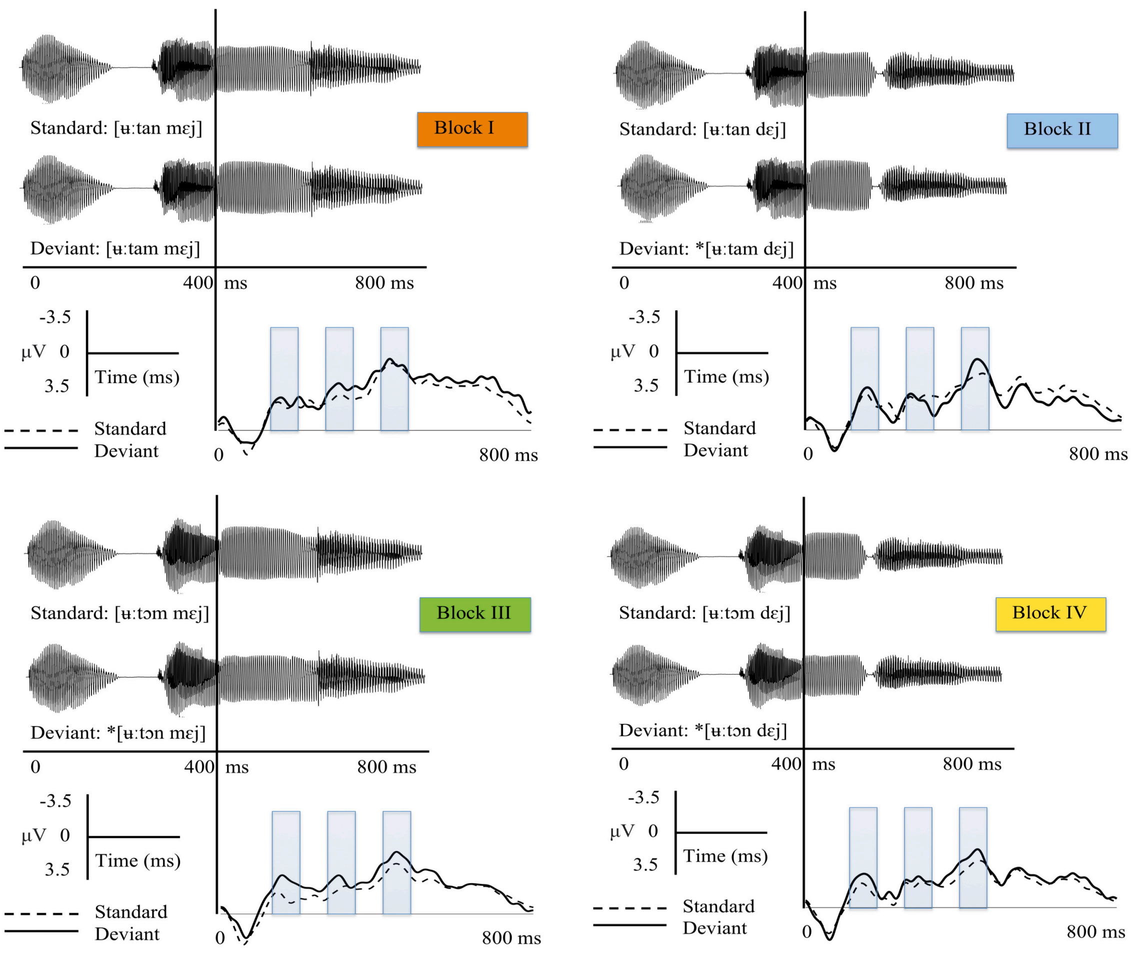


FIGURE 1 | Sound and grand-average ERP waveforms (from Fz) for the standard and deviant stimuli in each block. Blocks are color-coded in line with the bar graphs presented in **Figure 2**. The black lines show the ERPs for the deviant stimuli, the dashed lines the ERPs for the standard stimuli. The divergence point was used as zero point in the ERP figures given that the standards and deviants were identical up to the assimilation point. The shaded bars represent time windows selected for statistical analysis. Asterisks mark inappropriate/unattested deviant sequences.

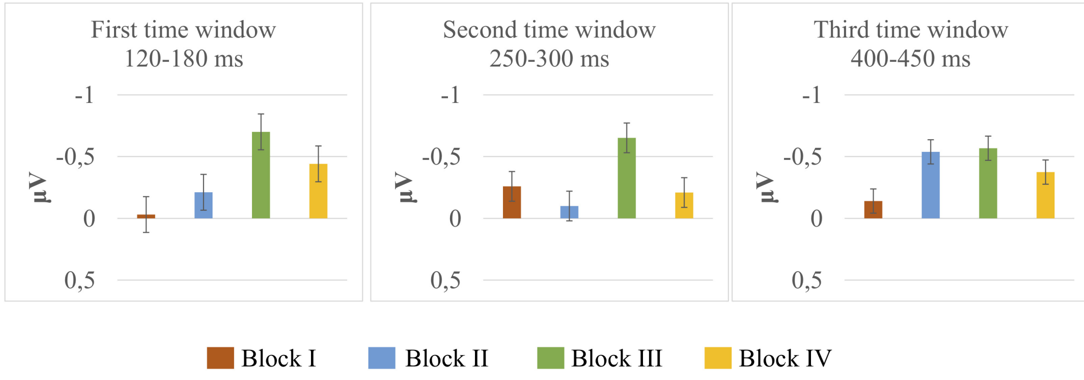


FIGURE 2 | Mean and the standard error of the mean for deviant-minus-standard subtraction amplitudes extracted from the frontal electrodes (F3, Fz, and F4) in microvolts (µV) of Block I (Orange bar), Block II (Blue bar), Block III (Green bar), and Block IV (Yellow bar) at three time windows.

mean for deviant-minus-standard difference amplitudes are illustrated in **Figure 2**. The results in the first time window (120–180 ms) indicated that the amplitudes did not differ from zero in Block I [$t(14) = -0.122$, $p = 0.904$]; Block II [$t(14) = -0.760$, $p = 0.460$]; and Block IV [$t(14) = -1.879$, $p = 0.081$]. A significant difference was present only in Block III [$t(14) = -2.414$, $p = 0.030$, $M = -0.70$, and $SD = 1.12$]. The results in the second time window (250–300 ms) showed no significant differences in any of the blocks (Block I [$t(14) = -1.108$, $p = 0.286$]; Block II [$t(14) = -0.594$, $p = 0.562$]; Block III [$t(14) = -1.709$, $p = 0.109$]; and Block IV [$t(14) = -0.682$, $p = 0.506$]). In the third time window (400–450 ms), there was a significant difference in Block II [$t_{(14)} = -3.447$, $p = 0.004$, $M = -0.53$, and $SD = 0.60$], whereas the amplitudes did not differ from zero in Blocks I [$t(14) = -0.590$, $p = 0.565$]; III [$t(14) = -1.503$, $p = 0.155$]; and IV [$t(14) = -1.137$, $p = 0.274$].

Results From ANOVAs

Two-way repeated-measures ANOVAs with factors Block (I–IV) and Stimuli (Standard and Deviant) were carried out in each time window. The results of two-way repeated measures ANOVA in the first time window (120–180 ms) indicated no significant main effect Block [$F_{(3, 42)} = 1.143$, $p = 0.343$, and $\eta 2 = 0.075$]. However, a significant main effect of Stimuli [$F_{(1, 14)} = 5.555$, $p = 0.034$, and $\eta 2 = 0.284$] was found. The analysis yielded no significant interaction between these factors [$F_{(3, 42)} = 1.300$, $p = 0.287$, and $\eta 2 = 0.085$]. In the second time window (250–300 ms), neither the main effects Block [$F_{(3, 42)} = 0.162$, $p = 0.921$, and $\eta 2 = 0.011$] and Stimuli [$F_{(1, 14)} = 4.078$, $p = 0.063$, and $\eta 2 = 0.226$] nor interaction between them [$F_{(3, 42)} = 0.751$, $p = 0.528$, and $\eta 2 = 0.051$] reached significance. Similar to the first time window, in the third time window (400–450 ms) there was no significant main effect of Block [$F_{(3, 42)} = 1.616$, $p = 0.199$, and $\eta 2 = 0.104$] but we found a significant main effect of Stimuli [$F_{(1, 14)} = 6.063$, $p = 0.027$, and $\eta 2 = 0.302$]. There was no significant interaction between these factors either [$F_{(3, 42)} = 0.516$, $p = 0.674$, and $\eta 2 = 0.036$]. Significant main effects of Stimuli in the first and third time windows indicated larger negative deflections for the deviant stimuli. To validate the results from ANOVAs, LMMs with Block and Stimuli as fixed factors were carried out both unstructured and with compound symmetry structure. In all cases, the results were identical to those from ANOVAs. The unstructured LMM from the first time window indicated a significant effect of Stimuli ($p = 0.028$). Block ($p = 0.308$) and Interaction ($p = 0.371$) effects did not, however, reach significance. The results from the third time window revealed similar patterns (Block, $p = 0.181$; Stimuli, $p = 0.022$; and Interaction, $p = 0.557$). The results of the compound symmetry structure in the first time window yielded a significant effect of Stimuli ($p = 0.009$). There were, however, no significant effects of Block ($p = 0.289$) and Interaction ($p = 0.295$). Similarly, in the third time window, there was a significant effect of Stimuli ($p = 0.017$) while the effects of Block ($p = 0.069$) and Interaction ($p = 0.792$) displayed no significance.

MMN Interpretations

The results of the ANOVAs showed no significant interactions between the stimuli and blocks, indicating that the variations are treated in the same way across the blocks. Although the ANOVAs did not show any significant interactions, the grand average waveforms and the MMN responses verified by one-sample *t*-tests suggested differences in MMN elicitation that may be influenced by the variation type. The results indicated, for instance, that an unattested change (i.e., a labial-to-coronal change in a labial context, [ʉːtɔm] > [ʉːtɔn] + [mɛjː]) elicited a significant MMN response at an early stage (Block III), and an attested change in an inappropriate context (i.e., a coronal-to-labial change followed by a coronal context, [ʉːtan] > [ʉːtam] + [dɛjː]) elicited a significant MMN response at a later time window (Block II), whereas the other variations did not elicit significant MMNs (see **Figure 2** and **Table 1**). In the early time window, there was further a tendency for an MMN response to another unattested change (i.e., a labial-to-coronal change in a coronal context, [ʉːtɔm] > [ʉːtɔn] + [dɛjː]), yet this response was not robust enough to reach significance (Block IV).

DISCUSSION

Transformation of auditory input into a meaningful representation is affected by several constraints, including attested and unattested phonological variations in the speech signal. The present paper investigated the consequences of attested phonological assimilations and unattested phonological variations in lexical access, and elaborated on different theoretical accounts for phonological assimilation. The attested case of phonological variation introduced by Swedish nasal regressive place assimilation was scrutinized in appropriate and inappropriate phonological contexts. For comparison, an instance of unattested phonological variation that does not appear naturally in the language was investigated in the relevant contexts. The results showed no significant interactions between the variations, indicating that the correct forms were retrieved from the signal, irrespective of the variations. However, there were differences in MMN elicitation that may be influenced by the nature of variations and phonological contexts. In the rest of the paper, we discuss these findings in light of various theoretical accounts for phonological assimilation and their MMN predictions as presented in the Introduction section (see **Table 2**).

According to the simple top-down lexical compensation account (e.g., Marslen-Wilson and Welsh, 1978; Samuel, 2001; Bölte and Coenen, 2002; Darcy et al., 2009; see also the tolerance-to-mismatch approach in Gow, 2001), no MMN difference is expected across the experimental blocks since the deviation from the canonical form will be compensated for at a lexical level, using semantic cues, irrespective of the nature of variation and the following phonological context. Given that there was no significant interaction between any of the deviations and the phonological context, the results are claimed to be in line with this lexical compensation account. It can be argued that the attested and unattested changes were treated in the same way across the appropriate and inappropriate contexts, and the variations did not incur an apparent cost in lexical access. The listeners may have successfully repaired the deviations since the extracted inputs from the deviants differed from the lexical representations formed by the standards only in one feature. Given also that this difference occurred at the end of the words, which underwent a phonological change, and that the difference between/m/and/n/is subtle, the brain might have corrected and compensated for the differences between these forms after several repetitions. These results are in line with the theories of spoken word recognition that assume a top-down influence of lexical representations on the activation of smaller perceptual units rather than a fully bottom-up flow of information (Marslen-Wilson and Welsh, 1978; Samuel, 2001; Bölte and Coenen, 2002; Darcy et al., 2009).

In contrast to the previous MMN studies, which argue for assimilatory processes operating on a pre-lexical level and argue in favor of the feature parsing and inference accounts (Mitterer and Blomert, 2003; Mitterer et al., 2006; Tavabi et al., 2009), the present paper indicated that the attested phonological assimilations as well as unattested phonological variations are compensated for at the lexical level. Although the current results do not provide unequivocal support for the other accounts reviewed in the present paper, they should not be dismissed fully. The MMN responses verified by one-sample *t*-tests suggested differences in MMN elicitation that may be affected by the nature of variation and the phonological context. A late MMN response to Block II (an attested change in an inappropriate context) and an early MMN response to Block III (an unattested change in the labial context), are partially in line with the MMN predictions of the FUL, regressive inference and feature parsing accounts, which are further discussed below.

The FUL account (e.g., Lahiri and Marslen-Wilson, 1991; Lahiri and Reetz, 2002, 2010) predicts different MMN responses to [ʉːtan] > [ʉːtam] and [ʉːtɔm] > [ʉːtɔn], yet regardless of the following phonological contexts, [mɛjː] and [dɛjː]. According to this account, a smaller MMN response is predicted to the deviants in both Block I and II, since the deviants – a change from [n] to [m] as in [ʉːtan] > [ʉːtam] – will be tolerated given the underspecification of coronal/n/and therefore a no-mismatch condition. The significant MMN response in Block II thus contradicts the FUL account. The MMN response in Block III is, however, in line with the FUL account, which predicts a clear MMN response to the deviants in Blocks III and IV, since the deviants – a change from [m] in [ʉːtɔm] to [n] as in [ʉːtɔm] > [ʉːtɔn] – will not be tolerated given that nasal

TABLE 2 | Excerpts from each block and the relevant theoretical accounts and their MMN predictions.

Block	Change	Context	Standard	Deviant
Block I	Coronal [n] > Labial [m] Attested	Labial [m] Appropriate	[ʉːtan mɛjː]	[ʉːtam mɛjː]
Block II	Coronal [n] > Labial [m] Attested	Coronal [d] Inappropriate	[ʉːtan dɛjː]	[ʉːtam dɛjː]
Block III	Labial [m] > Coronal [n] Unattested	Labial [m]	[ʉːtɔm mɛjː]	[ʉːtɔn mɛjː]
Block IV	Labial [m] > Coronal [n] Unattested	Coronal [d]	[ʉːtɔm dɛjː]	[ʉːtɔn dɛjː]
Theoretical account	**MMN predictions**			
Top-down lexical compensation	No difference across Blocks			
Feature underspecification (FUL)	Smaller MMN for Blocks I and II in comparison to Blocks III – IV			
Regressive inference	Smaller MMN for Block I in comparison to Blocks II – III – IV			
Feature parsing	Smaller MMN for Blocks I and IV in comparison to Blocks II – III			

Block I: an attested change in an appropriate context – a coronal-to-labial change followed by a labial context; Block II: an attested change in an inappropriate context – a coronal-to-labial change followed by a coronal context; Block III: an unattested change – a labial-to-coronal change followed by a labial context; and Block IV: an unattested change – a labial-to-coronal change followed by a coronal context.

assimilation only applies to the coronal nasal and a change from [m] to [n] creates a mismatch condition. However, according to the FUL account, this MMN response should be present in both phonological contexts, yet the response was not robust enough to reach significance in the coronal context (see Block IV).

Given the early MMN response in Block III and the marginally significant early MMN response in Block IV, one can still argue that the labial-to-coronal change was, in fact, directly perceived as incorrect prior to the following context, providing support for the FUL account. One can, however, also argue that the MMN response in Block II was late likely because the coronal-to-labial change remained acceptable until the onset of the following context; [ʉːtam] was perceived as incorrect only after encountering the [dɛjː] context, which, in turn, provides evidence for the regressive inference account.

In contrast to the FUL account, the regressive inference account argues that assimilatory changes are processed faster and more accurately in phonologically appropriate contexts (e.g., Gaskell and Marslen-Wilson, 1996, 2001; Coenen et al., 2001; Mitterer and Blomert, 2003; Mitterer et al., 2006; Gaskell and Snoeren, 2008; Tavabi et al., 2009). The regressive inference account predicts different MMN responses to [ʉːtan] > [ʉːtam] depending on the following context, [mɛjː] and [dɛjː]. In Block I, the deviant will be tolerated by assimilation of the [n] in [ʉːtan] to [m] due to the following phonological context [mɛjː], and accordingly no MMN response is predicted to the deviant. In Block II, on the other hand, the deviant will not arise by assimilation of the /n/ in [ʉːtan] to [m], due to the lack of a following context for assimilation, and an MMN response is predicted to be elicited to the deviant. The MMN response in Block II therefore provides support for the regressive inference account. This finding is also in line with previous research, which has indicated larger MMN response to an inappropriate context for assimilation (Mitterer and Blomert, 2003; Mitterer et al., 2006).

The reported late MMN response to Block II and early MMN response to Block III provides support for the feature parsing account, which predicted an attenuated MMN response to the deviants in Blocks I and IV compared to the deviants in Blocks II and III. The feature parsing account argues that an assimilated segment accomodates information not only about the original place of articulation present in the signal but also about the following segment (Gow, 2003). In this account, as long as they follow the grouping principles, no difference is expected between an attested phonological assimilation and an unattested phonological variation (see the priming experiment in Gow, 2001). The current MMN findings do not provide direct evidence for the feature parsing account since the unmodified coronals and labials were used as assimilated segments rather than intermediate, phonetically ambiguous segments, as used in Gow, 2003. However, given that the labiality of the assimilated segment in [ʉːtam] will be associated with the labiality of the following context in [mɛjː], leaving only the coronal property to be associated with the final segment of the word candidate as in [ʉːtan], and the coronality of the assimilated segment in [ʉːtɔn] will group with the coronality of the following segment in [dɛjː], leaving thus only the labial property to be related to the final segment of the word candidate in [ʉːtɔm], smooth word recognition was possible in variations as in Blocks I and IV, and accordingly smaller MMN responses were elicited to these variations.

The present pattern of results is in line with previous research, which argues that general perceptual preferences and phonetic details have an impact on the kind of assimilation rules applied (e.g., Mitterer et al., 2006). For instance, an indifference to contextual appropriateness, reported for the second experiment in Mitterer et al. (2006), was shown to depend on the acoustic details of the stimuli; the authors could in fact document the impact of context on assimilatory processes after changing the phonetic details of the stimuli (see for instance the fifth experiment). One can also argue that the current results indicate that the consonant sequences might favor the same place of articulation; if the change leads to a mismatch between the assimilated segment and the following segment with regard to the place of articulation, a larger MMN response is elicited, indicating a low-level perceptual processing independent of the nature of variation.

To conclude, the processing of phonological variations is contributed by lexical representations. For successful lexical access, there is no need for a close match between the auditory information extracted from the signal and lexical representations. Even unattested phonological variations successfully activate lexical representations, and a minimal mismatch between the features that are extracted from the signal and the features that comprise the lexical representations is compensated for at the lexical level. The results, however, indicate a hint of pre-lexical processing and point out context sensitivity to some extent in a similar fashion suggested in the feature parsing account. These findings thus raise the need for further comparisons, which can be obtained by changing the nature of the stimuli by introducing gradient modification of place of articulation, and by testing other target languages. By establishing the neural correlates of attested and unattested phonological variations and their consequences in lexical processing, the present study contributes to the understanding of inherently variable spoken language communication and automatic lexical access, which is particularly important given the rapid nature of spoken communication. The findings are relevant for explaining our ability to effectively recognize words despite variations as a result of assimilatory process as well as variations introduced by other factors such as speech rate, dialect and background noise. Most importantly, the present study attempts to provide a unified account of spoken language processing by deriving and testing the predictions of competing theoretical accounts on assimilatory processes. Revealing not only low-level perceptual processing of phonological units, but also higher-level lexical processing, the present pattern of results harmonizes the bottom-up and top-down theories of speech processing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Stockholm Regional Ethics Committee (2019/05501). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HZ, TR, SY, and VC: conception and design of the work. HZ: experimental work and drafting the manuscript. HZ, TR, SY, and VC: revision and final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Research Collaboration between Stockholm University and University of Helsinki [SU FV-5.1.2-0757-15] and Knut and Alice Wallenberg Foundation (Reg. nr. SU FV-2.1.9-0171-19).

ACKNOWLEDGMENTS

We are grateful to the Phonetics Laboratory at the Department of Linguistics, Stockholm University for providing the experimental facilities. We would like to thank Ambika Kirkland for her proofreading, and the reviewers for their constructive and insightful comments.

REFERENCES

Boersma, P., and Weenink, D. (2014). *Doing phonetics by computer, Version 5.3.24*.

Bölte, J., and Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner? *Brain Lang.* 81, 384–397. doi: 10.1006/brln.2001.2532

Coenen, E., Zwitserlood, P., and Bölte, J. (2001). Variation and assimilation in German: consequences of assimilation for word recognition and lexical representation. *Lang. Cogn. Process.* 16, 535–564. doi: 10.1080/01690960143000155

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *J. Mem. Lang.* 32, 193–210. doi: 10.1006/jmla.1993.1011

Corretge, R. (2012). *Praat Vocal Toolkit*. Available online at: http://www.praatvocaltoolkit.com (accessed July, 2020).

Darcy, I., Ramus, F., Christophe, A., Kinzler, K. D., and Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. *Variat. Grad. Phonet. Phonol.* 14:265.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Felder, V. (2009). *Neurolinguistic Evidence for the Representation and Processing of Tonal and Segmental Information in the Mental Lexicon*. PhD Dissertation, Universität of Konstanz, Konstanz.

Garami, L., Ragó, A., Honbolygó, F., and Csépe, V. (2017). Lexical influence on stress processing in a fixed-stress language. *Int. J. Psychophysiol.* 117, 10–16. doi: 10.1016/j.ijpsycho.2017.03.006

Gaskell, G., and Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 144–158. doi: 10.1037/0096-1523.22.1.144

Gaskell, G., and Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1632–1647. doi: 10.1037/a0011977

Gaskell, M. G., and Marslen-Wilson, W. D. (2001). Lexical ambiguity and spoken word recognition: bridging the gap. *J. Mem. Lang.* 44, 325–349. doi: 10.1006/jmla.2000.2741

Gow, D. W. (2000). "Assimilation, ambiguity, and the feature parsing problem," in *Proceedings of the Sixth International Conference of Speech and Language Processing*, Beijing, 535–538.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *J. Mem. Lang.* 45, 133–159. doi: 10.1006/jmla.2000.2764

Gow, D. W. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Percept. Psychophys.* 65, 575–590. doi: 10.3758/bf03194584

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178. doi: 10.1111/1469-8986.3720163

Jusczyk, P. W., and Luce, P. A. (2002). Speech perception and spoken word recognition: past and present. *Ear Hear.* 23, 2–40. doi: 10.1097/00003446-200202000-00002

Lahiri, A. (1995). Undoing place assimilation. *J. Acoust. Soc. Am.* 97:3333. doi: 10.1121/1.412780

Lahiri, A., and Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the lexicon. *Cognition* 38, 245–294. doi: 10.1016/0010-0277(91)90008-r

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Labphon 7*, eds C. Gussenhoven, N. Werner, and T. Rietveld (Berlin: Mouton de Gruyter), 637–676. doi: 10.1515/9783110197105.2.637

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonet.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lahiri, A., and van Coillie, S. (1999). *Non-Mismatching Features in Language Comprehension*. Konstanz: University of Konstanz.

Marslen-Wilson, W., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognit. Psychol.* 10, 29–63. doi: 10.1016/0010-0285(78)90018-x

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

Mitterer, H., and Blomert, L. (2003). Coping with phonological assimilation in speech perception: evidence for early compensation. *Percept. Psychophys.* 65, 956–969. doi: 10.3758/bf03194826

Mitterer, H., Csépe, V., Honbolygo, F., and Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cogn. Sci.* 30, 451–479. doi: 10.1207/s15516709cog0000_57

Näätänen, R., Gaillard, A. W., and Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nolan, F. (1992). "The descriptive role of segments," in *Papers in Laboratory Phonology II*, eds G. J. Docherty and D. R. Ladd (Cambridge: Cambridge University Press), 261–280. doi: 10.1017/cbo9780511519918.011

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Paavilainen, P., Arajärvi, P., and Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport* 18, 159–163. doi: 10.1097/wnr.0b013e328010e2ac

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., and Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clin. Neurophysiol.* 118, 177–185. doi: 10.1016/j.clinph.2006.09.001

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage* 14, 607–616. doi: 10.1006/nimg.2001.0864

Ranbom, L., and Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *J. Mem. Lang.* 57, 273–298. doi: 10.1016/j.jml.2007.04.001

Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychol. Sci.* 12, 348–351. doi: 10.1111/1467-9280.00364

Tavabi, K., Elling, L., Dobel, C., Pantev, C., and Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: a magnetoencephalography study. *PLoS One* 4:e4452. doi: 10.1371/journal.pone.0004452

Wheeldon, L., and Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain Lang.* 90, 401–412. doi: 10.1016/s0093-934x(03)00451-6

Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163. doi: 10.1027/0269-8803.21.34.147

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., and Paavilainen, P. (2016). Predictive coding of phonological rules in auditory cortex: a mismatch negativity study. *Brain Lang.* 162, 72–80. doi: 10.1016/j.bandl.2016.08.007

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., et al. (2010). Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *J. Cogn. Neurosci.* 22, 1319–1332. doi: 10.1162/jocn.2009.21272

Zora, H., Heldner, M., and Schwarz, I. C. (2016a). Perceptual correlates of Turkish word stress and their contribution to automatic lexical access: evidence from early ERP components. *Front. Neurosci.* 10:7. doi: 10.3389/fnins.2016.00007

Zora, H., Riad, T., Schwarz, I. C., and Heldner, M. (2016b). Lexical specification of prosodic information in Swedish: evidence from mismatch negativity. *Front. Neurosci.* 10:533. doi: 10.3389/fnins.2016.00533

Zora, H., Riad, T., and Ylinen, S. (2019). Prosodically controlled derivations in the mental lexicon. *J. Neurolinguist.* 52:100856. doi: 10.1016/j.jneuroling.2019.100856

Zora, H., Schwarz, I. C., and Heldner, M. (2015). Neural correlates of lexical stress: mismatch negativity reflects fundamental frequency and intensity. *NeuroReport* 26, 791–796. doi: 10.1097/wnr.0000000000000426

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 03 June 2021
doi: 10.3389/fnhum.2021.607148

Neurophysiological Correlates of Asymmetries in Vowel Perception: An English-French Cross-Linguistic Event-Related Potential Study

Linda Polka[1,2], Monika Molnar[3], T. Christina Zhao[4,5] and Matthew Masapollo[6]*

[1] School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada, [2] Center for Research on Brain, Language, and Music, McGill University, Montreal, QC, Canada, [3] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada, [4] Institute for Learning and Brain Sciences, University of Washington, Seattle, WA, United States, [5] Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, United States, [6] Department of Speech, Language, and Hearing Sciences, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Arild Hestvik,
University of Delaware, United States

Reviewed by:
Yan H. Yu,
St. John's University, United States
Andreas Højlund,
Aarhus University, Denmark

***Correspondence:**
Linda Polka
linda.polka@mcgill.ca

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: 16 September 2020
Accepted: 03 May 2021
Published: 03 June 2021

Citation:
Polka L, Molnar M, Zhao TC and Masapollo M (2021) Neurophysiological Correlates of Asymmetries in Vowel Perception: An English-French Cross-Linguistic Event-Related Potential Study. Front. Hum. Neurosci. 15:607148. doi: 10.3389/fnhum.2021.607148

Behavioral studies examining vowel perception in infancy indicate that, for many vowel contrasts, the ease of discrimination changes depending on the order of stimulus presentation, regardless of the language from which the contrast is drawn and the ambient language that infants have experienced. By adulthood, linguistic experience has altered vowel perception; analogous asymmetries are observed for non−native contrasts but are mitigated for native contrasts. Although these directional effects are well documented behaviorally, the brain mechanisms underlying them are poorly understood. In the present study we begin to address this gap. We first review recent behavioral work which shows that vowel perception asymmetries derive from phonetic encoding strategies, rather than general auditory processes. Two existing theoretical models–the Natural Referent Vowel framework and the Native Language Magnet model–are invoked as a means of interpreting these findings. Then we present the results of a neurophysiological study which builds on this prior work. Using event-related brain potentials, we first measured and assessed the mismatch negativity response (MMN, a passive neurophysiological index of auditory change detection) in English and French native-speaking adults to synthetic vowels that either spanned two different phonetic categories (/y/vs./u/) or fell within the same category (/u/). Stimulus presentation was organized such that each vowel was presented as standard and as deviant in different blocks. The vowels were presented with a long (1,600-ms) inter-stimulus interval to restrict access to short-term memory traces and tap into a "phonetic mode" of processing. MMN analyses revealed weak asymmetry effects regardless of the (i) vowel contrast, (ii) language group, and (iii) MMN time window. Then, we conducted time-frequency analyses of the standard epochs for each vowel. In contrast to the MMN analysis, time-frequency analysis revealed significant differences in brain oscillations in the theta band (4–8 Hz), which have been linked to attention and processing efficiency. Collectively, these findings suggest that early-latency (pre-attentive) mismatch responses may not be a strong neurophysiological correlate of asymmetric behavioral vowel discrimination. Rather, asymmetries may reflect differences in neural processing efficiency for vowels with certain inherent acoustic-phonetic properties, as revealed by theta oscillatory activity.

Keywords: vowel perception, mismatch negativity, prototypes, natural referent vowel framework, native language magnet model, brain rhythms

INTRODUCTION

A central goal of research in the field of speech perception is to explicate how listeners map the input acoustic signal onto the phonetic categories of language (for reviews, Cleary and Pisoni, 2001; Fowler, 2003; Diehl et al., 2004; Samuel, 2011). Within this overarching agenda, developmentalists have addressed how this mapping between acoustic and phonetic structures dynamically changes via early language experience in the first year of life (Werker and Curtin, 2005; Kuhl et al., 2008; Best et al., 2016). This emphasis on describing infant attunement to native speech derived in large part from experimental investigations by Kuhl and colleagues (Grieser and Kuhl, 1989; Kuhl, 1991; Kuhl et al., 1992; Iverson and Kuhl, 1995, 2000; Iverson et al., 2003). Their studies with human infants, human adults, and rhesus macaques revealed that early language experience profoundly alters speech perception by reducing discrimination sensitivity close to phonetic category prototypes and boosting sensitivity at the boundaries between categories (Lotto et al., 1998; Guenther et al., 1999, 2004; Feldman et al., 2009).

In more recent years, however, it has become increasingly clear that from infancy onward, speech processing involves generic as well as language-specific perceptual biases. It is now known that infants from across diverse linguistic communities initially display generic, "language-universal" biases or preferences in their perception of phonetic segments (Polka and Bohn, 2003, 2011; Nam and Polka, 2016). Moreover, these generic or "all-purpose" speech biases, which are distinct from "language-specific" prototype categorization processes, have been identified in adults (Masapollo et al., 2017a; Liu et al., 2021). These generic speech biases are evident in studies showing that young infants exhibit robust listening preferences for some speech sounds over others (Polka and Bohn, 2011; Nam and Polka, 2016), and that some phonetic contrasts are poorly distinguished early on (Polka et al., 2001; Best and McRoberts, 2003; Larraza et al., 2020) or show directional asymmetries in discrimination (Polka and Bohn, 2003, 2011; Kuhl et al., 2006; Pons et al., 2012; Nam and Polka, 2016).

The present research aims to improve our understanding of the neural mechanisms and processes underlying vowel perception biases observed in adults. It has been known for years that, early in development, infant perception is biased toward articulatorily and acoustically extreme vowels. These findings have been reviewed and discussed extensively by Polka and Bohn (2003, 2011), and have also been reinforced in recent meta-analyses (Tsuji and Cristia, 2017; Polka et al., 2019). Evidence supporting this view initially emerged from research revealing that infants show robust directional asymmetries in vowel discrimination tasks. More specifically, infants perform better at discriminating a change from a relatively less peripheral to a relatively more peripheral vowel within F_1–F_2 acoustic space, regardless of the language from which the contrast is drawn. As an example, Bohn and Polka (2001) used the head-turn conditioning procedure to test German-learning infants' discrimination of the German/i/-/e/vowel contrast (Werker et al., 1998). In this task, infants hear a repeating background stimulus and are assessed on their ability to distinguish a change from the background to a target stimulus. In the Bohn and Polka study, they counterbalanced presentation of each vowel; half of infant subjects were tested with one direction of change (from /i/ to /e/) and half were tested with a change in the reverse direction (from /e/ to /i/). The results revealed that infants performed better at discriminating the change from /i/ to /e/, compared to the reverse. Similar directional effects have been found with infants tested using numerous behavioral tasks and a wide range of vowel contrasts (Polka and Bohn, 2003, 2011). By adulthood, linguistic experience has altered vowel perception; similar asymmetries are observed for other non—native contrasts but not for native contrasts which are typically perceived with near-perfect accuracy (Polka and Bohn, 2011; Dufour et al., 2013; Tyler et al., 2014).

Over the last decade, Polka and colleagues have formulated and experimentally tested a theoretical framework, termed the Natural Referent Vowel (NRV) framework, for explicating the processes underlying directional asymmetries (Polka and Bohn, 2011; Masapollo et al., 2017a, 2018a; Polka et al., 2019). The NRV framework incorporates ideas across several existing phonetic theories, namely Steven's Quantal Theory (Stevens, 1989), and Schwartz's Dispersion-Focalization Theory (Schwartz and Escudier, 1989; Schwartz et al., 1997, 2005). Quantal Theory posits that vocalic articulations affiliated with the extremes of vowel space result in acoustic signals with obvious spectral prominences created by the convergence of adjacent formant frequencies. For example, when producing /i/ (the highest front vowel) F_2, F_3, and F_4 converge, when producing /y/ (the highest front rounded vowel) F_2 and F_3 converge, when producing /a/ (the lowest back vowel) and /u/ (the highest back vowel) F_1 and F_2 converge. These convergence points have also been referred to as "focal points" (Boë and Abry, 1986). According to the Dispersion-Focalization Theory, the strong tendency for vowel systems to select members found at the extremes of articulatory/acoustic vowel space is driven by two factors. First, dispersion ensures that vowels are acoustically distant from one another within vowel space, which enhances perceptual differentiation. Second, focalization ensures that vowels have salient and stable phonetic structures making them strong anchors for perception and production. Focal vowels will be easier for listeners to detect, encode, and retain in phonological working memory.

Concurring with these fundamental principles, the NRV framework (Polka and Bohn, 2011) offers additional insights into the aforementioned developmental findings by proposing that asymmetries in infant and adult vowel discrimination reflect a default, generic perceptual bias favoring focal vowels. In this account, the focalization of acoustic energy boosts perceptual salience, which in turn, biases perception and gives rise to the directional asymmetries observed in phonetic discrimination tasks (Schwartz and Escudier, 1989; Masapollo et al., 2017a). In advancing this viewpoint, Polka and Bohn do not mean to imply that perceptual asymmetries are attributable to low-level auditory or psychoacoustic processes. As highlighted in Masapollo et al. (2017b, 2018a), *NRV assumes that the effects of formant convergence on vowel perception reflect a phonetic bias that emerges when listeners are perceiving speech*, *rather than a low-level sensitivity to raw acoustic energy*. Compatible with this view, perception experiments have demonstrated that asymmetries predicted by differences in formant proximity are observed whether vowels are heard or perceived visually in a lip-reading task (Masapollo et al., 2017b, 2018a; Masapollo and Guenther, 2019), confirming that the "focal vowel" bias derives from phonetic processing rather than low-level psychoacoustic processes (Masapollo et al., 2019).

Polka and Bohn (2011) have further argued that the focal vowel bias plays an important role in the acquisition and processing of vowels across the lifespan. Asymmetries that point to a focalization bias are observed in infants in the first few months of life for both native and non-native vowel contrasts alike. Across the first year, as infants accrue specific linguistic experience, they begin tuning to native vowel contrasts. This will increase or diminish the initial focalization bias depending on the vowel inventory of their native language. This generic bias is thought to provide a scaffold to support the acquisition of a more detailed vowel system. Thus, according to NRV, both generic/focalization biases and language-specific biases influence vowel perception in mature, adult language users.

An alternative, but not mutually exclusive, account of asymmetries derives from Kuhl's Native Language Magnet (NLM) model (Kuhl, 1991; Iverson and Kuhl, 1995; Kuhl et al., 2008). This model, which combines principles from categorization and prototype theory (Rosch, 1975, 1977; Samuel, 1982) with statistical learning theory (Aslin and Newport, 2012), posits that directional asymmetries reflect biases favoring native language phonetic category prototypes (i.e., adult-defined "best" instances of a category). NLM assumes that phonetic categories emerge early in development as infants track distributional patterns in speech input during social interactions. Like other cognitive/perceptual categories, phonetic categories have an internal structure organized around a central, prototypic member (Kuhl, 1991). Furthermore, Kuhl claims that these prototypes have a "magnet-like" effect, which shrinks the immediate perceptual space making it more difficult to discriminate variants surrounding a prototype compared to variants surrounding a non-prototype of the same category (Kuhl, 1991; Kuhl et al., 1992; Iverson and Kuhl, 1995; cf. Miller and Eimas, 1996; Guenther et al., 1999, 2004; Feldman et al., 2009). Although NLM applies to both consonants and vowels, most of the research supporting the idea that there is a "warping" of within-category perceptual space that is tied to variation in category goodness has focused on vowels (Kuhl, 1991; Kuhl et al., 1992; Iverson and Kuhl, 1995; Lotto et al., 1998). Moreover, NLM posits that speech perception relies on general auditory mechanisms applied to acoustic rather than specifically phonetic information. Nevertheless, in the NLM model, directional asymmetries are viewed as an experience-dependent bias favoring native prototype; asymmetries arise because listener sensitivity is reduced when discriminating a change from a more-prototypic to a less-prototypic vowel compared to the reverse. In line with this view, Kuhl (1991) reported a directional asymmetry in which English-learning infants performed better at discriminating a change from a non-prototypic /i/ to a prototypic /i/, compared to a change from prototypic /i/ to non-prototypic /i/. Notably, in this case the prototypic /i/ was more focal (between F_2 and F_3) compared to the non-prototypic /i/. Thus, the observed asymmetry could be due to prototypicality and/or focalization effects.

Several English-French cross-linguistic studies assessed the competing NRV and NLM accounts of asymmetries in vowel perception (Masapollo et al., 2017a,b; Liu et al., 2021). The vowel /u/, as in *boo*, was chosen for use in these studies for several reasons. First prior research established that Canadian French speakers consistently produce more extreme /u/ gestures (resulting in lower and spectrally closer F_1 and F_2 values) than Canadian English speakers. Accordingly, in the standard vowel space, the mean location of French /u/ is more peripheral than that of English /u/ (Escudero and Polka, 2003; MacLeod et al., 2009; Noiray et al., 2011). This means that French /u/ has a more focal acoustic-phonetic form (with closer convergence of F_1 and F_2) compared to English /u/. These differences in focalization and language-specific phonetic categorization between English and French speakers provided an ideal opportunity to assess how these factors influence adult vowel perception.

In an initial study, Masapollo et al. (2017a) synthesized a set of vowels that were consistently identified as /u/ by native speakers of English and of French but that nevertheless varied in their stimulus goodness ratings, such that the best French /u/ exemplars were more focal (between F_1 and F_2) compared to the best English /u/ exemplars. In an AX (same/different) discrimination task, both English and French listeners were found to perform better at discriminating changes from the less to the more focal /u/ compared to the reverse, regardless of variation in prototypicality. Similar results were obtained using natural productions of English /u/ and French /u/ in tests with adults (Masapollo et al., 2017b) and infants (Polka et al., 2018). These findings established the focal vowel bias in adults, demonstrated that this perceptual bias favors vowels with greater formant convergence and established that this bias operates independently of biases related to language-specific prototype categorization.

In a subsequent study, Liu et al. (2021) presented Canadian English listeners with a finer grained series of vowels varying from the less-focal/English prototypic /u/ to the more-focal/French prototypic /u/ identified in the prior Masapollo et al. (2017a) study. In an AX discrimination task, the stimulus pairings included one-step, two-step, and three-step intervals along the series. The results revealed that focalization and prototype effects were both present but were differentially influenced by the size of the acoustic intervals along the stimulus series. More specifically, asymmetries favoring the English /u/ prototype emerged when subjects were discriminating small stimulus differences (1-step) close to the prototype stimulus. When stimulus differences were larger (2- or 3-steps) discrimination asymmetries favored more focal exemplars of /u/ (Masapollo et al., 2015). Collectively, these findings demonstrate, at the behavioral level, that directional asymmetries in adult vowel perception reveal a generic "focal vowel" bias that shapes the global structure of the vowel space (explained by NRV) as well as a more subtle experience-dependent bias that alters perception of the local internal structure of native vowel categories (as described by NLM).

Although the existing behavioral data indicate that directional asymmetries may be well predicted from a combination of salient spectral information and category "goodness" ratings, the neural underpinnings of these effects remain to be determined. Here, we present data from neurophysiological experiments with adults from different language backgrounds to begin to uncover these "brain-to-perception" relations and generate new hypotheses within the NRV and NLM theoretical frameworks. We wish to provide data that help to characterize what aspects of neural processing corroborate extant behavioral findings. Toward this end, we investigated whether we can observe asymmetries in the neurophysiological correlates of adult vowel perception, focusing on two neural measurements at the cortical level: (1) the mismatch negativity (MMN) that indexes neural sensitivity to vowel change; and (2) brain oscillatory activity in the theta (4–8 Hz) frequency band that indexes processing efficiency. While focalization biases have always been tested and demonstrated behaviorally by directional effects in discrimination tasks, examining the neural responses to vowels may provide us with a new window to understand vowel processing and the representation of "central" versus "peripheral" vowels in a more direct manner.

We recorded auditory event-related potentials (ERPs) and first computed the MMN response to within-category and cross-category vowel contrasts in native English- and French-speaking listeners. The vowel stimuli were previously used in an ERP study (Molnar et al., 2013) that compared vowel processing in bilingual and monolingual adults. The experimental design of this study also permitted an exploration of perceptual asymmetries, which is our present goal. Four stimuli were chosen from an acoustic vowel continuum (described below) ranging perceptually from /i/ to /y/ to /u/; the selected tokens include variants of /u/ and /y/ that form cross-category stimulus pairs (in French) and within-category stimulus pairs (in both languages). The psychophysical distances between the cross-category and within-category stimulus pairs were equated.

Prior studies examining the MMN in auditory oddball paradigms (Näätänen et al., 2007) have typically employed relatively short inter-stimulus-intervals (ISIs) (approximately 500 ms) in order to "build up" or "strengthen" the short-term memory "trace" for the repeated standard stimulus that develops online during the course of the experiment. The MMN is generally thought to reflect activity differences in neurons in or near the auditory cortex that detect a discrepancy (or mismatch) between the deviant percept and short-term trace of the standard (Näätänen et al., 2007). In tasks using relatively short ISIs, there will be less time for the short-term memory trace to decay between successive stimuli, and thus brain responses will reflect the basic resolution of the auditory system. Conversely, when the ISI is longer, the length of time that each stimulus is buffered in memory increases, short-term traces will decay, and brain responses will reflect stimulus encoding processing and long-term representations of phonological units (for discussion, see Strange, 2011). In the current research, we used a long ISI to better elicit a "phonetic mode" of processing and to restrict access to short-term memory traces. As previously discussed, NRV posits that focalization biases reflect phonetic processes rather than auditory processes in speech perception (Polka and Bohn, 2011; Masapollo et al., 2017a,b, 2018a,b). In keeping with this view, several previous behavioral studies have shown that in AX discrimination tasks, vowel order effects emerge or increase as the ISI increases, whereas overall perceptual performance improves and asymmetries decrease when the ISI decreases (Polka and Bohn, 2011; Masapollo et al., 2018b; Polka et al., 2019). For example, when testing adult discrimination of a non-native contrast, Polka and Bohn (2011) observed a directional asymmetry when they used a 1,500 ms ISI, but not when they used a 500 ms ISI.

On the basis of the aforementioned behavioral findings (Masapollo et al., 2017a; Liu et al., 2021), we generated several hypotheses concerning MMN responses measured using a long ISI. First, for relatively large (cross-category /u/ vs /y/) phonetic differences, we predicted (à la NRV) that the MMN will exhibit greater amplitude (and/or a shorter latency) in response to changes from less-focal to more-focal vowels compared to the reverse, but that these asymmetries will be weaker in French listeners because the /u/ - /y/ contrast is native in French but non-native in English. Second, for relatively small (within-category) phonetic differences, we hypothesized (à la NLM) that the MMN would exhibit greater amplitude (and/or a shorter latency) in response to changes from less-prototypic to more-prototypic vowels compared to the reverse. This hypothesis would be supported if the MMN showed opposite asymmetries across the two language groups. More specifically, English listeners would be expected to show a larger (and/or earlier) MMN when the English-prototypic /u/ occurs as a deviant among French-prototypic /u/ standards compared to the reverse, whereas French listeners would be expected to show a larger (and/or earlier) MMN when the French-prototypic /u/ occurs as a deviant among the English-prototypic /u/ standards compared to the reverse. Similarly, for the French group only, the MMN should be larger (and/or earlier) when the French-prototypic /y/ occurs as a deviant among the French-non-prototypic /y/ standards compared to the reverse. Yet another possibility is that directional asymmetries observed at the behavioral level may be reflected by ERP components with latency differences relative to the MMN. Because the MMN is thought to reflect "pre-attentive" processes, it may be too early of a cortical response to reflect asymmetries, which do not appear to derive from early stages of acoustic processing (Polka and Bohn, 2011; Masapollo et al., 2018b; Polka et al., 2019). Recently it has also been demonstrated that MMNs recorded in an oddball paradigm with a longer ISI (e.g., 600 vs. 2,600 ms) reflect sensitivity to language-specific phonological information rather than the acoustic information in speech sounds (Yu et al., 2017, 2018). An additional goal of the current study was to go beyond examination of the classic MMN response and track cortical oscillations. While the MMN response may seem like a more direct comparison with the existing behavioral discrimination findings, comparing the neural responses to the standard trials across the four vowels may provide us with a more direct look at how vowels with different acoustic characteristics are processed in the brain. We identified the theta band neural oscillation (4–8 Hz) to be a good measure to characterize vowel processing, as it has been argued to provide a measure of "neural efficiency" during speech processing (Zhang et al., 2011; Bosseler et al., 2013). We hypothesized that, if formant convergence influences attention or cognitive effort (à la NRV), then theta rhythms should show reduced power in response to more-focal compared to less-focal vowels. If, on the other hand, stimulus prototypicality influences cognitive effort (à la NLM), then theta rhythms should show reduced power in response to more prototypic native vowel exemplars compared to less prototypic ones.

MATERIALS AND METHODS

Participants

Thirty normal-hearing right-handed adults participated in the current experiment: 15 were native Canadian English speakers (average age = 25, seven females) and 15 were native Canadian French speakers (average age = 26 years, seven females). All were healthy young adults with no history of a speech, language, or other neurological impairment. Informed consent was obtained according to the McGill University human research committee. Four additional participants were tested but excluded from the analysis due to technical problems with the data acquisition (2) and poor data quality caused by artifacts (2). The EEG/ERP data for these participants was collected in a previous study (Molnar et al., 2013).

Participants' language background was assessed using two measures: (1) The *Language Experience and Proficiency Questionnaire* (LEAP-Q) which was developed specifically to evaluate bilingual and multilingual individuals' linguistic experience (Marian et al., 2007); and (2) a speech sample evaluated by monolingual speakers of Canadian English (n = 3) and Canadian French (n = 3) using a scale from 1 ("no ability in the given language") to 5 ("native-like ability").

Participants had to meet the following criteria to be included in the study: (1) no prior linguistic or phonetics training; (2) raised in a monolingual home and educated in a monolingual school in their respective language; (3) no experience learning a second language before 10 years of age; (4) no experience conversing in a second language on a regular basis, having rated their speaking and listening abilities in a second language with a maximum of 4 out of 10 on the LEAP-Q; and (5) their speech samples were rated 5 (native-like) on average.

Stimuli

In our previous behavioral studies (Molnar et al., 2010; Masapollo et al., 2017a), we synthesized a broad array of 128 vowels that covered the entire upper region of vowel space and ranged in F_1 (from 275 to 330 Hz) and F_2 (from 476 to 2,303 Hz) in equal psychophysical steps on the bark scale (Zwicker and Terhardt, 1980). All stimuli were synthesized using the Variable Linear Articulatory Model (Maeda, 1979, 1990; Boë, 1999; Ménard et al., 2004, 2009). The variants were created by manipulating the values of F_1 and F_2; the values of F_0, F_3, F_4, and F_5 remained constant for all vowels at 120, 2,522, 3,410, and 4,159 Hz, respectively. Each stimulus was 400 ms in duration and had the same intonation and intensity contours. In pilot studies, these stimuli were presented to native, monolingual Canadian English (n = 5) and Canadian French (n = 5) listeners, who were asked to give their phonetic identification and goodness ratings on a 5-point-scale (1 = very poor, 5 = very good). We found, as expected, that vowel judgments systematically varied as a function of F_2: For English listeners, the vowels varied perceptually from /u/("oo") to /i/("ee") as F_2 values increased, whereas for French listeners, the vowels varied perceptually from /u/("oo") to /y/ (as in the French word "but") to /i/ (ee) as the F_2 values increased (Note that in Canadian English, /y/ does not occur (Escudero and Polka, 2003; MacLeod et al., 2009).

Based on the results of these initial tests, we then selected 34 vowels to present to larger groups of English (n = 13) and French (n = 13) listeners in a subsequent experiment for identification and goodness ratings. This reduced stimulus set included 22 high back vowels targeting English /u/ and French /u/ vowel (F_1 = 275 and 300 Hz; F_2 = 4,548 to 979 Hz), and 12 high front vowels targeting English /i/ and French /i/ and /y/ (F_1 = 275 and 300 Hz; F_2 = 1,753 to 2,202 Hz). Note that we also synthesized two additional filler vowels (/o/["oh"] and /ə/["uh"]) to include in the stimulus set to provide some variation in vowel quality. This made it easier to assess whether participants were successful in identifying vowel quality differences using key words. The results of these tests were then used to select four vowel tokens (shown in **Figure 1**) for use in the current neurophysiological study: a good exemplar of French /u/ (F_1 = 275 Hz, F_2 = 745 Hz), a good exemplar of English /u/ (F_1 = 300 Hz, F_2 = 979 Hz), a good exemplar of French /y/ (F_1 = 275 Hz, F_2 = 2,011 Hz), and a poor exemplar of Canadian French /y/ (F_1 = 300 Hz, F_2 = 1,597 Hz). The selected variants of /u/ (V1 and V2) and /y/ (V3 and V4) were equally distant on the bark scale (Zwicker and Terhardt, 1980) along both the F_1 and F_2 dimensions. The French-prototypic/u/is the most focal between F_1 and F_2 (**Figure 1**, top panel), whereas the French-prototypic /y/ is the most focal between

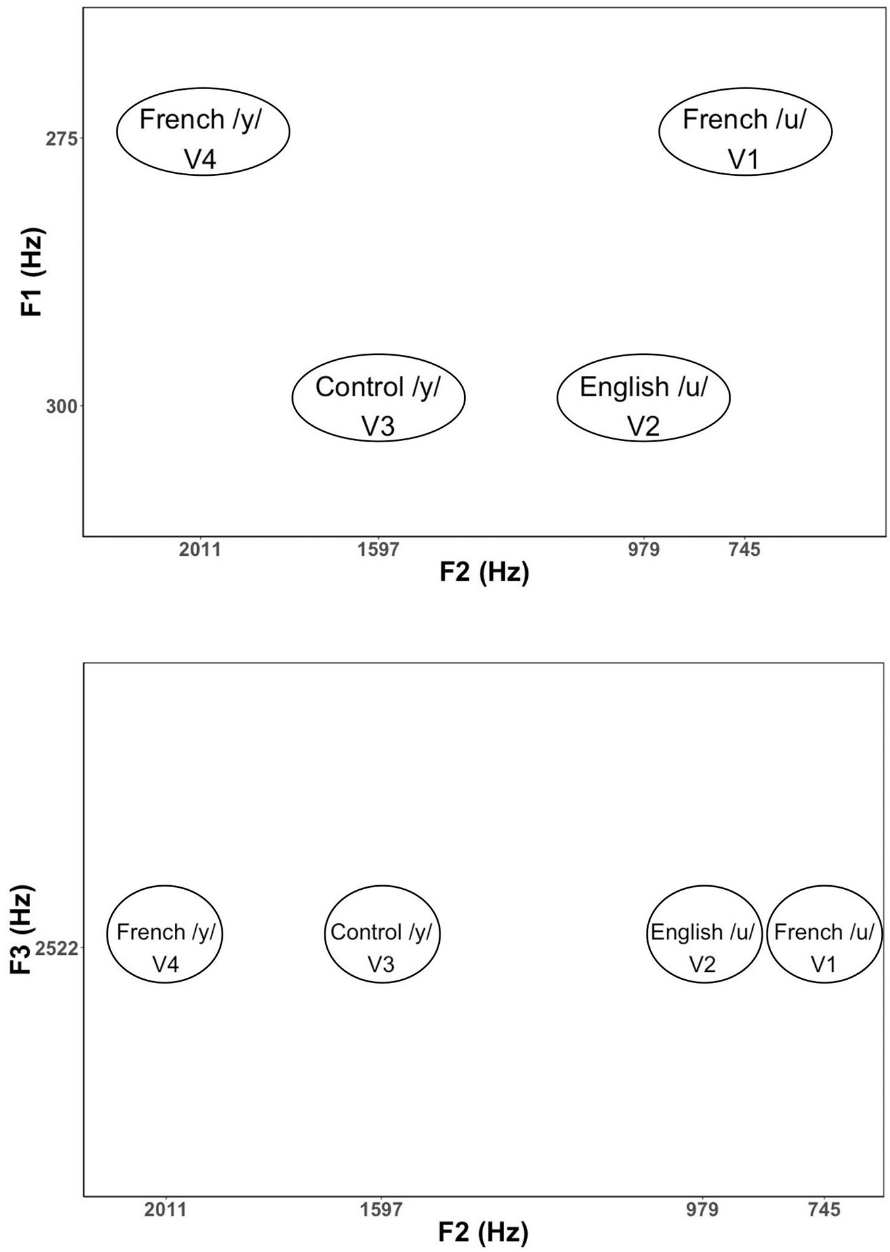


FIGURE 1 | Formant values for the vowel stimuli in F_1–F_2 and F_2–F_3 spaces. F_1 is related to the degree of constriction formed by the tongue in the vocal tract with lower F_1 values corresponding to a tighter constriction formed by a higher tongue position. F_2 is related to the location of the tongue constriction along the length of the vocal tract, with higher F_2 values corresponding to constrictions closer to the lips. Lip rounding (lip compression and protrusion) increases vocal tract length, which in turn, has the effect of lowering all formants, especially F_3.

F_2 and F_3 (**Figure 1**, bottom panel). **Table 1** gives the lower formants [F_1–F_3 (Hz)] and their corresponding bandwidths for each vowel stimulus. **Figure 2** schematizes the underlying perceptual vowel spaces for each language group [English (top) vs. French (bottom)].

Procedure and Design

Vowel perception was assessed with ERPs. The stimuli were presented across four different experimental blocks using a passive "multi-deviant oddball" task (Näätänen et al., 2004), as schematized in **Figure 3**. As shown in **Figures 3A,B**, different from the multi-feature paradigm, we also made sure there were at least two standards prior to each deviant. Further, rather than assigning the role of standard to one specific stimulus alone and the role of deviant to the other (remaining) stimuli of interest, as is typically done, all the vowels were presented both as standards and as deviants across the four different presentation blocks. This provided a way to control for potential

TABLE 1 | Formant frequency and bandwidth (Hz) for the lower formants (F1, F2, and F3) for each vowel stimulus.

Stimulus	F_1	F_2	F_3	B_1	B_2	B_3
French/u/(V1)	275	745	2522	85	30	35
English/u/(V2)	300	979	2522	85	30	35
Control/y/(V3)	300	1597	2522	85	30	35
French/y/(V4)	275	2011	2522	85	30	35

differences in the N1 and P2 components (which overlap with the MMN) due to physical differences among the evoking stimuli. Within each block (in **Figure 3B**), a standard vowel alternated with three deviant vowels that differed in their first and second formant frequencies. The sequences of the four blocks were counter-balanced across subjects and language groups (English vs. French). The deviant and standard ratio was roughly 20:80 (each block contained 1,000 stimuli; 790 standards, 210 deviants [70 of each deviant vowel token]), and the inter-stimulus interval (ISI) was 1,600 ms. Within each block, deviants and standards were presented in a pseudo-random order ensuring that at least two standards preceded each deviant. During the recording sessions, participants sat in a comfortable armchair in an electrically shielded sound-attenuated booth and watched a silent movie under the instruction to ignore the stimuli. The stimulus output intensity was 65 decibels in hearing level (dB HL) and delivered to both ears through insert earphones (Etymotic Research). The experimental sessions lasted approximately 3.5 h including preparation time (approximately 40 min) and breaks (approximately 30 min).

EEG Recording

EEG data were continuously recorded (500 Hz/32 bit sampling rate; Neuroscan Synamps2 amplifier) from 20 sites on the scalp with cap-mounted Ag-Ag/Cl electrodes (Electro-cap International, Inc., Eaton, OH, United States), based on the international 10–20 system of electrode placement: Fp1/2, F7/8, F3/4, T3/4, C3/4, T5/6, P3/4, O1/2, Fpz, Fz, Cz, Pz, and Oz. Eye movements and blinks were detected using electro-oculography (EOG). Vertical and horizontal EOG were recorded from bipolar electrodes placed above and below the left eye, and at the outer corner of each eye, respectively. All EEG electrodes were referenced against the right mastoid, and an electrode located between Fz and Fpz provided the ground. Electrode impedances were kept under 3 kOhm.

Data Preprocessing and Analysis

EEG data were analyzed using Brain Vision Analyzer software (Brain Products GmbH, Germany), including offline band-pass filtering (0.5–30 Hz) and artifact rejection with a ±50 microvolts (μV) deviation criterion at all channels except for Fp1 and Fp2, which were clearly more affected by eye movements than the rest of the channels. Consequently, Fp1 and Fp2 were excluded from any further analysis and data processing. Artifact rejection resulted in data loss within a range of 3.45 and 11.09% across participants. Note that analyses with other band-pass settings (0.4–40, 0.4–100 Hz) were also computed. They resulted in the same findings reported here, but data included more noise.

Event-related potentials were time-locked to vowel onset and were computed separately for the standard and deviant conditions of each vowel. Only the standard immediately preceding a deviant stimulus was included in the calculation of ERPs for standards in order to use the same number of stimuli in forming the standard and the deviant. The epochs were 850-ms long (-50 ms pre-stimulus and 800 ms post-stimulus onset) and were baseline corrected to the time period from 50 ms of pre-stimulus onset to 50 ms of post-stimulus activity. We opted for this baseline correction (instead of the typical −100 ms to 0) because there was a 50 ms (±4 ms) silence at the beginning of each sound file that we had realized once the experiment was completed. Future studies that wish to replicate our procedure should select time windows based on the actual stimulus onset, not the specific values reported here. **Figure 4** shows the obtained ERP responses to each vowel token (V1 vs. V2 vs. V3 vs. V4) when presented in the contextual role of standard versus deviant for each language group [English (left panels) vs. French (right panels)].

Mismatch Negativity Response Analyses

A directional asymmetry is essentially a context effect, i.e., a difference found when the same stimuli are presented in a different order (or context). In behavioral discrimination tasks, directional asymmetries are assessed by comparing outcome measures (e.g., accuracy) across different orders (AB vs. BA). With ERP recordings, we can track neural processing of the individual stimuli within a sequence, which allows us to examine order/context effects at a deeper level. This was optimized by the current study design which ensured that subjects were presented each vowel within a pair as the deviant and also as a standard. For example, for vowel pair V1–V2, we can ask whether the processing of V1 is different (faster, stronger) when it follows V2 (serving as the deviant) than when it proceeds V2 (serving as the standard). With this in mind, MMN waveforms were calculated by subtracting standard ERPs from deviant ERPs of the same vowel, which allows us to target the context effects. For example, in the present study, to characterize the neural processing of the V1–V2 vowel pair we first calculated two MMN responses. The first MMN indexes the processing of V1 (in context of V2); to do so we took the ERPs recorded when V1 was the deviant (and V2 was the standard) minus the ERPs recorded when V1 was the standard (and V2 was the deviant); the result is plotted in **Figure 3C** (purple line). The second (or reverse) MMN indexes the processing of V2 (in the context of V1); to do so we took the ERPs recorded when V2 was the deviant (and V1 was the standard) minus the ERPs recorded when V2 was the standard (and V1 was the deviant); the result is plotted in **Figure 3C** (turquoise line).

We then conducted two types of analyses to examine whether asymmetries emerged in the neurophysiological responses to each of the vowel pairs: (i) a hypothesis-based analysis focused to examine possible directional effects on MMN responses in three *a priori* time windows identified in Molnar et al. (2013), and (ii) an exploratory temporal clustering analysis to reveal additional

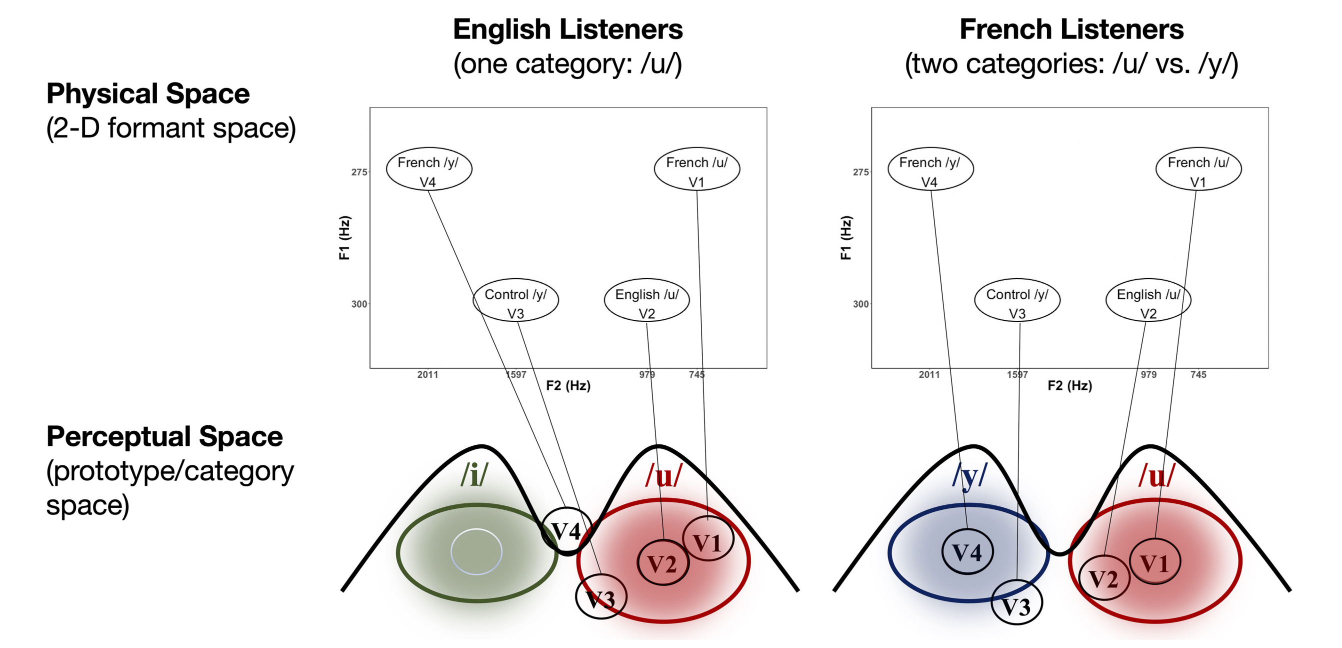


FIGURE 2 | Schematic of the hypothesized relationship between physical (2D F_1–F_2 formant) space and perceptual (prototype/category) space in the case of the current vowel stimuli for both English (*left*) and French (*right*) listeners (see text for explanation).

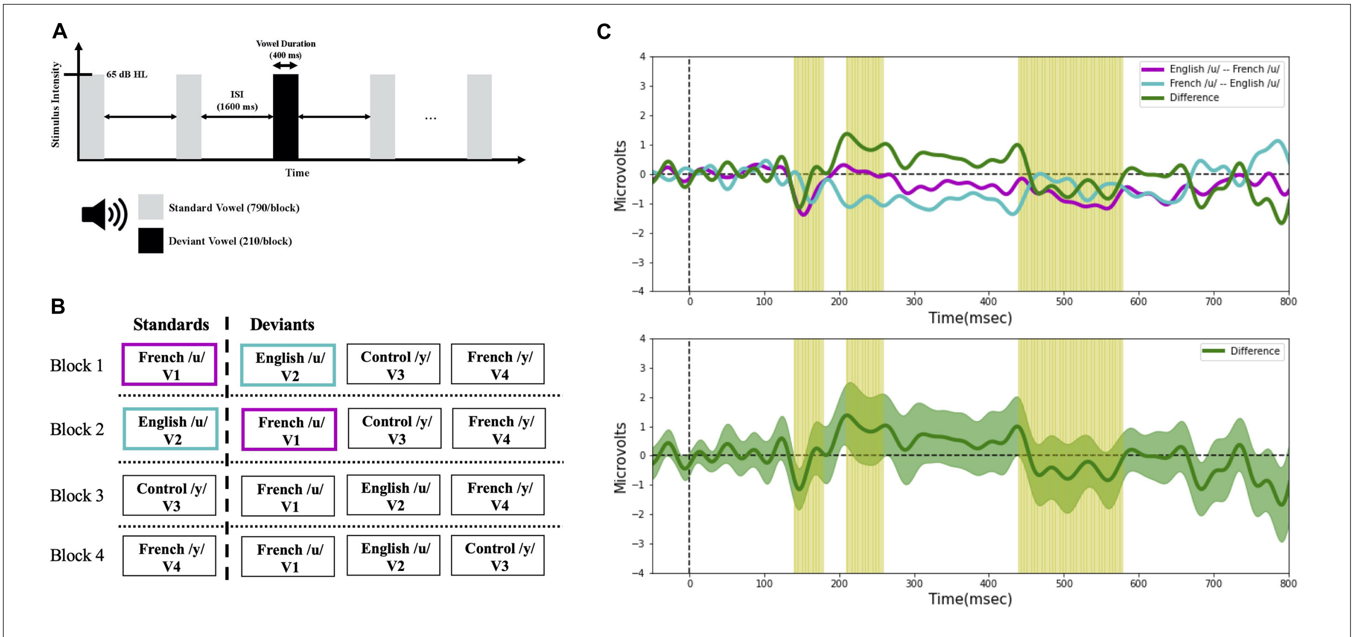


FIGURE 3 | (A) Timeline of the stimulus presentation in the oddball task. **(B)** Schematic of the experimental design, and **(C)** example MMN waveforms. Four experimental blocks were presented to each participant in a randomized order. The color coding in panel **(B,C)** illustrates the ERP averaging technique applied in the current study. For example, to characterize the detection of V1 among V2 (i.e., V2 → V1 direction), the MMN (purple line) was calculated by subtracting the ERPs to French /u/ (V1) as deviant measured in a block where the English /u/ (V2) was the standard, from the ERPs to French /u/ (V1) as standard in a separate block. Conversely, to characterize the detection of V2 among V1 (i.e., V1 → V2 direction), the MMN (turquoise line) was calculated by subtracting the ERPs to English /u/ (V2) as deviant measured in a block where the French /u/ (V1) was the standard, from the ERPs to English /u/ (V2) as standard in a separate block. In this way, we obtained an average ERP for the standards and for the deviants that was unaffected by the physical characteristics of the stimuli, and only the oddball effect was present when comparing the standard and deviant waveforms. The same averaging technique was applied with the rest of the contrasts. To test for directional asymmetries in the MMN, difference waves (green line) were then computed by subtracting the MMN waveform for the two opposite stimulus orders (e.g., V2 → V1 vs. V1 → V2); this is shown in the bottom panel of 3C with the green shaded region representing the standard error.

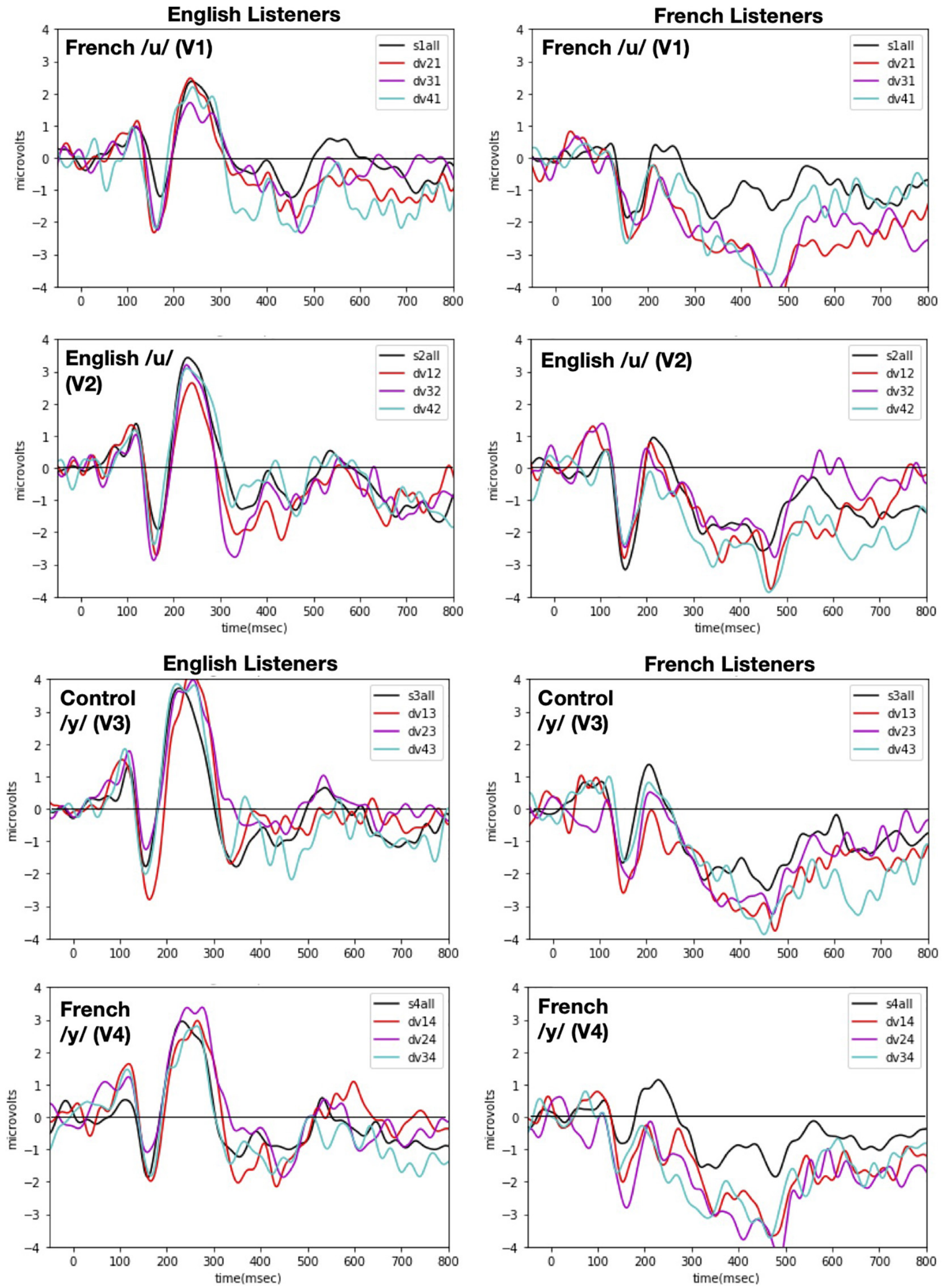


FIGURE 4 | ERP responses recorded with each of the four vowel tokens as the standard (V1, V2, V3, and V4 from top panel to bottom panel) and for each language group (English Listeners: *left panels* and French Listeners: *right panels*). Each panel shows ERP responses recorded when one vowel (indicated in the left hand corner) was presented as the standard with each of the other vowels as the deviant. As indicated in the right hand corner, the deviants are plotted as different color lines and as a black line when all deviants are combined. Note: in labeling the waves, the first number designates the deviant vowel and the second number designates the standard, e.g., for dv21 and V2 is the deviant when V1 is the standard.

time windows where MMN responses may be asymmetrical within the entire epoch (Maris and Oostenveld, 2007). The time windows for the hypothesis-based analysis were selected by the visual inspection of the grand average waveforms (that included all the conditions across all the participants) on the Fz electrode, and corresponding to time points associated with the N1, MMN, and the late negativity. The visual inspection yielded three consecutive latency time windows: 140–180 ms, 210–260 ms, and 440–580 ms. For each vowel pair (e.g., V1–V2) we then computed the average MMN (V1 in context of V2) and reverse MMN (V2 in context of V1) value within each of these latency windows for each participant. These values were submitted to separate analyses of variance (ANOVAs)–latency window (140–180 vs. 210–260 vs. 440–580 ms) × direction (MMN vs. reverse MMN)–for each vowel pair (V1–V2, V1–V3, V1–V4, V2–V3, V2–V4, and V3–V4).

An additional hypothesis-based analysis, was conducted based on the amplitude of the MMN *difference wave* computed by subtracting the MMN waveforms for each vowel within a pair; the amplitude difference was computed at each latency window (140–180 vs. 210–260 vs. 440–580 ms). For example, as shown for vowel pair V1–V2 in **Figure 3C**, the green line represents the difference between the MMN for V1 (MMN above) and the MMN for V2 (reverse MMN above); the green shaded area in the lower panel represents the standard error, and the yellow shading corresponds to the three latency windows. Within each language group, MMN difference waves were calculated for each of the six vowel pairs (V1–V2, V1–V3, V1–V4, V2–V3, V2–V4, and V3–V4) and averaged across six electrode sites (F3, F4, Fz, C3, C4, and Cz). Within each latency window, we averaged the values across the time window for each participant and then submitted these values to separate mixed ANOVAs–latency window (140–180 vs. 210–260 vs. 440–580 ms) × language group (English vs. French)–for each vowel pair.

Finally, the exploratory temporal clustering analysis was conducted to determine whether there were any additional latency windows (not tested in the aforementioned analysis) in which the MMN difference waves were significantly different from 0 (the value expected if there is no order/context effect) within the entire epoch. This was a data-driven approach with no *a priori* hypotheses with regard to the latency window(s) where the difference waveforms would significantly differ from 0 μV (see Maris and Oostenveld, 2007). Specifically, we deployed the threshold-free cluster enhancement (TFCE) extension method (Smith and Nichols, 2009) which allows for improved sensitivity, but more interpretable output than traditional cluster-based thresholding. First, the TFCE values were generated by summing across a series of thresholds, thus avoiding selecting an arbituary threshold and then the p values for each time sample were calculated through permutation. The analysis was performed using the TFCE cluster test with a start = 0, step = 0.01, and 3,000 permutations, implemented in MNE python software (Gramfort et al., 2014).

Time-Frequency Analysis

Finally, to characterize theta activity (4–8 Hz) during vowel processing, we conducted time-frequency analyses on the ERPs of the standard trials for each vowel, using the multi-taper method implemented in MNE python (Gramfort et al., 2014). Similar to the MMN analyses, the ERPs of the standard trials were also averaged across the F3, F4, Fz, C3, C4, and Cz electrode sites. The mean theta-band activity for each vowel and each participant was further extracted by averaging across the time window between 0 and 600 ms and across the frequency band between 4 and 8 Hz. Repeated ANOVAs and paired-sample *t*-tests were then performed on the individual means to test for effects of formant proximity and stimulus prototypicality on theta activity. Greenhouse-Geisser corrections were applied when appropriate and partial eta-squared (η_p^2) was calculated for main effects and interactions.

RESULTS

Mismatch Negativity Response Analyses

We assessed possible asymmetric patterns in the neurophysiological responses to all six vowel pairs within each language group and also cross-linguistically. Each of the six vowel pairs fell into one of three stimulus types: (1) *cross-category* pairs (V1–V4 and V2–V4) with relatively large acoustic differences; (2) *within-category* pairs (V1–V2 and V3–V4) with relatively small acoustic differences; and (3) *mixed-category* pairs (V1–V3 and V2–V3) with intermediate acoustic differences. **Figures 5–7** show the MMN results averaged across six electrode sites (F3, F4, Fz, C3, C4, and Cz) and plotted as a function of language group (English vs. French) and vowel contrast (V1–V2; V2–V3;V3–V4; V1–V4; V2–V4; and V1–V3), grouped by stimulus type (*cross-category* vs. *within-category* vs. *mixed-category;* the dark green shaded area represents the standard error, and the yellow shading corresponds to the three latency windows). The ANOVA results comparing the MMN waves in each direction for each of the three *a priori* latency windows are summarized in **Table 2**; none of the vowel pairs showed a main effect of direction or an interaction effect ($p > 0.05$). The ANOVA results comparing the MMN *difference waves* for each language group and for each of the three *a priori* latency windows are summarized in **Table 3**; none of the vowel pairs showed any main or interaction effects ($p > 0.05$). The temporal cluster analyses also failed to reveal asymmetric MMN response in other temporal windows, except for a small region of the MMN response to the V2–V3 vowel pair (described below).

Cross-category pairs (**Figure 5**). For the *cross-category* (/u-y/) vowel pairs (V1–V4 and V2–V4), recall the following NRV predictions: (1) for non-native (English) listeners, the MMN response will be greater (and/or earlier) when the more focal variants (V1 and V4) serve as the deviant stimulus, whereas (2) for native (French) listeners, vowel processing for *cross-category* pairs is predicted to be more symmetrical. No reliable asymmetry was found for either vowel pair in any of the pre-selected latency windows for either language group. No additional significant time windows were revealed by the exploratory temporal cluster analyses.

Within-category pairs (**Figure 6**). For the *within-category* pairs (V1–V2 and V3–V4), recall that NRV predicts that the MMN response will be greater (and/or earlier) when the more focal variants (V1 and V4) serve as the deviant stimulus compared to the reverse, regardless of native language. In contrast, NLM predicts that the MMN response will be greater (and/or earlier) when the more prototypical variant of a native vowel category serves as the deviant stimulus compared to the reverse. According

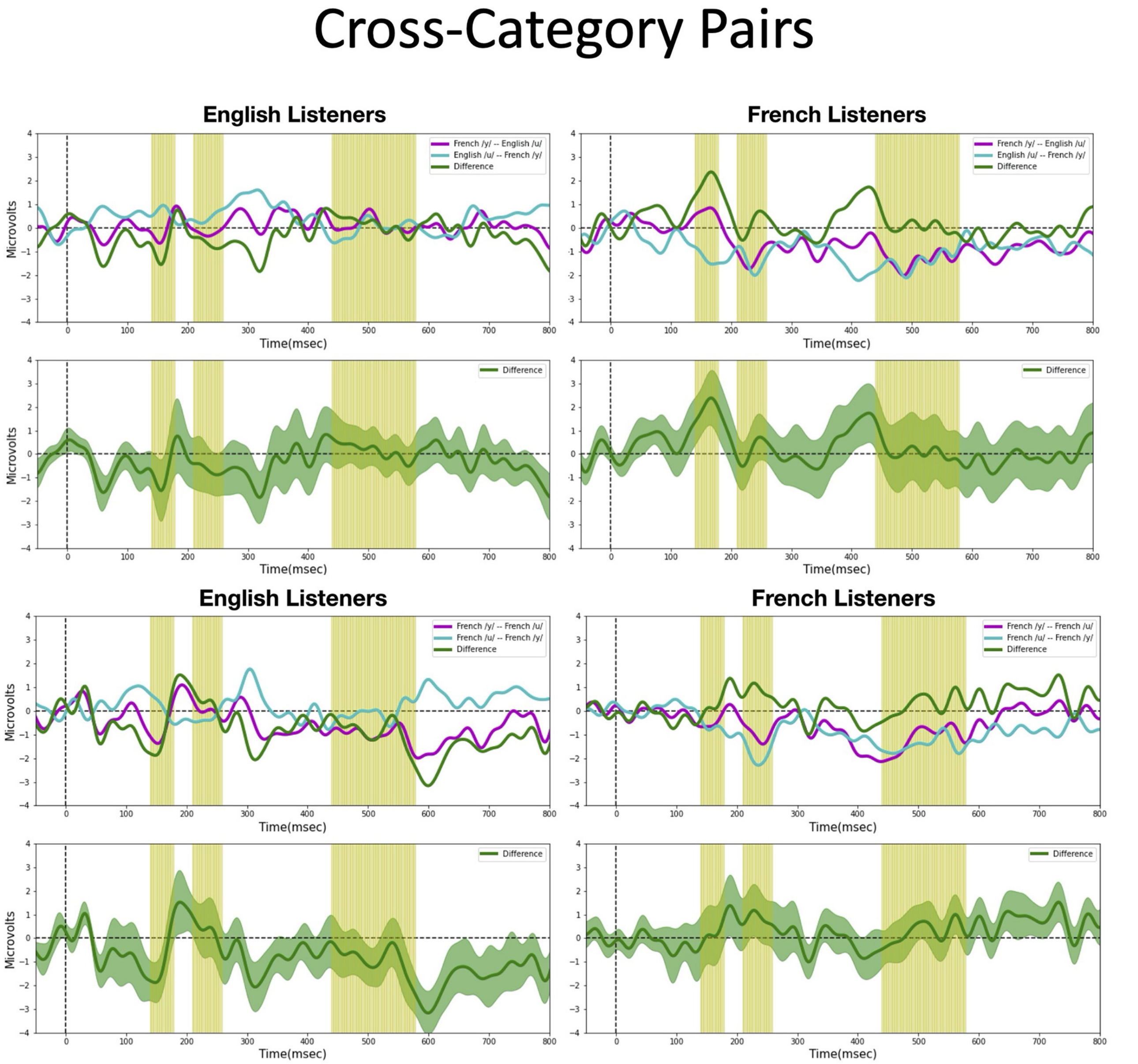


FIGURE 5 | Mismatch negativities (as described in **Figure 3C**) for each stimulus type: *Cross-category* pairs, the mean MMNs for each stimulus presentation order (purple and turquoise lines) and difference waveform (green lines) (in the top half of each panel) and the difference waveform plotted with the standard error for the group (as the dark green shaded region) in the bottom half of each panel. The light green shading indicates the pre-selected time windows used in the hypothesis-based anayses. The MMNs are plotted for each language group wth English Listeners on the left and French Listeners on the right. The vowel pair for each MMN is indicated in the legend at right hand corner of each panel with the standard followed by the deviant (e.g., French /y/ – French /u/ denotes an MMN with French /u/ (V1) as the deviant in the context of French /y/ (V4) as standard).

to this view, the MMN is expected to be stronger in the English listeners when V2 serves as the deviant among V1 standards, whereas for the French listeners, the MMN is expected to be stronger in the French listeners when V1 serves as the deviant among V2 standards and when V4 serves as the deviant among V3 standards. Although there was a trend for the MMN responses to pattern in the manner predicted by NLM across both language groups for the V1–V2 vowel pair, these differences did not reach statistical significance in any of the predetermined latency windows. However, for the V3–V4 pair in the 210–260 ms time window, the difference wave was significantly above 0 [$t(14) = 2.20$, $p = 0.04$], such that the MMN was stronger in the French group when V4 was the deviant compared to when V3 was the deviant). This finding is consistent with both NRV and NLM since the French prototypic /y/ (V4) is more focal (between F_2 and F_3) than the non-prototypic French /y/ (V3). No additional significant time windows were revealed by the cluster analyses.

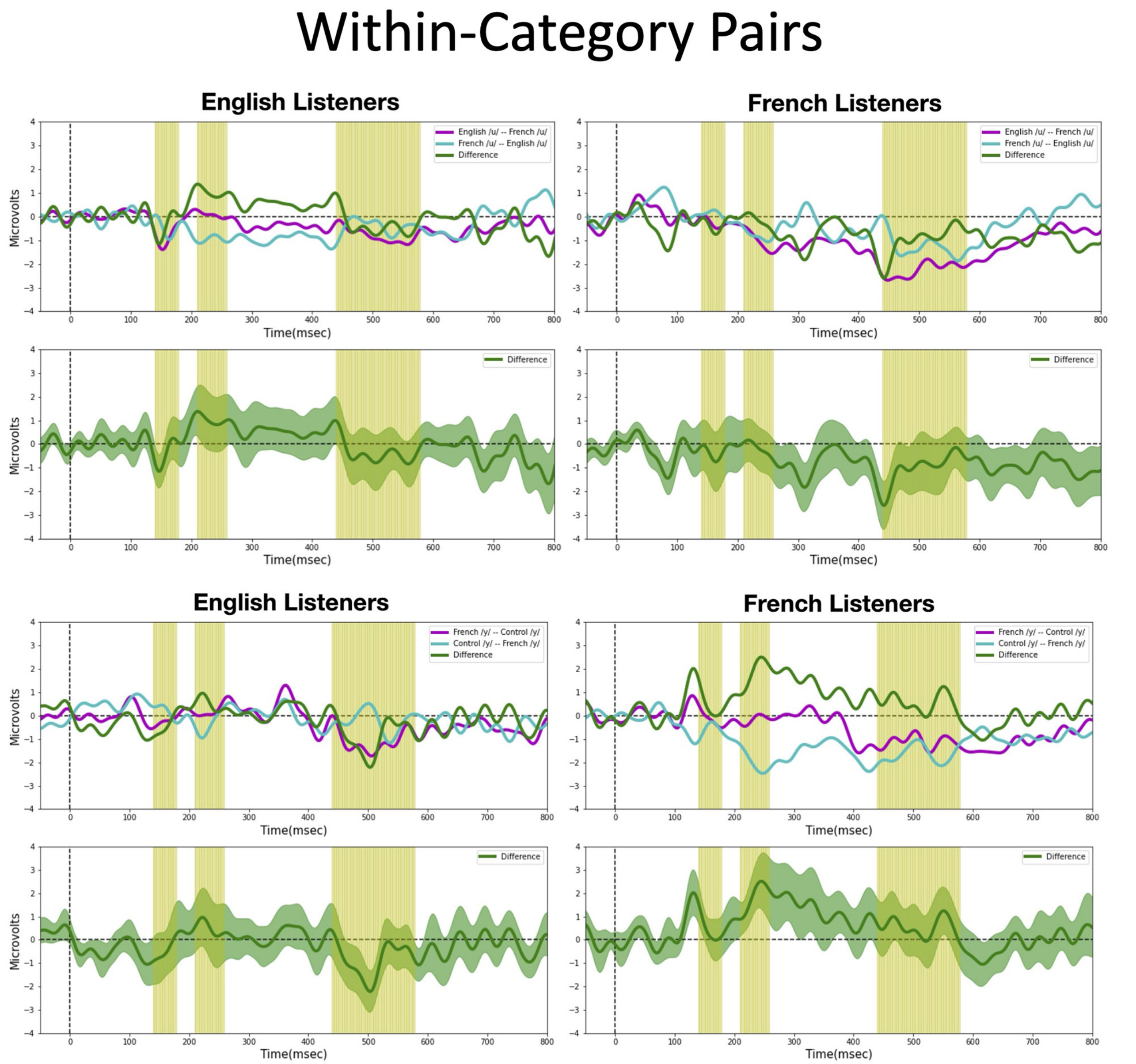


FIGURE 6 | Mismatch negativities (as described in **Figure 3C**) for each stimulus type: *Within-category pairs*, the mean MMNs for each stimulus presentation order (purple and turquoise lines) and difference waveform (green lines) (in the top half of each panel) and the difference waveform plotted with the standard error for the group (as the dark green shaded region) in the bottom half of each panel. The light green shading indicates the pre-selected time windows used in the hypothesis-based anayses. The MMNs are plotted for each language group wth English Listeners on the left and French Listeners on the right. The vowel pair for each MMN is indicated in the legend at right hand corner of each panel with the standard followed by the deviant (e.g., French /y/ – French /u/ denotes an MMN with French /u/ (V1) as the deviant in the context of French /y/ (V4) as standard).

Mixed-category pairs (**Figure 7**). For the *mixed-category* pairs (V2–V3 and V1–V3), NRV predicts that the MMN response will be greater when the more focal stimulus (V1 and V2) serves as the deviant compared to the reverse. NLM would predict that the MMN should be greater in the English group (in which V1, V2, and V3 are perceived as variants of the native /u/ category) when V2 serves as the deviant among V3 standards since V2 is the category prototype. Here, no significant asymmetries were found for either vowel contrast for either language group in the analyses using pre-selected time windows. However, the exploratory, temporal cluster permutation tests revealed an asymmetric MMN response between V2 and V3 or both English and French listeners but in opposite directions. For the English group, the time windows between 292–300 and 324–334 ms in the difference wave were significantly below 0 ($p < 0.05$), whereas for the French group, the time window in the difference wave between 66 and 104 ms was significantly above 0 ($p < 0.05$). These asymmetries do not align across language groups; for the English group, the

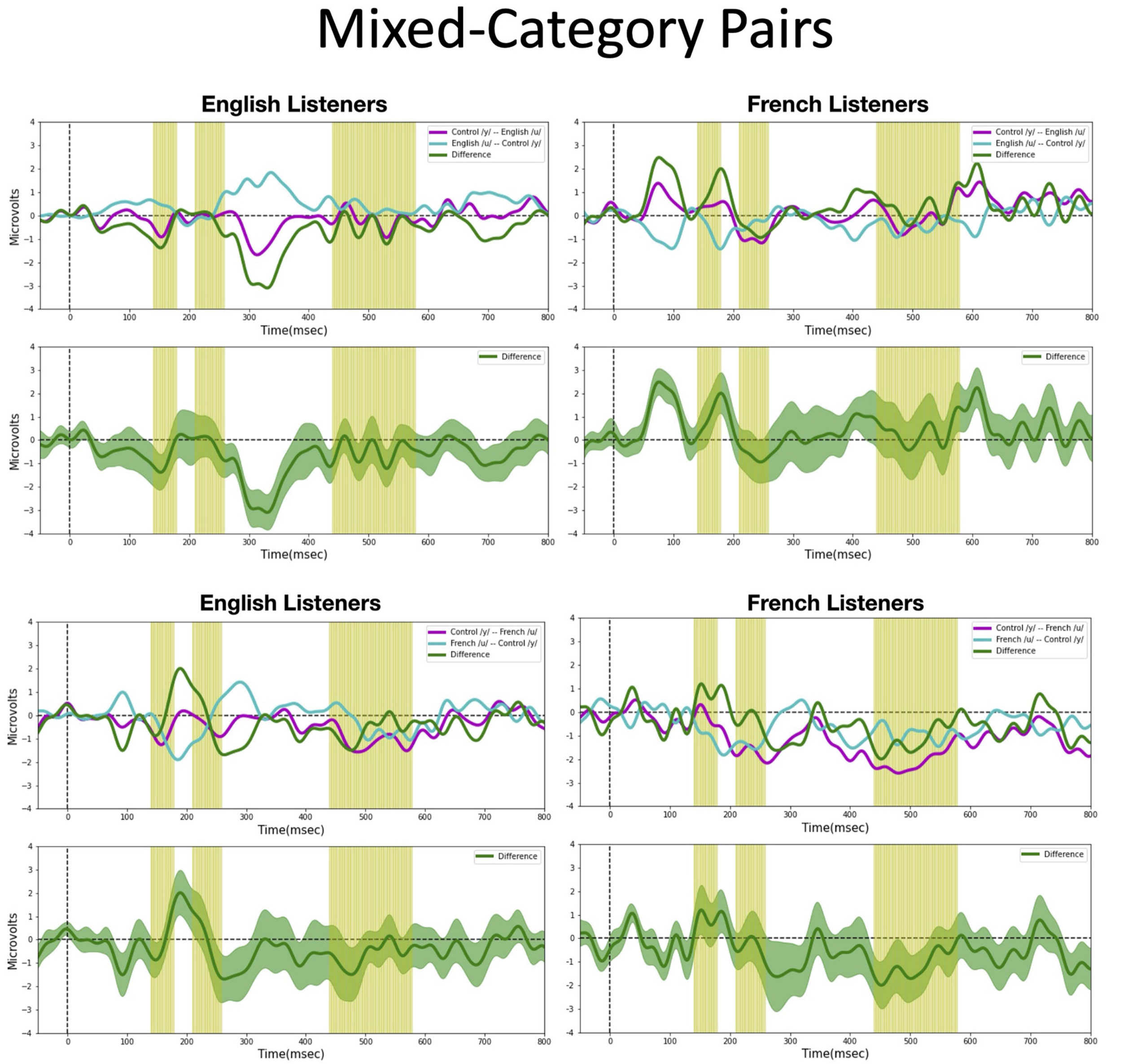


FIGURE 7 | Mismatch negativities (as described in **Figure 3C**) for each stimulus type: *Mixed-category pairs*. the mean MMNs for each stimulus presentation order (purple and turquoise lines) and difference waveform (green lines) (in the top half of each panel) and the difference waveform plotted with the standard error for the group (as the dark green shaded region) in the bottom half of each panel. The light green shading indicates the pre-selected time windows used in the hypothesis-based anayses. The MMNs are plotted for each language group wth English Listeners on the left and French Listeners on the right. The vowel pair for each MMN is indicated in the legend at right hand corner of each panel with the standard followed by the deviant (e.g., French /y/ – French /u/ denotes an MMN with French /u/ (V1) as the deviant in the context of French /y/ (V4) as standard).

MMN was stronger when the stimuli were presented in the direction going from V3 to V2, whereas for the French group, the MMN was stronger when the stimuli were presented in the direction going from V2 to V3. While the English results may be interpreted as a prototype effect, neither NRV nor NLM explicitly predicted the directional effect observed in the French group.

Overall, both the hypothesis-based and the exploratory analyses failed to provide evidence that the MMN responses to these vowel stimuli are asymmetric. The threshold-free cluster enhancement method that we applied is designed to isolate meaningful, non-random effects, through a data-driven approach. Given that the time windows tagged to have significant asymmetries using this method are quite short and also do not temporally align across the language groups these results should be interpreted with caution. These MMN findings alone do not provide sufficient evidence to draw conclusions about the neural processes that underlie the asymmetries in vowel processing that are observed in behavior.

TABLE 2 | Repeated measures an analysis of variance (ANOVA) on mismatch negativities (MMN) amplitude presented for each stimulus type (*cross-category* vs. *within-category* vs. *mixed-category*).

Effect	*F*	*df*	*p*	*np*^{2}
Cross-category pairs				
Vowel pair: more-focal French/u/(**V1**) vs. more-focal French/y/(**V4**)				
Latency window	1.497	2	0.233	0.049
Direction	0.003	1	0.954	<0.001
Latency window × Direction	1.588	2	0.214	0.052
Vowel pair: less-focal English/u/(**V2**) vs. more-focal French/y/(**V4**)				
Latency window	3.206	2	0.058	0.100
Direction	0.101	1	0.753	0.003
Latency window × Direction	0.773	2	0.461	0.026
Within-category pairs				
Vowel pair: more-focal French/u/(**V1**) vs. less-focal English/u/(**V2**)				
Latency window	3.297	2	0.054	0.102
Direction	0.191	1	0.665	0.007
Latency window × Direction	1.308	2	0.277	0.043
Vowel pair: less-focal/control/y/(**V3**) vs. more-focal/control/y/(**V4**)				
Latency window	6.092	2	0.007**	0.174
Direction	0.491	1	0.489	0.017
Latency window × Direction	2.326	2	0.112	0.074
Mixed-category pairs				
Vowel pair: less-focal English/u/(**V2**) vs. more-focal control French/y/(**V3**)				
Latency window	0.409	2	0.614	0.014
Direction	0.1	1	0.754	0.003
Latency window × Direction	0.414	2	0.635	0.014
Vowel pair: more-focal French/u/(**V1**) vs. more-focal control French/y/(**V3**)				
Latency window	1.239	2	0.296	0.041
Direction	0.345	1	0.561	0.012
Latency window × Direction	2.212	2	0.119	0.071

*Mismatch negativities for each vowel pair were analyzed using an ANOVA with the main factors of latency window (140–180 vs. 210–260 vs. 440–580 ms), direction (MMN vs. reverse MMN), and the interaction effect. Shown are the F value, the degrees of freedom, p value, and partial-eta-squared value for each effect, *p < 0.05; **p < 0.01.*

Theta Rhythms

Time-frequency plots showing brain oscillations (averaged across the F3, F4, Fz, C3, C4, and Cz electrode sites, 0–600 ms time window) for each vowel stimulus (V1 vs. V2 vs. V3 vs. V4) and language group (English vs. French) are given in **Figure 8**. We tested two specific hypotheses that derive from NRV and NLM, respectively: (1) mean theta activity, which has been argued to reflect attention and processing efficiency during speech processing (Bosseler et al., 2013), will be lower for the relatively more-focal vowel stimuli (V1 and V4) compared to the relatively less-focal vowel stimuli (V2 and V3); and (2) mean theta activity will be lower for the more-prototypical native-language vowel stimuli compared to the less-prototypical vowel stimuli. We first examined whether mean theta-band activity was influenced by formant proximity, independent of variation in stimulus prototypicality, in a mixed ANOVA with language group (English vs. French) as a between-subjects factor and vowel type [more-focal (V1–V4) vs. less-focal (V2–V3)] as a within-subjects factor. **Figure 9A** shows the theta rhythm results for the less-focal versus the more-focal vowels for each language group. Consistent with the predictions derived from NRV, there was a main effect of vowel type [$F(1,28) = 5.786$, $p = 0.023$, $\eta_p^2 = 0.171$], such that theta activity was lower for the more-focal vowels compared to the less-focal vowels. Neither the effect of language group [$F(1,28) = 2.855$, $p = 0.102$, $\eta_p^2 = 0.093$] nor the interaction effect [$F(1,28) = 0.676$, $p = 0.418$, $\eta_p^2 = 0.024$] reached statistical significance.

Next, to test whether mean theta-band activity was influenced by stimulus prototypically, we conducted a mixed ANOVA with language group (English vs. French) as a between-subjects factor

TABLE 3 | Mixed ANOVA on MMN Amplitude difference (MMN minus reversed MMN) for each stimulus type (*cross-category* vs. *within-category* vs. *mixed-category*).

Effect	*F*	*df*	*p*	*np*^{2}
Cross-category pairs				
Vowel pair: more-focal French/u/(**V1**) vs. more-focal French/y/(**V4**)				
Latency window	1.542	2	0.224	0.052
Language group	0.718	1	0.404	0.025
Latency window × Group	0.148	2	0.856	0.005
Vowel pair: less-focal English/u/(**V2**) vs. more-focal French/y/(**V4**)				
Latency window	0.802	2	0.451	0.028
Language group	1.308	1	0.262	0.045
Latency window × Group	2.067	2	0.137	0.069
With in-category pairs				
Vowel pair: more-focal French/u/(**V1**) vs. less-focal English/u/(**V2**)				
Latency window	1.279	2	0.285	0.044
Language group	0.354	1	0.557	0.012
Latency window × Group	0.349	2	0.683	0.012
Vowel pair: less-focal/control/y/(**V3**) vs. more-focal/control/y/(**V4**)				
Latency window	2.253	2	0.12	0.074
Language group	1.554	1	0.223	0.053
Latency window × Group	0.094	2	0.894	0.003
Mixed-category pairs				
Vowel pair: less-focal English/u/(**V2**) vs. more-focal control French/y/(**V3**)				
Latency window	0.431	2	0.618	0.015
Language group	0.823	1	0.372	0.029
Latency window × Group	2.176	2	0.132	0.072
Vowel pair: more-focal French/u/(**V1**) vs. more-focal control French/y/(**V3**)				
Latency window	2.185	2	0.122	0.072
Language group	0.002	1	0.965	<0.001
Latency window × Group	0.652	2	0.523	0.023

Mismatch negativity difference scores for each vowel pair were analyzed using ANOVA with the main factors of latency window (140–180 vs. 210–260 vs. 440–580 ms), language group (English vs. French), and the interaction effect. Shown are the F value, the degrees of freedom, p value, and partial-eta-squared value for each effect.

and vowel type [more-focal (V1) vs. less-focal (V2)] as a within-subjects factor. **Figure 9B** shows the theta rhythm results for the more-focal /u/ (V1) versus the less-focal /u/ (V2) for each language group. These results of the analysis revealed that there was a main effect of vowel type [$F(1,28) = 6.871$, $p = 0.014$, $\eta_p^2 = 0.197$], such that theta activity was lower for the more-focal variant ($M = 1.85$; $SD = 0.29$) compared to the less-focal variant ($M = 1.98$; $SD = 0.32$). Neither the effect of language group [$F(1,28) = 2.684$, $p = 0.113$, $\eta_p^2 = 0.087$] nor the interaction effect [$F(1,28) = 0.040$, $p = 0.844$, $\eta_p^2 = 0.001$] reached statistical significance. This finding aligns with the prior analysis showing that theta activity is lower during the processing of more-focal vowel stimuli, regardless of variation in category "goodness." It is important to note that because the magnitude of this difference was not greater for the French group compared to the English group, it further demonstrates that the aligned effects of focalization and stimulus prototypicality are not greater than effects of focalization alone. In a final analysis, we compared theta activity for the less-focal, non-prototypic /y/ (V3) vs more-focal prototypic /y/ (V4) for French listeners only because this is not a within-vowel category difference for English listeners. These results are shown in **Figure 9C**. Here, theta activity did not significantly differ between V3 and V4 [$t(14) = 1.052$, $p = 0.310$], indicating that the aligned effects of focalization and prototypicality is weak for this vowel pair.

DISCUSSION

In the current research we asked the following question: can we observe directional asymmetries in the neurophysiological correlates of vowel processing, and, if so, can these effects be attributable to processing differences related to generic phonetic biases, as predicted by the NRV framework (Polka and Bohn,

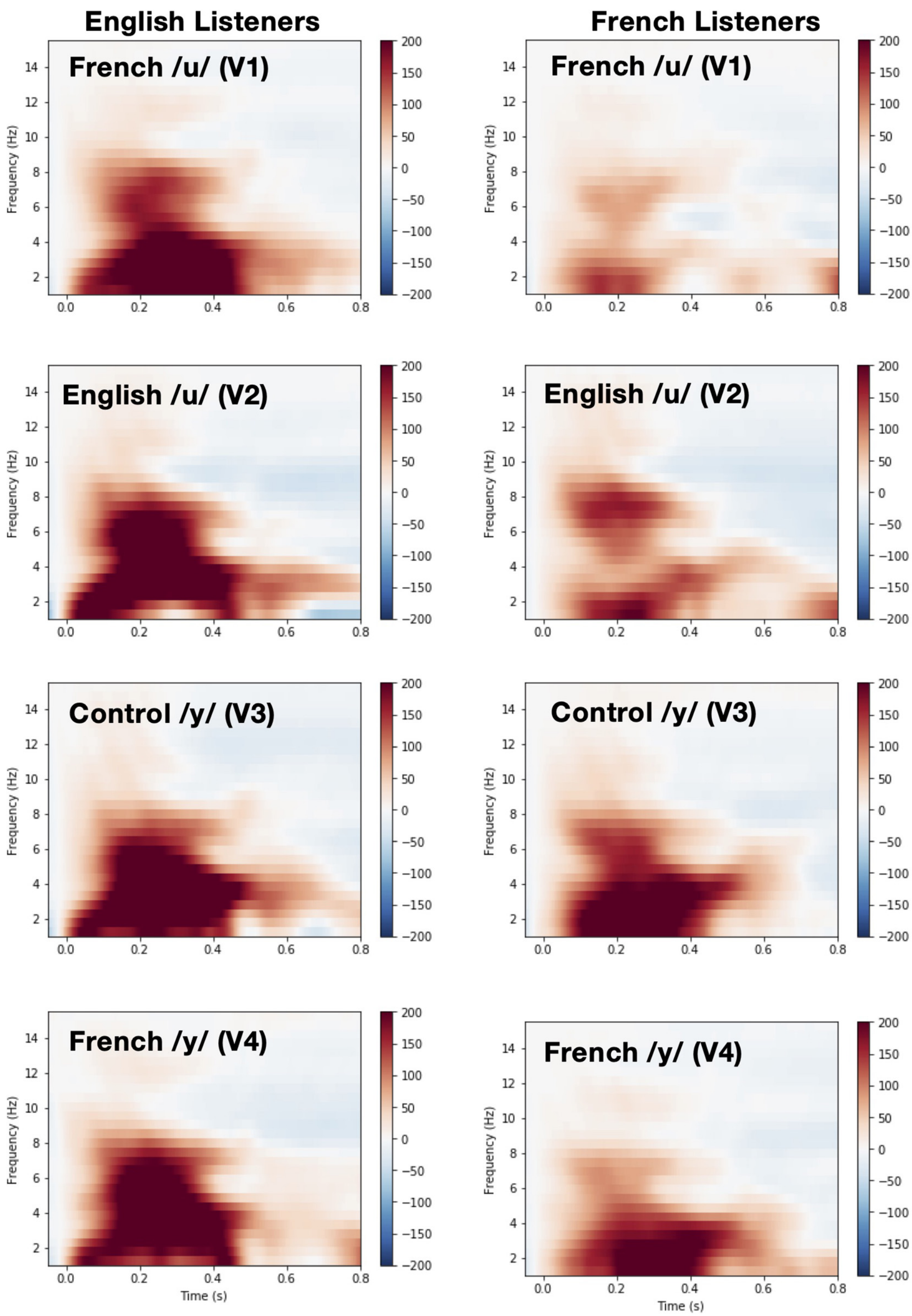


FIGURE 8 | Time-Frequency plots showing brain oscillations for each vowel during pre-deviant standard trials (averaged across the F3, F4, Fz, C3, C4, and Cz electrode sites, 0–600 ms time window and between 4 and 8 Hz) and for each language group [English (*left panel*) vs. French (*right panel*)].

2011), and/or stimulus prototypicality, as predicted by the NLM model (Kuhl, 1991; Kuhl et al., 2008)? To address this, we focused on the MMN and theta brain rhythms elicited in response to cross-category and within-category vowel pairs by English- and French-speaking adults. Recent behavioral research using very similar /u/ stimuli (Masapollo et al., 2015, 2017a; Liu et al., 2021) has shown directional asymmetries that are consistent with NRV when acoustic differences are relatively large, and that follow

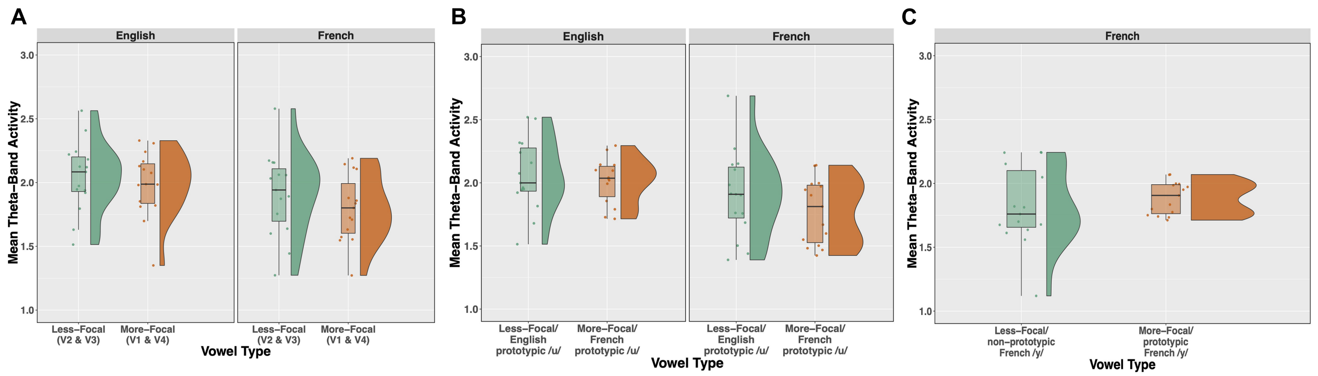


FIGURE 9 | Raincloud plots, which combine boxplots and split-half violin plots, of theta-band activity elicited in response to the vowel stimuli. **(A)** Theta activity for the relatively less-focal vowels (V2 and V3) versus more-focal vowels (V1 and V4) plotted as a function of each language group (English vs. French). **(B)** Theta for the more-focal /u/ (V1) versus less-focal /u/ (V2) for each language group. **(C)** Theta for the less-focal, non-prototypic /y/ (V3) vs. more-focal prototypic /y/ (V4) for French listeners. There was a main effect of vowel type in panels **(A,B)**, such that theta was lower for the more-focal vowel variants, and this effect was significant within each language group (see text for further explanation).

NLM predictions when acoustic differences are relatively small and very close to the location of a native category prototype in psychophysical space. The present study extends this work by showing that the pre-attentive MMN response may not be a reliable neurophysiological correlate of the asymmetric behavioral vowel discrimination. Rather, the data suggest that asymmetries, as revealed by theta oscillatory activity, may reflect differences in neural processing efficiency for vowels with varying degrees of formant convergence.

One possible explanation for the lack of robust directional differences in the MMN responses (shown in **Figures 5–7** and **Tables 2**, **3**) is that the present neurophysiological testing procedures require less cognitive demands than the prior behavioral paradigms (Masapollo et al., 2015, 2017a; Liu et al., 2021). More specifically, the ERPs were elicited by *passive* listening to the vowel stimuli; participants were not required to actively attend to the stimuli or make overt behavioral responses. In fact, they were distracted by a silent movie and instructed to ignore the auditory stimuli. Thus, the failure to observe asymmetric MMN responses could be due to the minimal processing demands in the ERP task in comparison to behavioral discrimination tasks where perceptual asymmetries are observed. For example, the AX discrimination task used by Masapollo et al. (2017a) required participants to attend to subtle acoustic differences in the vowel stimuli, to encode and buffer this information across a 1,500 ms delay, and then arrive at a "same" or "different" judgment within a brief time interval.

Although the MMN response failed to reveal asymmetries, the neural efficiency data suggests that cognitive demands are decreased when listeners process relatively more focal compared to relatively less focal vowels. Across both language groups, theta band activity was lower in response to the more focal variants but not to the more prototypic variants (**Figure 9**). This is consistent with theoretical accounts that vowels with a greater degree of formant convergence are easier to process and are more stable in phonological working memory compared to less focal vowels (Schwartz and Escudier, 1989; Schwartz et al., 1997, 2005; Polka and Bohn, 2011; Masapollo et al., 2017a).

The results also revealed a trend (albeit non-significant) for theta-band activity to be lower for the French listeners than the English listeners (**Figure 9**). These cross-linguistic processing differences align with behavioral data showing that French listeners are more sensitive to acoustic variations in this part of vowel space (Masapollo et al., 2017a,b). This may be attributable to differences in the vowel systems of English and French. More specifically, French has a richer inventory of high vowels (/i y u/) than English (/i u/; see, Escudero and Polka, 2003). Thus, this high region of the vowel space is denser in French than in English, which may explain French listeners' enhanced sensitivity to spectral differences in this part of the articulatory/acoustic vowel space. An alternative, but not mutually exclusive, explanation is that among our French subjects, who are functionally monolingual, at least some had experienced passive exposure to spoken English, and also heard English-accented French, while growing up and living in Quebec. In comparison, our English participants were university students from outside of Quebec studying in Montreal,; they are less likely to have gained passive experience listening to French. Nevertheless, the current theta-band activity results raise the possibility that the roots of these cross-linguistic differences in behavioral discrimination lie in more efficient use of underlying neural mechanisms.

The present neurophysiological study builds on prior behavioral studies that seek to understand universal and experience-dependent factors that interact to shape vowel perception across the lifespan. Our findings add to an existing body of research that has employed the MMN or other neural measures to assess asymmetric patterns in speech processing. Unlike the present study, much of the prior MMN work related to asymmetries was motivated and designed to assess theoretical perspectives on phonological processing, typically focusing on evaluating models that posit different feature-based approaches such as abstract/under-specification versus detailed/full specification views (Eulitz and Lahiri, 2004; Cornell et al., 2011, 2013; Scharinger et al., 2012; Hestvik and Durvasula., 2016; Hojlund et al., 2019). For example, Eulitz and Lahiri (2004) examined German adults' discrimination of several *native* vowel contrasts (i.e., /e/-/o/, /ø/-/o/, /e/-/ø/) in a passive oddball task, using event-related potentials. Not-surprisingly, listeners showed mismatch negativity (MMN) responses to all of the contrasts. However, the results also indicated that for the /ø/-/o/ contrast, listeners showed larger and earlier MMN amplitudes when /ø/ was the deviant, compared to when /o/ was the deviant. Listeners did not show asymmetric MMN responses for either of the other two vowel contrasts. The authors interpret these findings as suggesting that /ø/-/o/ has a different *phonological* status than the other vowel contrasts tested, and that this difference in phonological representation might explain the neural processing differences. Specifically, they postulate that the place of articulation feature [coronal] is universally absent or "underspecified" from phonemic representations in the lexicon–for vowels and consonants alike. In this view, the listeners tested in their oddball task may have elicited a larger and earlier MMN response when /ø/ was presented as a deviant in a train of /o/ standards, compared to the reverse, because /ø/ has an underspecified [coronal] place of articulation. These results raise the possibility that the structure of phonological representations for vowels in the lexicon must be considered along with formant proximity (Polka and Bohn, 2011) to account for what is currently known about native-language vowel processing in adults.

In general, though, prior research on asymmetries in speech processing has focused more on consonants rather than vowels and has not always included behavioral findings, making it unclear how to interpret some MMN findings. More importantly, prior MMN studies investigating asymmetric patterns in vowel processing have typically not considered specific perceptual or phonetic biases. Accordingly, vowel contrasts/stimuli used often do not allow for clear predictions à la NRV or NLM and the acoustic phonetic details (e.g., F_3 values) needed to consider NRV predictions are often not reported. Some methodological choices also make it difficult or impossible to connect reported MMN findings with predictions based on more perceptually oriented models like NRV or NLM. For example, most prior MMN studies present stimuli using ISI (inter-stimulus interval) values that are very short compared to the values used in perception studies; this is problematic given that an NRV bias is not expected when the ISI is short because this promotes low-level acoustic biases rather than phonetic encoding of vowels (Masapollo et al., 2018b).

Similarly, the present study was also not motivated or designed to assess conceptual views on phonological processing. We choose (sub-lexical) stimuli and task conditions that allow us to make clear predictions about the role of focalization and prototypicality in neural speech processing, but these choices do not permit clear predictions about competing feature-based phonological models, such as the FUL (Featurally Underspecified Lexicon) model (Eulitz and Lahiri, 2004). Importantly, although our work has a somewhat narrower focus, we do not assume that the phonetic biases hypothesized within NRV or NLM are incompatible with phonological processing models, including more abstract views like FUL that propose under-specification of phonological representations. Rather, we expect that phonetic biases must eventually align with and support mature phonological processing. Explaining how this developmentally unfolds is a critical goal of future research. Achieving this goal will require us to design studies that assess divergent theoretical perspectives in an integrated rather than a parallel fashion. Going forward, it will also be important to examine and control relevant low-level acoustic parameters and task parameters. This is needed to clarify when we are measuring simple order or context effects that are possible procedural artifacts and when we are measuring aspects of stimulus salience that also operate under natural speech processing conditions.

We clearly need further research into the relations between behavioral and neural levels of vowel processing in cross-linguistic studies with adults as well as infants early in development. Several recent ERP studies (Slabu et al., 2012; Zhao et al., 2019) focusing on *subcortical* auditory processing of speech sounds in adults have found evidence to corroborate some aspects of the existing behavioral data on directional asymmetries (Masapollo et al., 2017a). In the Zhao et al. study, English listeners were presented with resynthesized (shortened) versions of the less-focal/English-prototypic /u/ and more-focal/French-prototypic /u/ identified by Masapollo et al. (2017a) in a passive listening task. The researchers recorded the frequency-following response (FFR) with the two stimuli arranged in oddball and reversed-oddball blocks. It is generally assumed that the FFR reflects the encoding of acoustic energy in the fundamental frequency (F_0) and lower harmonics of vowel stimuli (Krishnan, 2002; Skoe and Kraus, 2010; Bidelman et al., 2013; Bidelman, 2018). Accordingly, Zhao and colleagues used the FFR as a way to assess whether there was more robust neural encoding in the frequency range of the F_1 region for the more focal versus the more prototypical variant of the native vowel category/u/. They found that English listeners show enhanced power at the frequencies corresponding to F_1 when listening to the more-focal/French prototypic /u/, but only when it served as the deviant stimulus. However, this pattern was not found in the neural encoding of F_1 in response to the less-focal/English prototypic/u/; for the less focal /u/ the neural encoding was comparable regardless of whether the vowel served as the standard or deviant stimulus. These findings suggest that focality impacts the neural encoding of vowels.

While Zhao et al.'s (2019) results revealed an intriguing parallel between subcortical ERP and behavioral measures of auditory vowel processing, the precise nature of the relations between these measures remain unclear in part because they were obtained under very different task demands and stimulus presentation speeds. The behavioral perceptual tasks utilized by Masapollo et al. (2017a) and Liu et al. (2021) were more demanding in terms of attention and memory than the passive oddball task used by Zhao et al. (2019). In the AX discrimination tasks, listeners were required to make overt similarity judgements about pairs of stimuli separated by relatively long inter-stimulus intervals (ISIs) (i.e., 1,500 ms), whereas in the ERP task, listeners attended to a silent video while passively listening to trains of vowel stimuli separated by very short ISIs (i.e., ∼50 ms).

In future work, it will be informative to explore how task and stimulus factors modulate neurophysiological responses to vowel stimuli. If asymmetries evident in behavior are supported by pre-attentive processes, this may be more clearly observed in the standard passive task using a more typical short ISI. If, however, asymmetries are boosted by active and attentive processing of phonetic structure in the speech signal, as posited by NRV, then asymmetric patterns in brain activity should be more prominent during active processing tasks. Given that vowel perception asymmetries have been observed during unimodal (visual-only) as well as bimodal (audio-visual) vowel processing (Masapollo et al., 2017b, 2018a; Masapollo and Guenther, 2019), it will also be informative to record and compare neurophysiological responses in adults using visual-only and auditory-visual vowel stimuli. This could bring new insights into the brain networks linked to the processing of speech signals as phonetic units.

Apart from those differences in task demands, it is also unclear whether the subcortical ERP measures themselves directly relate to asymmetries in vowel perception documented at the behavioral level. Although the functionality and origins of the FFR are a topic of active debate, there is mounting evidence that the FFR arises from activity generated by both cortical and subcortical structures along the auditory pathway, and reflects pre-attentive neural tracking of sustained periodic information; namely, the fundamental frequency and higher harmonics in vowels (Skoe et al., 2013; Coffey et al., 2016; Tichko and Skoe, 2017; Bidelman, 2018). However, even if we assume this to be the case, we are still left with the question of whether and how such components from deep in the brainstem relate to cortical levels of processing and perception (for recent discussion, Zhao and Kuhl, 2018).

Finally, although the preponderance of evidence suggests that asymmetries in vowel perception derive from cognitive encoding strategies involving attention and working memory rather than general auditory processes (Polka and Bohn, 2011; Masapollo et al., 2017b, 2018b), analogous effects have also been reported with non-speech tonal analogues of vowels that approximate some of the temporal characteristics of naturally-produced /u/ vowels executed with more versus less extreme lip gestures (Masapollo et al., 2019). While such findings may be interpreted as evidence that asymmetries reflect (at least in part) general auditory processing biases, it is also possible that they reflect fundamentally different types of processes than those captured using speech stimuli (see also, Bishop et al., 2005; Timm et al., 2011). Future studies could directly compare subcortical and cortical responses to vowels and non-speech tones that track the center of the formant paths of the vowels. Such analyses will help to further uncover how adult and infant brains process vowels.

DATA AVAILABILITY STATEMENT

The data and analysis code supporting the conclusions of this article is available at: https://osf.io/5ubsh/?view_only=08560bc5fdac4d7f82af7418a6820d7d.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by McGill University, Faculty of Medicine Institutional review board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MMo, TCZ, and MMa analyzed the data. LP, TCZ, and MMa wrote the manuscript. LP and MMo designed the ERP experiment. MMo conducted the ERP experiment. All authors contributed to the article and approved the submitted version.

FUNDING

The research reported here was supported by two grants from the Natural Sciences and Engineering Research Council of Canada, one to LP (RGPIN-2018-05823), and another to MMo (RGPIN-2019-06523) and by a grant from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health to TCZ (Award Number R21NS114343). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

We are grateful to Rachael Carson and Nathalie Belanger at McGill University for their assistance with subject recruitment and data collection, and Lucie Ménard for her assistance with the preparation of the speech stimuli.

REFERENCES

Aslin, R. N., and Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Curr. Direct. Psychol. Sci.* 21, 170–176. doi: 10.1177/0963721412436806

Best, C. T., Goldstein, L. M., Nam, H., and Tyler, M. D. (2016). Articulating what infants attune to in native speech. *Ecol. Psychol.* 28, 216–261. doi: 10.1080/10407413.2016.1230372

Best, C. T., and McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Lang. Speech* 46(Pt 2–3), 183–216. doi: 10.1177/00238309030460020701

Bidelman, G. M., Moreno, S., and Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212. doi: 10.1016/j.neuroimage.2013.04.093

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69.

Bishop, D. V. M., O-Reilly, J., and McArthur, G. M. (2005). Electrophysiological evidence implicates automatic low-level feature detectors in perceptual asymmetry. *Cogn. Brain Res.* 2005, 177–179. doi: 10.1016/j.cogbrainres.2004.12.007

Boë, L.-J. (1999). "Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: consequences for ontogenesis and phylogenesis," in *Proceedings of the International Congress of Phonetic Sciences*, 2501–2504.

Boë, L.-J., and Abry, C. (1986). Nomogrammes et systemes vocaliques, Actes des 15 e 'mes journés d' étude sur la parole. *Sociéte Francaise d'Acoustique* 1986, 303–306.

Bohn, O.-S., and Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *J. Acoust. Soc. Am.* 110, 504–515. doi: 10.1121/1.1380415

Bosseler, A., Taulu, S., Pihko, E., Makela, J. P., Imada, T., Ahonen, A., et al. (2013). Theta brain rhythms index perceptual narrowing in infant speech perception. *Front. Psychol.* 4, 1–12. doi: 10.3389/fpsyg.2013.00690

Cleary, M., and Pisoni, D. B. (2001). "Speech perception and spoken word recognition: Research and theory," in *Blackwell handbook of sensation and perception*, ed. E. B. Goldstein (Malden, MA: Blackwell), 499–534. doi: 10.1002/9780470753477.ch16

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., and Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Comm.* 7, 1–11. doi: 10.1007/978-3-319-62657-4_32-1

Cornell, S. A., Lahiri, A., and Eulitz, C. (2011). What you encode is not necessarily what you store: evidence for sparse feature representations from mismatch negativity. *Brain Res.* 1394, 79–89. doi: 10.106/j.brainres.2011.04.001

Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Ann. Rev. Psychol.* 55, 149–179.

Dufour, S., Brunelliere, A., and Nguyen, N. (2013). To what extent do we hear phonemic contrasts in a non-native regional variety? Tracking the dynamics of perceptual processing with EEG. *J. Psycholinguist. Res.* 42, 161–173. doi: 10.1007/s10936-012-9212-8

Escudero, P., and Polka, L. (2003). "A cross-language study of vowel categorization and vowel acoustics," in *Proceedings of the International Congress of Phonetic Sciences*, eds M. J. Sole, D. Recansens, and J. Romero (Barcelona), 861–864.

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.116/089892904323057308

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752–782. doi: 10.1037/a0017196

Fowler, C. A. (2003). "Speech production and perception," in *Handbook of psychology Experimental psychology*, eds A. F. Healy, R. W. Proctor, and I. B. Weiner (Hoboken, NJ: Wiley), 237–266.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Grieser, D., and Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Dev. Psychol.* 25, 577–588. doi: 10.1037/0012-1649.25.4.577

Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *J. Acoustical Soc. Am.* 106, 2900–2912. doi: 10.1121/1.428112

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hearing Res.* 47, 46–57. doi: 10.1044/1092-4388(2004/005)

Hestvik, A., and Durvasula. (2016). Neurobiological evidence for voicing underspecificaton in English. *Brain Lang.* 152, 28–43. doi: 10.1016/j.bandl.2015.10.007

Hojlund, A., Gebauer, L., McGregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. *Lang. Cogn. Neurosci.* 34, 545–560. doi: 10.1080/23273798.2019.1572204

Iverson, P., and Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* 97, 553–562. doi: 10.1121/1.412280

Iverson, P., and Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Percept. Psychophys.* 62, 874–886. doi: 10.3758/bf03206929

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.

Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hear. Res.* 166, 192–201. doi: 10.1016/s0378-5955(02)00327-1

Kuhl, P. K. (1991). Human adults and human infants show a "Perceptual Magnet Effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, F13–F21.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosoph. Transac. R Soc. B* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Larraza, S., Molnar, M., and Samuel, A. G. (2020). Phonemic contrasts under construction? Evidence from Basque. *Infancy* 25, 304–318. doi: 10.1111/infa.12330

Liu, Y. Y., Polka, L., Masapollo, M., and Ménard, L. (2021). Disentangling the roles of formant proximity and stimulus prototypicality in adult vowel perception. *J. Acoust. Soc. Am. Exp. Lett.* 1:003041.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *J. Acoust. Soc. Am.* 103, 3648–3655. doi: 10.1121/1.423087

MacLeod, A., Stoel-Gammon, C., and Wassink, A. B. (2009). Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *J. Phonet.* 37, 374–387. doi: 10.1016/j.wocn.2009.07.001

Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *J. Acoust. Soc. Am.* 65:S22.

Maeda, S. (1990). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, eds W. L. Hardcastle and A. Marshal (Dordrecht: Kluwer Academic), 131–149. doi: 10.1007/978-94-009-2037-8_6

Marian, V., Blumfeld, H. K., and Kaushanskaya, M. (2007). Language experience and proficiency questionnaire (LEAP-Q). *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)

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

Masapollo, M., Polka, L., and Ménard, L. (2015). "Asymmetries in vowel perception: effects of formant convergence and category "goodness."," in *Proceedings of the 18th International Congress of Phonetic Sciences*, (Glasgow), 1–5.

Masapollo, M., Polka, L., Molnar, M., and Ménard, L. (2017a). Directional asymmetries reveal a universal bias in adult vowel perception. *J. Acoust. Soc. Am.* 141, 2857–2869. doi: 10.1121/1.4981006

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

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., and Morgan, J. L. (2018a). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation and configuration. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1103–1118. doi: 10.1037/xhp0000518

Masapollo, M., Franklin, L., Morgan, J. L., and Polka, L. (2018b). "Asymmetries in vowel perception arise from phonetic encoding strategies," in *Poster presented at the 175th Meeting of the Acoustical Society of America*, (Minneapolis, MN).

Masapollo, M., and Guenther, F. H. (2019). Engaging the articulators enhances perception of concordant visible speech movements. *J. Speech, Lang. Hear. Sci.* 62, 3679–3688. doi: 10.1044/2019_jslhr-s-19-0167

Masapollo, M., Zhao, T. C., Franklin, L., and Morgan, J. L. (2019). Asymmetric discrimination of non-speech tonal analogues of vowels. *J. Exp. Psychol.* 45, 285–300. doi: 10.1037/xhp0000603

Ménard, L., Schwartz, J.-L., and Boë, L.-J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *J. Speech Lang. Hear. Sci.* 47, 1059–1080. doi: 10.1044/1092-4388(2004/079)

Miller, J. L., and Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Percept. Psychophys.* 38, 1157–1167. doi: 10.3758/bf03207549

Ménard, L., Davis, B., Boë, L.-J., and Roy, J.-P. (2009). Producing American-English during vocal tract growth: A perceptual categorization study of synthesized vowels. *J. Speech Lang. Hear. Sci.* 52, 1268–1285. doi: 10.1044/1092-4388(2009/08-0008)

Molnar, M., Polka, L., Baum, S., and Steinhauer, K. (2010). "Vowel perception: How simultaneous bilinguals do it," in *Poster presented at the Neurobilingualism Conference*, (Donostia).

Molnar, M., Polka, L., Baum, S., and Steinhauer, K. (2013). Learning two languages from birth shapes pre-attentive processing of vowel categories: electrophysiological correlates of vowel discrimination in monolingual and simultaneous bilinguals. *Lang. Cogn.* 2013, 1–16.

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

Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115, 140–144. doi: 10.1016/j.clinph.2003.04.001

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

Noiray, A., Cathiard, M.-A., Meìnard, L., and Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *J. Acoust. Soc. Am.* 129, 340–349. doi: 10.1121/1.3518452

Polka, L., Colantonio, C., and Sundara, M. (2001). A cross-language comparison of /d/-/th/ perception: evidence for a new developmental pattern. *J. Acoust. Soc. Am.* 109, 2190–2201. doi: 10.1121/1.1362689

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Comm.* 41, 221–231. doi: 10.1016/s0167-6393(02)00105-x

Polka, L., and Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *J. Phonet.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Polka, L., Masapollo, M., Noriega, P., and Liu, Y. Y. (2018). "Directional asymmetries reveal a universal bias in infant auditory vowel perception," in *Proceedings of the 16th Meeting of the Association for Laboratory Phonology*, (Lisbon).

Polka, L., Ruan, Y., and Masapollo, M. (2019). Understanding vowel perception biases: a meta-analytic approach. *Sound Appr. Lang. Matter.* 2019, 561–582.

Pons, F., Albareda-Castellot, B., and Sebastian-Gallés, N. (2012). The interplay between input and initial biases: Asymmetries in vowel perception during the first year of life. *Child Dev.* 83, 965–976. doi: 10.1111/j.1467-8624.2012.01740.x

Rosch, E. (1975). Cognitive reference points. *Cogn. Psychol.* 7, 532–547. doi: 10.1016/0010-0285(75)90021-3

Rosch, E. (1977). "Human categorization," in *Studies in cross-cultural psychology*, ed. N. Warren (London: Academic Press), 1–49.

Samuel, A. G. (1982). Phonetic prototypes. *Percept. Psychophys.* 31, 307–314. doi: 10.3758/bf03202653

Samuel, A. G. (2011). Speech perception. *Ann. Rev. Psychol.* 62, 49–72.

Scharinger, M., Bendixen, A., Trujilllo-Barreto, N. J., and Obleser, J. (2012). A sparse neural code for some speech sounds and not for others. *PLoS One* 7:340953. doi: 10.1371/journal.pone.0040953

Schwartz, J.-L., and Escudier, P. (1989). A strong evidence for the existence of a large- scale integrated spectral representation in vowel perception. *Speech Comm.* 8, 235–259. doi: 10.1016/0167-6393(89)90004-6

Schwartz, J.-L., Boë, L.-J., Valleìe, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *J. Phonet.* 25, 255–286. doi: 10.1006/jpho.1997.0043

Schwartz, J.-L., Abry, C., Boë, L.-J., Meìnard, L., and Valleìe, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalization Theory. *Speech Commun.* 45, 425–434. doi: 10.1016/j.specom.2004.12.001

Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/AUD.0b013e3181cdb272

Skoe, E., Krizman, J., Anderson, S., and Kraus, N. (2013). Stability and plasticity of auditory brainstem function across the lifespan. *Cereb. Cortex* 25, 1415–1426. doi: 10.1093/cercor/bht311

Slabu, L., Grimm, S., and Escera, C. (2012). Novelty detection in the human auditory brainstem. *J. Neurosci.* 32, 1447–1452. doi: 10.1523/jneurosci.2557-11.2012

Smith, S. M., and Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98. doi: 10.1016/j.neuroimage.2008.03.061

Stevens, K. N. (1989). On the quantal nature of speech. *J. Phonet.* 17, 3–46. doi: 10.1016/s0095-4470(19)31520-7

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *J. Phonet.* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001

Tichko, P., and Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hear. Res.* 348, 1–15. doi: 10.1016/j.heares.2017.01.014

Timm, J., Weise, A., Grimm, S., and Schronger, E. (2011). An asymmetry in the automatic detection of the presence or absence of a frequency modulation within a tone: a mismatch negativity study. *Front. Psychol.* 2:1–8. doi: 10.3389/fpsyg.2011.00189

Tsuji, S., and Cristia, A. (2017). Which acoustic and phonological factors shape infants' vowel discrimination? Exploiting natural variation in InPhonDB. *Proc. Interspeech* 2017, 2108–2112.

Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21. doi: 10.1159/000356237

Werker, J. F., Sji, R., Desjardins, R., Pegg, J. E., Polka, L., and Patterson, M. (1998). "Three methods for testing infant speech perception," in *Perceptual development: Visual, auditory, and speech perception in infancy*, ed. A. M. Slater (London: UCL Press), 389–420.

Werker, J. F., and Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234. doi: 10.1080/15475441.2005.9684216

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Front. Neurosci.* 11:95. doi: 10.3389/fnins.2017.00095

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2018). The duration of auditory sensory memory for vowel processing: Neurophysiological and behavioral measures. *Front. Psychol.* 9:335. doi: 10.3389/fpsyg.2018.00335

Zhang, Y., Koerner, T., Miller, S., et al. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Dev. Sci.* 14, 566–581. doi: 10.1111/j.1467-7687.2010.01004.x

Zhao, T. C., and Kuhl, P. K. (2018). Linguistic effect on speech perception observed at the brainstem. *Proc. Natl. Acad. Sci.* 115, 8716–8721. doi: 10.1073/pnas.1800186115

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., and 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 Lang.* 194, 77–83. doi: 10.1016/j.bandl.2019.05.002

Zwicker, E., and Terhardt, E. (1980). Analytical expressions for critical band rate and critical bandwidth as a function of frequency. *J. Acoust. Soc. Am.* 68, 1523–1525. doi: 10.1121/1.385079

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fnhum.2021.629648

Contrast and Conflict in Dutch Vowels

*Nadine P. W. D. de Rue[1], Tineke M. Snijders[2,3] and Paula Fikkert[1]**

[1] *Centre for Language Studies, Radboud University, Nijmegen, Netherlands,* [2] *Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands,* [3] *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands*

The nature of phonological representations has been extensively studied in phonology and psycholinguistics. While full specification is still the norm in psycholinguistic research, underspecified representations may better account for perceptual asymmetries. In this paper, we report on a mismatch negativity (MMN) study with Dutch listeners who took part in a passive oddball paradigm to investigate when the brain notices the difference between expected and observed vowels. In particular, we tested neural discrimination (indicating perceptual discrimination) of the tense mid vowel pairs /o/-/ø/ (place contrast), /e/-/ø/ (labiality or rounding contrast), and /e/-/o/ (place and labiality contrast). Our results show (a) a perceptual asymmetry for place in the /o/-/ø/ contrast, supporting underspecification of [CORONAL] and replicating earlier results for German, and (b) a perceptual asymmetry for labiality for the /e/-/ø/ contrast, which was not reported in the German study. A labial deviant [ø] (standard /e/) yielded a larger MMN than a deviant [e] (standard /ø/). No asymmetry was found for the two-feature contrast. This study partly replicates a similar MMN study on German vowels, and partly presents new findings indicating cross-linguistic differences. Although the vowel inventory of Dutch and German is to a large extent comparable, their (morpho)phonological systems are different, which is reflected in processing.

Keywords: perceptual asymmetry, vowels, Dutch, MMN, conflict, phonological contrasts

OPEN ACCESS

Edited by:
Valerie L. Shafer,
The City University of New York, United States

Reviewed by:
Linda Polka,
McGill University, Canada
David Jackson Morris,
University of Copenhagen, Denmark
Mirjam J. I. de Jonge,
University of Amsterdam, Netherlands

***Correspondence:**
Paula Fikkert
p.fikkert@let.ru.nl

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *15 November 2020*
Accepted: *05 May 2021*
Published: *07 June 2021*

Citation:
de Rue NPWD, Snijders TM and Fikkert P (2021) Contrast and Conflict in Dutch Vowels. Front. Hum. Neurosci. 15:629648. doi: 10.3389/fnhum.2021.629648

INTRODUCTION

There is considerable acoustic variation in natural speech, making recognition of spoken words or even single vowels rather complex. For word recognition, it is important to perceive meaningful differences (i.e., phonological features), which are contrastive and to some extent predictable. The phonological underlying representations of words, and indeed phonemes, in our mental lexicon are made up of phonological features that play a crucial role in recognition. A vital issue in phonology is what information these representations contain and how they enable us to recognize and produce words the way we do. Some theories assume rich phonetic detail to be part of stored representations (e.g., Johnson, 1997; Goldinger, 1998; Bybee, 2001; Pierrehumbert, 2002; Polka and Bohn, 2003, 2011; Masapollo et al., 2017a,b), while others only assume the essential features needed to differentiate between lexical contrasts (e.g., Chomsky and Halle, 1968; Archangeli, 1988; Lahiri and Reetz, 2002; Dresher, 2009). One model that makes very explicit assumptions as to which phonological features are stored in the underlying representations in the mental lexicon is the Featurally Underspecified Lexicon (FUL) model (e.g., Lahiri and Reetz, 2002, 2010; Lahiri, 2018).

FUL aims to define and regulate a set of features which can cover the typology of all possible contrasts and alternations in the languages of the world (Lahiri, 2018). In addition, it is able to account for acquisition and language processing. The model assumes privative phonological features (i.e., presence or absence of features), and furthermore assumes only contrastive features to be stored in mental representations. For the contrasts under investigation in this paper, the features [LABIAL], [CORONAL], and [DORSAL] – which reside under the ARTICULATOR node – are in focus. [CORONAL] is assumed to be the universal default, and hence is underspecified in the underlying representation. In FUL, features are identical for vowels and consonants. There are no feature dependencies, and features underneath each node are mutually exclusive, although in the current model of FUL [LABIAL] can combine with [CORONAL] and [DORSAL] for vowels, an assumption we will further address in the discussion. Furthermore, the model assumes that listeners extract features from the acoustic signal which are mapped onto the features of the underlying representations in the mental lexicon. These perceived surface features can *match*, *mismatch* or *no-mismatch* with the underlying features. Matches and mismatches require features to be detected from the input and present in the underlying representation. A no-mismatch occurs when an extracted surface feature is underspecified in the underlying representation (e.g., default). This ternary mapping procedure predicts specific asymmetries in a Mismatch Negativity (MNN) experiment. A mismatch is predicted to result in a larger neural discrimination difference between two sounds than a no-mismatch, reflected by an enhanced or earlier MMN (Näätänen, 2001), indicating perceptual discrimination.

In their seminal MMN study, Eulitz and Lahiri (2004) showed that German listeners perceived vowel contrast asymmetrically, and argued that this asymmetry is due to phonological underspecification of coronal place of articulation. While the acoustic difference between the mid vowels /e, ø/ on the one hand, and /o, ø/ on the other is similar, their phonological representations differ, as shown in (1): while /e, ø/ are underspecified for place of articulation in the underlying representation, indicated by [—], /o/ is specified as [DORSAL]. Furthermore, /o, ø/ both are specified for [LABIAL]. However, the underspecified feature [CORONAL] can still be extracted from the acoustic signal, and thus be part of the surface representation, just like [LABIAL] and [DORSAL]. When in an oddball paradigm the standard (indicated by //) is presented, the underlying representation of the standard is pre-activated. Upon hearing a deviant (indicated by []), the surface representation of the deviant is mapped onto the pre-activated underlying representation of the standard. When the standard is /o/, the coronal place of articulation of the deviant [ø] will lead to a mismatch (marked with a red line), but the reverse will not, as the perceived [DORSAL] feature of [o] will form a no-mismatch with the underspecified [CORONAL] in the underlying representation of /ø/, as shown in **Figure 1**.

Based on these perceptual asymmetries, Eulitz and Lahiri (2004) argue that coronal is underspecified in German. These perceptual asymmetries of place of articulation (in particular [CORONAL] and [DORSAL]) have also been found in various

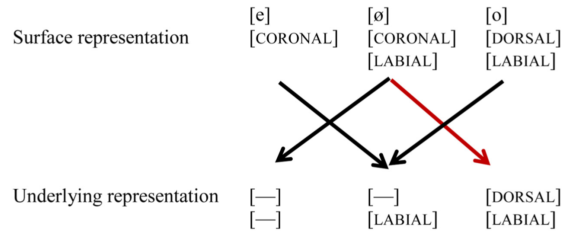


FIGURE 1 | Black lines indicate a no-mismatch, the red line indicates a mismatch. [—] indicates underspecification.

subsequent studies for vowels in isolation, as well as in words and non-words (Cornell et al., 2011). The role of [LABIAL] is not much discussed in the paper, which is something we will pay attention to in the current study. The contrast /e, ø/ shares the same place of articulation [CORONAL], but differs in the value [LABIAL]: /ø/ is specified for [LABIAL], while /e/ is not. The contrast /o, ø/ differs in place of articulation, /ø/ being [CORONAL] on the surface, while /o/ is [DORSAL], but they share the feature [LABIAL] (see **Figure 1**).

Like German, Dutch has a three-way contrast between /e, ø, o/, and this suggests that the same underlying representations may be at stake. If so, we should replicate the perceptual asymmetries in German for Dutch, which is our first aim. Yet, looking at the linguistic system as a whole, the nature of the front rounded vowels may be different in Dutch and German. In German, many front rounded vowels arise because of morphological umlaut, which fronts back vowels in certain plurals, diminutives, adjectival and verbal forms. For example, German has the singular-plural alternation *Vogel – Vögel* "bird(s)," whereas Dutch has *vogel – vogels*, without vowel alternation. Although the original motivation for Umlaut was phonological (Twadell, 1938), in modern German, Umlaut is no longer phonologically transparent, and has become a morphological rule (Wiese, 1996). Consequently, the /ø/ predominantly occurs in morphologically derived contexts, and might not be part of the underlying vowel inventory of German. Scharinger (2009) has argued that the underlying representations of stems that may alternate between front rounded and back rounded vowels in morphologically derived context may differ from stems that do not alternate. Non-alternating stems with front rounded vowels are thus much less frequent in German than in Dutch. Modern Dutch does not have morphological Umlaut, and front rounded vowels are truly part of the Dutch vowel inventory. The second aim of this paper is to investigate whether this difference between German and Dutch is reflected in differences in phonological processing.

Another difference between the German and Dutch vowels /e, ø, o/ is that the Dutch realization of vowels can be considered semi-diphthongized (Adank et al., 2004), while the German ones have more stable formant frequencies. Phonologically, however, these tense mid vowels are not diphthongs in Dutch; the formant transition is not obligatory. Diphthongization strengthens place features toward the end of vowels, as front vowels become even more front, impacting the second formant (F2), which is the acoustic cue for the front-back dimension. Furthermore, vowels may become higher toward the end, acoustically reflected by a lower first formant (F1). Although the diphthongization may not immediately affect the phonological representations of Dutch vowels, it may make place of articulation contrasts larger, and hence more difficult to find asymmetries. No effect is expected on labiality. Plots and Tables of the Dutch and German vowel formats are given in **Supplementary Appendix 3**.

The current paper therefore replicates the study of Eulitz and Lahiri (2004) and investigates the processing of the same three-way vowel contrast in Dutch. Similar to Eulitz and Lahiri (2004), we measured the MMN in an electroencephalography (EEG)-experiment, but in Dutch listeners. Like in the original study, we tested discrimination of tense mid vowels /e/, /ø/, and /o/ in isolation in a passive oddball paradigm. We predicted to find the same perceptual asymmetry for place (i.e., in the vowel pair /o/-/ø/ we expect to find a larger MMN when [ø] is the deviant, than when [o] is the deviant), supporting underspecification of [CORONAL]. We are less certain about the predictions in the vowel pair /e/-/ø/, which are both coronal, but differ in labiality (rounding), as these are not addressed in Eulitz and Lahiri (2004). There are two reasons to want to investigate potential asymmetries for these vowels. The first reason is that for vowels [LABIAL] is not mutually exclusive with [CORONAL] or [DORSAL], while the latter two are mutually exclusive. This suggests that [LABIAL] has a different status. The second reason is that, although Eulitz and Lahiri do not report an asymmetry for the pair /e/-/ø/, it is not entirely straightforward how the model could account for the discrimination of these vowels: the perceived [CORONAL] of /e/ does not seem to mismatch with [LABIAL] of /ø/ in this case, as /ø/ itself is coronal. This too suggests a different status of [LABIAL]. Hence, investigating listeners' brain responses to this contrast may provide new insights into the nature of the representations of these vowels.

MATERIALS AND METHODS

We replicated the German electroencephalographic (EEG) study by Eulitz and Lahiri (2004) in Dutch. We measured differences in brain responses to three tense mid vowels /e, ø, o/ in Dutch by means of a MMN experiment, using a passive oddball paradigm. The three vowels form three vowel pairs, for which we tested both directions of discrimination. This way, we aim to gain insight in the underlying representations of these sounds with respect to place of articulation and labiality (also known as rounding).

This experiment was conducted at the Donders Centre for Cognitive Neuroimaging in Nijmegen (Netherlands). The study was approved by the local ethics committee "Commissie Mensgebonden Onderzoek" (CMO) Arnhem-Nijmegen, Netherlands, under the general ethics approval (Imaging Human Cognition, CMO 2014/288), and the experiment was conducted in accordance with these guidelines.

Participants

Seventeen right-handed adult native Dutch speakers (10 female, aged 18–30) were included in the final analysis. Participants were recruited using the Radboud Research Participation System (SONA Systems Ltd.). Subjects grew up as monolingual Dutch speakers at least until the age of twelve. Dialect speakers were excluded from participation, as several Dutch dialects have morphological Umlaut, comparable to German. Participants had normal hearing, did not suffer any language or speech impediments (e.g., dyslexia, cleft palate, DLD, etc.), and had not received speech therapy at present or in the past. They had no background in linguistics. Prior to participating, subjects signed an informed consent and were screened for EEG-compatibility (e.g., with respect to claustrophobia or epilepsy). Subjects received a financial reimbursement or study course credits.

Power analysis was done in G*power, based on the values for the F-test for the mean amplitude of Fz reported in Eulitz and Lahiri (2004), using a partial omega squared of 0.245 [based on $F(1,11 = 5.21)$, see Lakens (2013)]. With the resulting effect-size $f(U)$ of 0.57, for a power of 0.8 fourteen participants were required as a minimum for the current study. We used six versions of the experiment to counterbalance order of presentation (each testing all conditions), and we preferred to test each version with an equal number of participants. We therefore aimed for eighteen participants – three participants per version.

In total, twenty-four participants completed this EEG study. Six subjects were excluded due to technical errors (4 subjects), the use of antidepressants (1 subject), and too noisy EEG data (1 subject). These subjects were replaced to get to the aimed 18 subjects. One final subject was excluded due to the failure to show an MMN response to any condition nor overall, resulting in 17 included participants in the final analysis.

Task and Apparatus

Subjects participated in a passive oddball paradigm. Their electrical brain activity was recorded by means of EEG while they listened to streams of vowels in isolation. Stimuli were presented using Presentation Software (version 18.2 02.18.16). The complete EEG-recording took roughly 1.5 h. In every block, one vowel category occurred frequently (the *standard* stimuli), interspersed with tokens of a vowel category that occurred infrequently (the *deviant* stimuli). For instance, participants heard /e/ as a standard and [o] as a deviant sound in one block. Apart from staying awake, there was no explicit task or overt response. During the experiment, participants were seated in front of a computer screen (Benq XL2420Z – 24 inch) in a sound-attenuated booth. Participants were instructed to sit as still and relaxed as possible, to reduce movement artifacts, eye movements and blinks as much as possible. Auditory stimuli were presented through circumaural passive noise canceling Sennheiser headphones. Participants watched a silent movie which kept them engaged and awake for the duration of the entire experiment. The film was presented at screen center with half screen width to minimize saccades. Viewing distance was approximately 100 cm.

Design and Procedure

We tested discrimination of three tense mid vowels /e, ø, o/, which were presented as vowels in isolation. These three vowel categories can be combined into three different vowel pairs (see **Table 1**). For each vowel pair, two conditions were tested, referring to different directions of change: e.g., for vowel pair /e, ø/, in one direction of change /e/ served as a standard and [ø] as a deviant, while in the opposite direction of change /ø/ is the standard and [e] serves as the deviant. With two directions of change for all three vowel pairs, this results in six conditions. Each condition was tested in a separate block. Each participant was tested on all six conditions.

In every block, 1,000 vowels were presented in a passive oddball paradigm, with 85% standards and 15% deviants. An inter stimulus interval of 700 ms was used. Note that due to misinterpretation this ISI is longer than the 500 ms ISI in Eulitz and Lahiri (2004). We used six different orders of blocks – counterbalanced across subjects. Two successive blocks never had the same standard vowel. Each block lasted for ∼15 min. Participants were free to take a break in between the blocks. Participants themselves indicated when they were ready to start the next block. For each of the six different block orders, different pseudo-randomized within-block stimulus lists were used. Pseudo-randomized stimulus lists ensured that every block started with at least three tokens of the standard, and a deviant was always followed by at least two and maximum eleven standards before another deviant was presented. Deviants occurred unpredictably.

Stimuli

Our standard and deviant stimuli were three different tokens of the three Dutch vowels [e] as in *zeef* ("sieve"), [ø] as in *deuk* ("dent"), and [o] as in *poot* ("paw"/"leg"), spoken by a male Dutch native speaker. By using three tokens of each vowel, acoustic variability was introduced to simulate more natural speech perception conditions and to force the processing system to map the incoming acoustic signals onto more abstract representations, rather than focus on properties unimportant in verbal processing (Eulitz and Lahiri, 2004). The vowels were recorded in isolation, and in /hV/ context, where the /h/ was spliced off. For more details regarding stimulus creation, see **Supplementary Appendix 1**.

Stimuli had a duration of 200 ms with 50 ms offset ramps. F0 was similar (∼112 Hz) in all vowel categories. dBA values measured at the headphones for all vowels were within a 1 dB range of 64 dB. Since different vowel categories have different frequency characteristics, which could lead to differences in perceptual loudness, we assessed differences in perceptual loudness in a behavioral pretest. In addition, we assessed the degree of within-category variation by means of a pretest as well. Both these pretests are reported in **Supplementary Appendix 2**.

The vowel categories mainly differ with respect to F2 and F3, which are related to place and labiality features. In **Figure 2**, the three tokens of each of the three vowel categories are placed in the F2/F3 vowel space. **Figure 2** shows that the acoustic distance between [e] and [ø] on the one hand, and between [o] and [ø] on the other, was quite similar. Acoustic stimulus characteristics are reported in more detail in **Supplementary Appendix 3**.

TABLE 1 | Overview of experimental conditions.

Vowel pair	Condition	Standard	Deviant	Contrastive feature(s)
/ø/-/o/	[ø]/o/	/o/	[ø]	Place
	[o]/ø/	/ø/	[o]	
/e/-/ø/	[e]/ø/	/ø/	[e]	Labiality
	[ø]/e/	/e/	[ø]	
/e/-/o/	[e]/o/	/o/	[e]	Place and labiality
	[o]/e/	/e/	[o]	

For each vowel pair two conditions are tested in order to test two directions of discrimination. During testing, surface features of the deviant are mapped onto the underlying features of the standard. A vowel in // represents the standard sound. Vowels in square brackets [] represent the deviant vowel. E.g., condition [e]/o/ measures the response to vowel [e] as a deviant in a stream of /o/-stimuli.

Electroencephalography Recording

Electroencephalography data was recorded in Brain Vision Recorder (Brain Products GmbH, Munich, Germany) with 64 active electrodes (ActiCAP, equidistant – Brain Products) against left mastoid as an online reference, using a sampling rate of 500 Hz. Eye movements and blinks were tracked by four EOG (electrooculography) electrodes: one above and one below the left eye tracking vertical movement and blinks, and one left of left eye and one right of right eye tracking horizontal eye movements. Impedance levels of <20 kΩ were adopted for each channel, and we employed <5 kΩ impedance levels for reference electrodes on both mastoids.

Data Analysis

Data were analyzed using Brain Electrical Source Analyses 6.0 (BESA; MEGIS Software GmbH, Gräfelfing, Germany). EEG data was re-referenced against linked mastoids. Filter

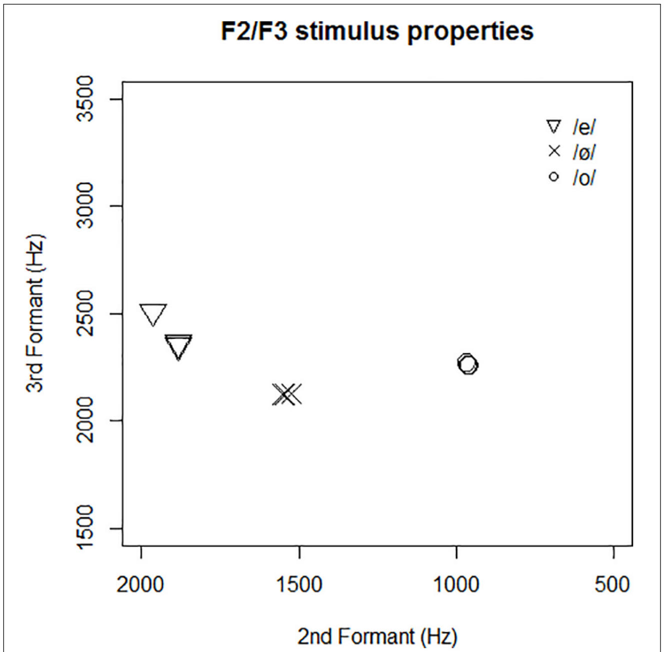


FIGURE 2 | Each of the three tokens per vowel category plotted in F2/F3 space.

and data cleaning parameters are based on analyses of the German experiment (Eulitz and Lahiri, 2004). Data was band pass filtered with low cutoff at 0.1 Hz (6 dB/oct slope) and high cut off 30 Hz (12 dB/oct slope). Epochs containing large non-eye artifacts found by visual inspection were discarded. Independent component analysis (ICA) was performed to correct for eye movement and blink artifacts. Remaining epochs with artifacts exceeding 100 μV within the −100 till 650 ms time window were discarded.

Each vowel category served as a standard in two blocks, for example: /e/ occurs as a standard in a block with [o] as a deviant as well as in a block with [ø] as a deviant. The first standard stimulus in a block was removed, as well as the first standard immediately after a deviant stimulus. Visual inspection showed that the event-related potential (ERP) waveforms of the same standard vowel in different blocks were completely overlapping in the 100–250 ms MMN time-window. Therefore, for standard vowels, the ERP was calculated over both blocks for every vowel. This way, we have the most robust measure for standards.

Each vowel has an inherent vowel specific auditory response due to its frequency characteristics, which is always there regardless of context. Event-related brain responses to different vowels show different waveform morphologies that may have nothing to do with any change detection process. The MMN response is a difference waveform between two ERPs (standard – deviant). To avoid effects due to the use of different vowels, the MMN is calculated by subtracting the ERP to a vowel when it is a deviant in a particular context (i.e., the vowel that is the standard in that block), from when that same vowel serves as a standard itself – so in a different block. For example, the MMN for the condition where [e] serves as a deviant and /o/ as standard is calculated as follows:

$$\text{MMN [e]/0/} = \text{[ERP to standard /e/]} - \text{[ERP to deviant [e] in context of /o/]}$$

As such, the MMN reflects *change* perception only, and we can assess the impact of the context of the deviant. We calculated MMNs for each condition for each individual participant. For each experimental condition, we used two dependent variables for our analyses, which are similar to the ones used in Eulitz and Lahiri (2004):

1. *MMN latency* measured at the negative maximum amplitude at frontal electrode (Fz) in the latency range from 100 to 250 ms post-stimulus onset, based on a window around the mean MMN latency over all conditions.
2. *MMN amplitude (μV)* at Fz position measured as the mean amplitude across 80 ms centered at the mean MMN latency across subjects in the corresponding experimental condition.

These two parameters were subjected to a two-way repeated measures ANOVA. The ANOVA was restricted to the two pairs of inversion with a similar acoustic change: /e, ø/ and /o, ø/. The third vowel pair /e, o/ shows markedly larger acoustic difference between the two vowels (see **Figure 2**), and differed on two phonological features instead of one, and was therefore statistically tested separately. The ANOVA had two within-subject factors:

1. *Pair-of-Inversion* showing a similar acoustic change: [e]/ø/ versus [ø]/e/ and [o]/ø/ versus [ø]/o/.
2. *Direction-of-Change* of F2 frequency between standard and deviant: ascending in [e]/ø/ and [ø]/o/, but descending in [ø]/e/ and [o]/ø/.

This ANOVA assesses whether the MMN differs between vowel pairs (Pair of Inversion) and whether there are general acoustic asymmetric influences on the MMN (Direction-of-Change), or different asymmetries for different vowel pairs (interaction Pair-of-Inversion and Direction-of-Change).

Asymmetries in the MMN for the different vowel pairs were subsequently assessed by directly comparing the MMN characteristics (latency and amplitude at Fz) for the different directions of change by means of paired samples *t*-tests (planned comparisons).

RESULTS

A clear grand average overall MMN was found with a peak latency of 218 ms, with a clear frontal topography. Amplitude maps as well as voltage topography maps show typical MMN topographies (e.g., Näätänen, 2001; Näätänen et al., 2007) as is displayed in **Figure 3** – with a predominant influence of left and right hemispheric temporal generators on the MMN, similar to Eulitz and Lahiri (2004). Visual inspection of grand average MMN waveforms showed a later peak latency than reported in Eulitz and Lahiri (2004), where all conditions had an MMN peak latency shorter than 170 ms, possibly due to the fact that we used a longer ISI. Because our grand average MMN was later, we used the time window of 100–250 ms to find peak latency values for each condition in each participant, covering the entire window of where a typical MMN peak would occur.

The two-way ANOVA revealed statistically significant interactions for *Pair of Inversion*Direction of Change* for MMN amplitude at Fz ($F(1,16) = 26.3$; $p < 0.001$) as well as for MMN peak latency at Fz ($F(1,16) = 6.3$; $p = 0.023$). In addition, a main effect for *Pair of Inversion* (i.e., vowel pair) appeared significant

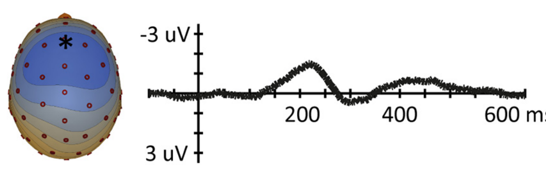


FIGURE 3 | (Left) Overall MMN voltage topography map (μV) for the 5 ms time-window around the average MMN peak latency (218 ms). Blue in topoplot indicates negative potential, red positive potential. **(Right)** MMN waveforms at frontal electrode position (Fz).

($F(1,16) = 12.09$; $p = 0.003$) for amplitude, but not for latency. *Direction of Change* appeared non-significant in both amplitude and latency measures.

The attested main-effect of *Pair of Inversion* in Fz amplitude shows that the two vowel pairs with similar acoustic distance behave differently. The lack of a main effect of *Direction of Change* implies that results cannot be explained merely on acoustic change in F2. The attested interaction in all measures implies that the impact of direction of change differs for different contrasts in processing of these vowel pairs.

Planned comparisons of directions of change to assess asymmetries were tested with *t*-tests. Results of MMN amplitude and latency measured at Fz for each condition are reported in **Table 2** below. MMN topographical plots and waveforms are presented in **Figures 4–6** for each contrast separately.

For the place contrast /o, ø/, a paired *t*-test showed a significantly larger MMN amplitude at Fz for condition [ø]/o/ ($M = -1.19$ μV; SD = 1.81) than for condition [o]/ø/ ($M = 0.28$ μV; SD = 1.31); $t(16) = 4.281$, $p < 0.001$. The peak latency measure resulted in a significant asymmetry in the same direction: a shorter latency was found for [ø]/o/ ($M = 210$ ms; SD = 27) than for [o]/ø/ ($M = 239$ ms; SD = 38); $t(16) = 2.823$, $p = 0.012$.

For the labiality contrast /e, ø/, a paired *t*-test on Fz amplitude revealed a significantly larger MMN response for [ø]/e/ ($M = -2{,}18$ μV; SD = 1.66) compared to condition [e]/ø/ ($M = -1.08$ μV; SD = 1.28); $t(16) = -4{,}053$, $p < 0.001$. Thus, the results show a significant asymmetry between [ø]/e/ and [e]/ø/ with respect to amplitude. Fz MMN peak latency showed no significant difference between [ø]/e/ ($M = 219$ ms; SD = 31) compared to [e]/ø/ ($M = 216$ ms; SD = 29); $t(16) = 0.284$, $p = 0.78$.

For two-feature contrast /e, o/, the paired *t*-test did not show significant results: the MMN for [e]/o/ did not differ from the MMN for [o]/e/, neither in Fz amplitude ([e]/o/: $M = -1.49$ μV; SD = 0.93; [o]/e/ : $M = -2.06$ μV; SD = 1.1; $t(16) = -1.857$, $p = 0.083$), nor in Fz latency ([e]/o/; $M = 203$ ms; SD = 33; [o]/e/ ($M = 212$ ms; SD = 23; $t(16) = 1.051$, $p = 0.309$).

DISCUSSION

The present paper set out to test whether Dutch listeners show the same perceptual asymmetries as German listeners in an MMN paradigm in which three vowel pairs were investigated. We found significant asymmetries for both one-feature contrasts (place contrast /o, ø/, labiality contrast /e, ø/), but no significant asymmetry was found for our two-feature contrast (/e, o/). In section 5.1 we argue that the phonological specifications of the vowels together with the matching procedure can account for these differences. We discuss each pair separately. In section 5.2, we argue that acoustic differences between the vowels in each pair cannot straightforwardly explain the attested asymmetries. Before turning to the general conclusion, we also discuss predictions based on theories that rely on experience or frequency, and argue that these do not offer a clear explanation for the attested asymmetries.

Explaining Asymmetries

Place of Articulation: /o/-/ø/

In the vowel pair /o/-/ø/, where the vowels share labiality, but differ in place of articulation, we replicated Eulitz and Lahiri (2004), in that a change from standard /o/ to deviant [ø] shows a larger MMN amplitude and shorter peak latency compared to the reverse [o]/ø/. In other words, there is a conflict in one direction. This provides evidence for [CORONAL] underspecification in Dutch, similar to what has been argued for German. A perceived [DORSAL] feature is a no-mismatch with an underspecified [CORONAL] feature, while a perceived [CORONAL] feature in the surface representation is a mismatch with a specified [DORSAL] feature in the underlying representation (see **Figure 7**). A similar finding has been found by De Jonge and Boersma (2015) for French, which also has a three-way contrast between the high vowels /i, y, u/ and mid vowels /e, ø, o/.

Labiality: /e/-/ø/

In the vowel pair /e/-/ø/, both vowels share place of articulation as both are [CORONAL]. They differ in labiality as /ø/ is specified for [LABIAL], while /e/ is not. The results from our study show a perceptual asymmetry: when /e/ is standard and [ø] deviant the data show a larger MMN amplitude than in the reverse condition [e]/ø/. This asymmetry was not reported for German. Eulitz and Lahiri (2004) argue that this is a non-conflicting situation with respect to place of articulation, which is why the MMN does not vary. In this condition, Dutch and German seem to differ. This finding raises at least two questions. First, how can we account for conflict based on surface and underlying features in the case of labiality, and second, why does this difference between the two languages arise, assuming the same underlying and surface features, as shown in **Figure 1**.

In the FUL model, a mapping procedure is proposed, where features extracted from the acoustics are compared to features stored in the underlying representation. The goal of this feature mapping is to deselect unwanted candidates, and to limit the number of word candidates. Of course, only the relevant comparisons are made. After all, when for instance a feature [CORONAL] is extracted, only mapping to a mutually exclusive underlying feature like [DORSAL] would result in a meaningful/informative mismatch, whereas mapping [CORONAL] onto a feature like [HIGH] is neither efficient nor informative. In order for the mapping process to result in meaningful (no-)(mis)matches, the scope of the comparison of surface features must be defined. In the FUL model, [LABIAL], [CORONAL] and [DORSAL] share a node in the feature tree, the ARTICULATOR node, as shown in **Figure 8A**, an assumption shared with many others models (e.g., Chomsky and Halle, 1968; Sagey, 1986; Clements and Hume, 1995; Clements, 1999). Alternatively, [CORONAL] and [DORSAL] could be placed under a separate node, the LINGUAL node, as shown in **Figure 8B**. This assumption is also shared by various researchers (Browman and Goldstein, 1989; Keyser and Stevens, 1994). Keyser and Stevens (1994) assume [CORONAL] and [DORSAL] to form a single constituent – LINGUAL – because they both involve the tongue as its articulator. The tongue blade and tongue body are not completely independent, due to their anatomical connection. In

TABLE 2 | Table of results: MMN amplitude and peak latency at Fz for all six experimental condition.

	Vowel pair	Distinctive feature	Condition	Standard	Deviant	Mean Fz Amplitude ±SD (μV)	Mean peak latency ± SD (ms)
1	/o/-/ø/	Place	[ø]/o/	/o/	[ø]	−1.19 ± 1.81*	210 ± 27*
			[o]/ø/	/ø/	[o]	0.28 ± 1.31*	239 ± 38*
2	/e/-/ø/	Labiality	[ø]/e/	/e/	[ø]	−2.18 ± 1.66*	219 ± 31
			[e]/ø/	/ø/	[e]	−1.08 ± 1.28*	216 ± 29
3	/e/-/o/	Place & Labiality	[e]/o/	/o/	[e]	−1.49 ± 0.93	203 ± 33
			[o]/e/	/e/	[o]	−2.06 ± 1.1	212 ± 23

Significance of paired-samples t-tests comparing directions within one vowel pair are indicated with an asterisk. Overview of experimental conditions: Vowels in square brackets [] represent the deviant vowel to which the MMN response is measured in a certain context. A vowel in // represents the standard (=context).

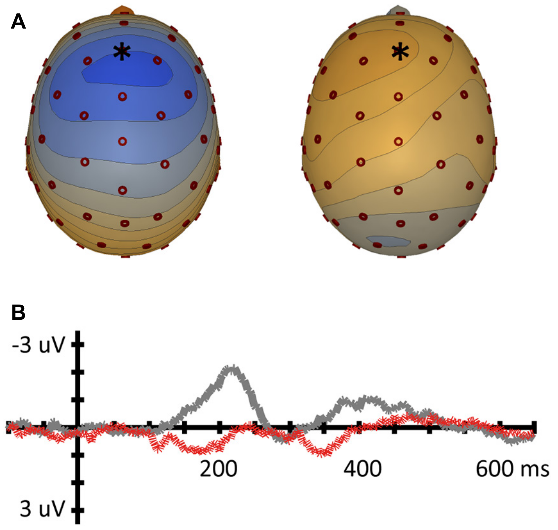


FIGURE 4 | Mismatch negativity for the place contrast (vowel pair /o-ø/). **(A)** Topographic voltage plots of the MMN at the average MMN peak latency for both conditions (left: [ø]/o/, 210 ms; right [o]/ø/, 239 ms). Blue in topoplot indicates negative potential (μV), red positive potential, electrode Fz is indicated with an asterisk. **(B)** MMN waveforms for the Fz electrode [ø]/o/ (gray) and [o]/ø/ (red dashed).

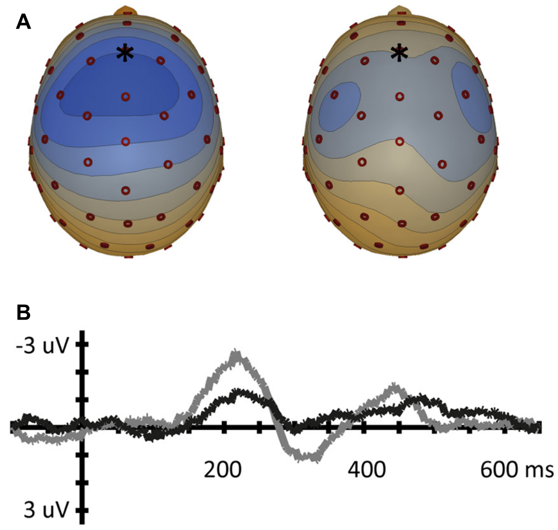


FIGURE 5 | Mismatch negativity for the Labiality contrast (vowel pair /e-ø/). **(A)** Topographic voltage plots of the MMN at the average MMN peak latency for both conditions (left: [ø]/e/, 219 ms; right [e]/ø/, 216 ms). Blue in topoplot indicates negative potential (μV), red indicates positive potential, electrode Fz is indicated with an asterisk. **(B)** MMN waveforms for the Fz electrode [ø]/e/ (gray) and [e]/ø/ (black).

contrast, the lips are anatomically and articulatorily independent from the tongue. For these reasons, Keyser and Stevens assume a separate LABIAL node, whereas [CORONAL] and [DORSAL] are parented by a LINGUAL node. Others have argued that features under the LINGUAL node often pattern together in phonological processes.

For our purposes, separating a LABIAL and a LINGUAL node solves the issue that [LABIAL] may be combined with [DORSAL] and [CORONAL] (in languages like Dutch), which violates the mutual exclusivity assumption of FUL. With separate nodes for [LABIAL] on the one hand and [DORSAL] and [CORONAL] on the other hand, the mutual exclusivity assumption is no longer violated, since [LABIAL] no longer shares the parental node with [DORSAL] and [CORONAL]. This new hierarchy does not affect the place features discussed above, as [DORSAL] and [CORONAL] still share a node, but it does affect labiality as [LABIAL] is the only feature under the LABIAL node. A vowel is either [LABIAL] or lacks a LABIAL node. [NON-LABIAL] is neither a phonological feature, nor does it have stable acoustical features that would enable the perceptual system to extract it from the acoustics as a surface representation feature. When the acoustic feature corresponding to [LABIAL] is extracted from the signal and hence part of the surface representation, mapping it to the underlying representation with [LABIAL] results in a match. When the underlying representation lacks a LABIAL node, mapping cannot take place. This implies a phonological discrepancy between the surface and underlying representation, and we argue that such a case also indicates a phonological conflict. This is shown in **Figure 9**.

A similar asymmetry regarding the feature [VOICE] has been reported by Van der Feest and Fikkert (2015) in Dutch toddlers' perception. They report that a change from a voiceless toward a voiced speech sound is perceptually more salient than vice versa, which resembles the case of labiality. A change from a non-feature (voiceless) toward a specified feature [VOICE] appears to result in phonological conflict, whereas the reverse does not.

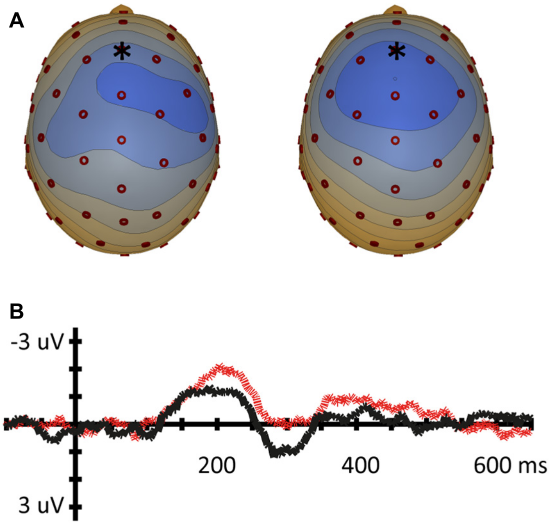


FIGURE 6 | Mismatch negativity for the two-feature contrast (vowel pair /e-o/). **(A)** Topographic voltage plots of the MMN at the average MMN peak latency for both conditions (left: [e]/o/, 203 ms; right [o]/e/, 212 ms). Blue in topoplot indicates negative potential (μV), red indicates positive potential, electrode Fz is indicated with an asterisk. **(B)** MMN waveforms for the Fz electrode [e]/o/ (black solid line) and [o]/e/ (red dashed).

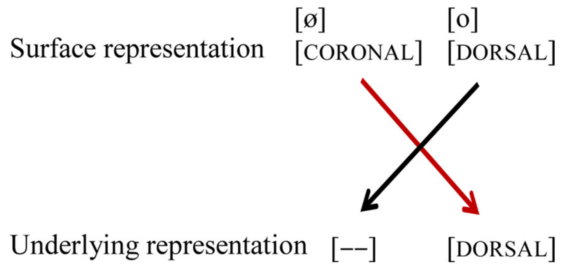


FIGURE 7 | Asymmetrical perception of the vowel pair /o-ø/. Black lines indicate no-mismatch, the red line indicates a mismatch. [– –] indicates underspecification.

Lahiri (2018) also points out that [VOICE] could have its own node, similar to what we propose for [LABIAL]. This line of reasoning is also compatible with the perceptual asymmetries presented in Hestvik and Durvasula (2016) for English, albeit the specified feature here is [SPREAD GLOTTIS] for /t/, while /d/ is unspecified (lacking laryngeal specification). They noticed a larger MMN for [d]/t/ than for [t]/d/, although they presented a somewhat different analysis. Using a similar mapping rule, this would explain the laryngeal asymmetry (where either [VOICE] is specified as in Dutch, or [SPREAD GLOTTIS] as in English): the mapping cannot be completed when a voiced sound is heard but mapped onto a voiceless underlying representation where the VOICING node is absent, hence resulting in a phonological conflict. Whether this idea holds true more generally should be assessed in future research.

The second question raised by the results is why Dutch and German listeners' brains react differently in this condition. Eulitz and Lahiri (2004) reported symmetrical perception for the vowel pair /e/-/ø/, which share place of articulation, and hence do not constitute a conflicting situation for place of articulation. This result was replicated for German listeners in an MMN passive oddball paradigm using both words and non-words (Cornell et al., 2011), who similarly report symmetrical results in this condition. This contrasts with the results in the current paper, which not only show a place asymmetry but also an asymmetry for labiality. One hypothesis is that the status of /ø/ as an underlying phoneme is different in Dutch and German. German has far fewer monomorphemic words with /ø/ than Dutch. In German, [ø] often arises as the result of morphological Umlaut as in the plural in *Vogel* [o] – *Vögel* [ø] "bird(s)," and in the comparative form of the adjective in *hoh* [o] – *höher* [ø] ("high – higher"). Hence, [ø] often occurs in derived environments, rather than in the lexicon. Although /ø/ occasionally also occurs in non-derived (lexical) environment, its contrastive value in German is limited, in comparison to Dutch, which may explain why German and Dutch listeners react differently to [LABIAL] in the context of [CORONAL] vowels: in German [LABIAL] is not a strong lexical contrast, while in Dutch it is. The implication is that the vowel inventory is based on lexical stems, rather than on all surface forms, and that hearing a front rounded vowel in German immediately activates the morphology. A perceived [LABIAL] will in those cases automatically activate back vowels in German, which is not the case in Dutch, where [LABIAL] is used to mark the contrast between /e/ and /ø/. While testing this is beyond the scope of this paper, it certainly warrants further research. There is some evidence in the literature that listeners are sensitive to critical phonological properties in parsing morpho-phonological forms (e.g., Post et al., 2008; Pliatsikas et al., 2014).

Scharinger (2009) and Scharinger et al. (2010) present an alternative approach to morphologically related words and assume that the vowels in lexical words that alternate in derived forms (i.e., certain morphological categories), such as the German word *Vogel*, have a different underlying representation than words with dorsal vowels that do not alternate. Specifically, they argue that the /o/ in *Vogel* is underspecified for [DORSAL], but specified for [LABIAL]. Under this view [ø] does not mismatch with the /o/ vowel in *Vogel*: the perceived [CORONAL] no longer mismatches with [DORSAL], as [DORSAL] is not part of the underlying representations. /ø/ and $/o/_{alternating}$ have the same underlying representation, i.e., [LABIAL], but at a later stage a default coronal fill-in rule and a specific dorsal fill-in rule apply, differentiating both vowels in the surface representation. In terms of the hierarchy proposed in this paper, this would mean that these vowels lack specification under the LINGUAL node in the underlying representation. The influence of morphological alternations on the underlying vowel system requires further investigation.

Two Feature Contrast: /e/-/o/

The vowel pair /e/-/o/ differs in two features, both place of articulation and labiality. This vowel pair is also acoustically more different than both /e/-/ø/ and /o/-/ø/. We found no significant asymmetry for this two-feature contrast. From the discussion of

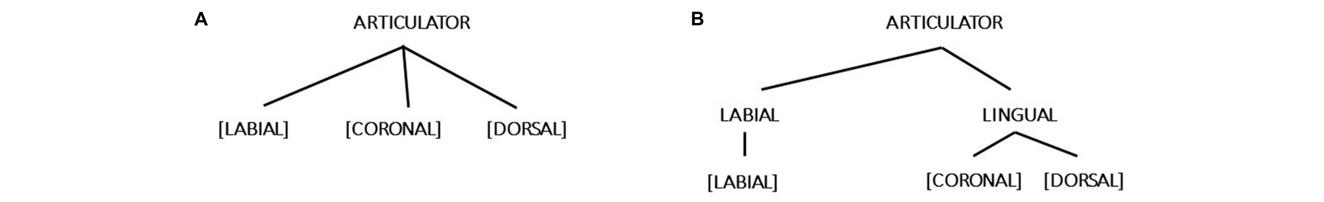


FIGURE 8 | Relevant features [LABIAL], [CORONAL], and [DORSAL] in the feature tree of the FUL-model **(A)**, and the revised model proposed in this paper **(B)**.

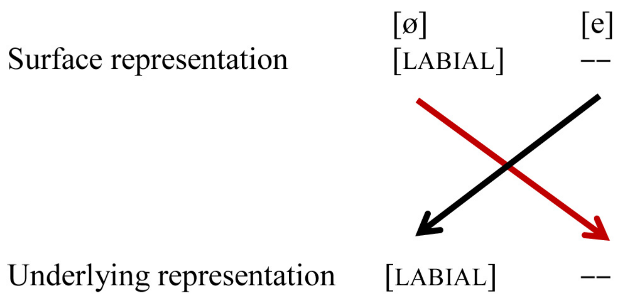


FIGURE 9 | Asymmetrical perception of the vowel pair /e-ø/. Black lines indicate no-mismatch, the red line indicates a mismatch. [– –] indicates no feature selected (SR) and node lacking (UR).

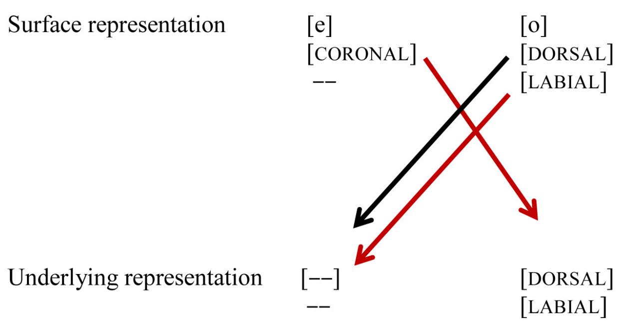


FIGURE 10 | Asymmetrical perception of the vowel pair /e-o/. Black lines indicate no-mismatch, the red line indicates a mismatch. [– –] indicates no feature selected (SR) and node lacking (UR). The matching for place of articulation and labiality shows conflicting results.

the two vowel pairs with one-feature contrast, this may not come as a surprise. For the place contrast, we would expect an enhanced MMN in the context [e]/o/, but not vice versa. However, for the labiality contrast, we would expect an enhanced MMN in the context [o]/e/, as shown in **Figure 9**.

One explanation for the lack of a significant asymmetry may thus be that these cancel each other out (see **Figure 10**). Another explanation could be that the two vowels are acoustically very different and easy to discriminate, as assumed in Eulitz and Lahiri (2004).

Alternative Explanations

The literature mentions at least two other types of explanations for attested asymmetries in vowel perception: phonetic saliency

TABLE 3 | Overview of predictions for FUL, NRV and relative frequency.

Vowel pair	FUL +	NRV F1–F2	NRV F2–F3	Relative frequency	Attested asymmetry
/ø/-/o/	[ø]/o/	[o]/ø/	[ø]/o/	[o]/ø/	[ø]/o/
/e/-/ø/	[ø]/e/	[ø]/e/	[e]/ø/	[e]/ø/	[ø]/e/

For each vowel pair two conditions are tested in order to test two directions of discrimination. For each vowel pair and for each theory, the table indicates the direction of change in which the largest MMN is expected. During testing, surface features of the deviant [] are mapped onto the underlying features of the standard //. The predictions are conform the asymmetries in this paper, but the NRV makes predictions depending on closeness or convergence (based on F1-F2, as is the case /o/, or on F1-F2 or F2-F3, as in the case of /e/, as can be seen in ***Supplementary Appendix 3****). The frequency predictions do make the opposite predictions. Marked in gray are the predictions that are not conform to our experimental findings, which are given in the final column.*

and frequency of use. Phonetic saliency or ease of discrimination has been used to account for asymmetries in vowel perception, among others by proponents of the Natural Referent Vowel (NRV) framework (Polka and Bohn, 2011; Masapollo et al., 2017a,b). It assumes that vowels with formant frequencies closer together have focalized energy, and hence, universally are more salient in perception than vowels with formants further apart. The most focalized vowels are /i/, /a/ and /u/, the cornerstones of the vowel space. Consequently, changes from less to more focal vowels are easier to discriminate. In light of our study, a larger MMN is expected when the standard is non-focal, and the deviant is focal. Based on the convergence or closeness of F2 and F3, /e/ is more focal than /ø/, although toward the end of the vowel the difference becomes small, and the formants are slightly overlapping, making them less distinguishable. Based on the F1-F2 dimension, /ø/ would be more focal than /e/, but the formants do not get close, and it remains the question whether the formants lead to focalized energy. Moreover, this latter prediction is contradicting the claim made in Polka and Bohn (2011). In other words, the predictions based on the NRV are not entirely straightforward. For the vowel pair /o/-/ø/ the predictions are not clear either. Based on F1-F2, /o/ is more focal than /ø/. Based on F2-F3 /ø/ is more focal than /o/. The predicted asymmetries can be seen in **Table 3**.

For frequency, or experienced-based theories, like for instance the Native Language Magnet (NLM) theory (Kuhl, 1991; Kuhl et al., 2008) it is usually assumed that category building is based on distributions in the input, and more frequent vowels have a stronger magnet effect, warping the perceptual space around them. Therefore, poorer discrimination is expected in the direction from more frequent to less frequent. However, this discrimination effect usually holds for within-category discrimination. If both are categories in the language, predictions are less clear. However, one could hypothesize that frequent vowels allow for more variation, and hence it is expected that a frequent standard and an infrequent deviant would be more difficult to discriminate, and hence show a smaller MMN effect, than vice versa. With respect to frequency, front round vowels are relatively infrequent in Dutch compared to front (unround) vowels and back (round) vowels. Baayen et al. (1995) report the following percentages for the relevant vowels in CELEX: [e] = 6.7%; [o] = 6.0%; [ø] = 2.3%, meaning /e/ and /o/ occur roughly twice as often as /ø/ in Dutch. The predictions based on these vowel distributions are thus opposite to the findings in our experiment. The different predictions are presented in the **Table 3**.

To summarize, the current NRV framework only partly predicts the attested asymmetries. It must be noted that the results reported in Polka and Bohn (2011) do not always conform to their own predictions. Notably, they report better discrimination in the direction from/ e/ to /ø/ in Danish infants. The frequency account does not make the right predictions, and it also raises questions as how large the frequency difference needs to be to predict asymmetries. The proposed geometry and mapping algorithm in this paper shows how listeners might evaluate the incoming signal based on its phonology, which is conform the attested asymmetries.

CONCLUSION

The current study showed evidence for asymmetrical processing of vowels, and the attested asymmetries provide further support for the FUL model, in which underlying phonological representations are underspecified. It replicated the place asymmetry in German listeners as reported in Eulitz and Lahiri (2004) for Dutch. While they did not find asymmetries between /e/ and /ø/, i.e., asymmetries based on [LABIAL], this asymmetry was shown in the current study. We propose to place [LABIAL] under a separate node from [CORONAL] and [DORSAL] (**Figure 8B**), which we group together under a LINGUAL node, following for instance Keyser and Stevens (1994). Our proposal has at least two benefits: features under the same node are mutually exclusive, a central assumption made by FUL, which however had to be relaxed to allow [LABIAL] to combine with [CORONAL] or [DORSAL] in the traditional FUL model. Moreover, as [LABIAL] is the only feature under the LABIAL node, its mapping must match, as a no-mismatch is no viable option. If the perceived [LABIAL] in the surface representation cannot be matched, the mapping is aborted, leading to a conflict, and hence a perceptual asymmetry. This is the case when the standard is /e/ (no labial node), and the deviant [ø], with a surface feature [LABIAL]. A consequence is that not only a mapping mismatch (as traditionally discussed in FUL), but also an aborted mapping implies a phonological conflict. Finally, this study shows that languages can show cross-linguistic differences if contrasts are implemented differently in the languages.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by “Commissie Mensgebonden Onderzoek” (CMO) Arnhem-Nijmegen, Netherlands, under the general ethics approval (Imaging Human Cognition, CMO 2014/288). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NR and PF: conceptualization, writing – original draft preparation, writing – review and editing, and project administration. NR and TS: methodology, software, validation, formal analysis, and visualization. NR: investigation and data curation. PF: resources and funding acquisition. TS and PF: supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was supported by grant NWO 322-75-005 “Asymmetries in West-Germanic vowel systems” from the Dutch Science Foundation (NWO) to PF. The project is carried out by NR under supervision of PF and TS.

ACKNOWLEDGMENTS

We would like to thank Aditi Lahiri, Sandra Kotzor, and Laura Hahn for their help and advice. Colin Brooks was very helpful with technical support. We thank Jeroen Dera for lending his voice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.629648/full#supplementary-material

REFERENCES

Adank, P., Hout, R., and Smits, R. (2004). An acoustic description of vowels of Northern and Southern standard dutch. *J. Acoust. Soc. Am.* 116, 1729–1738. doi: 10.1121/1.1779271

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

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia Linguistic Data Consortium, University of Pennsylvania.

Browman, C., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper and Row.

Clements, G. N. (1999). "'The geometry of phonological features,'" in *Phonological Theory: the Essential Readings*, ed. J. Goldsmith (Oxford: Blackwell), 201–223.

Clements, G. N., and Hume, E. (1995). "'The internal organization of speech sounds,'" in *Handbook of Phonological Theory. Oxford: Basil*, ed. J. Goldsmith (Oxford: Blackwell), 245–306.

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

De Jonge, M. J. I., and Boersma, P. (2015). "'French high-mid vowels are underspecified for height', in the scottish consortium for ICPhS 2015," in *Proceedings of the 18th International Congress of Phonetic Sciences*, (Glasgow: The University of Glasgow).

Dresher, B. E. (2009). *The Contrastive Hierarchy in Phonology. (Cambridge Studies in Linguistics* 121.). Cambridge: Cambridge University Press.

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Goldinger, S. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251

Hestvik, A., and Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain Lang.* 152, 28–43.

Johnson, K. (1997). "'Speech perception without speaker normalization: an exemplar model,'" in *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullenix (San Diego, CA: Academic Press), 145–166.

Keyser, S., and Stevens, K. (1994). Feature geometry and the vocal tract. *Phonology* 11, 207–236.

Kuhl, P. K. (1991). Human adults and human infants show a "Perceptual Magnet Effect" for the prototypes of speech categories: monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/BF03212211

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Lahiri, A. (2018). "Predicting universal phonological features," in *Phonological Typology*, eds L. Hyman and F. Plank (Berlin: Mouton de Gruyter), 229–272.

Lahiri, A., and Reetz, H. (2002). Underspecified recognition. *Lab. Phonol.* 7, 637–676.

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonet.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

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

Masapollo, M., Polka, L., Molnar, M., and Ménard, L. (2017b). Directional asymmetries reveal a universal bias in adult vowel perception. *J. Acoust. Soc. Am.* 141:2857. doi: 10.1121/1.4981006

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, 1–21. doi: 10.1017/s0048577201000208

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Pierrehumbert, J. (2002). Word-specific phonetics. *Lab. Phonol.* 7, 101–139.

Pliatsikas, C., Wheeldon, L., Lahiri, A., and Hansen, P. C. (2014). Processing of zero-derived words in English: an fMRI investigation. *Neuropsychologia* 53, 47–53. doi: 10.1016/j.neuropsychologia.2013.11.003

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural referent vowel (NRV) framework: an emerging view of early phonetic development. *J. Phonet.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Post, B., Marslen-Wilson, W. D., Randall, B., and Tyler, L. K. (2008). The processing of English regular inflections: phonological cues to morphological structure. *Cognition* 109, 1–17. doi: 10.1016/j.cognition.2008.06.011

Sagey, E. (1986). *The Representation of Features and Relations in Nonlinear Phonology*. PhD dissertation, Cambridge, MA: MIT Press.

Scharinger, M. (2009). Minimal representations of alternating vowels. *Lingua* 119, 1414–1425.

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

Twadell, W. (1938). A note on old high german umlaut. *Monatshefte Für Deutschen Unterricht* 30, 177–181.

Van der Feest, S. V. H., and Fikkert, P. (2015). Building phonological lexical representations. *Phonology* 32, 207–239. doi: 10.1017/S0952675715000135

Wiese, R. (1996). Phonological versus morphological rules: on german umlaut and ablaut. *J. Linguist.* 32, 113–135. doi: 10.1017/S0022226700000785

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fnhum.2021.643655

Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults

Natalia Nudga[1†], Josef Urbanec[2,3†], Zuzana Oceláková[1], Jan Kremláček[2,4] and Kateřina Chládková[5,6]*

[1] Faculty of Arts, Institute of Phonetics, Charles University, Prague, Czechia, [2] Department of Pathological Physiology, Faculty of Medicine in Hradec Králové, Charles University, Hradec Králové, Czechia, [3] Pediatrics Department, Havlíčkův Brod Hospital, Havlíčkův Brod, Czechia, [4] Department of Medical Biophysics, Faculty of Medicine in Hradec Králové, Charles University, Hradec Králové, Czechia, [5] Faculty of Arts, Institute of Czech Language and Theory of Communication, Charles University, Prague, Czechia, [6] Institute of Psychology, Czech Academy of Sciences, Brno, Czechia

OPEN ACCESS

Edited by:
Arild Hestvik,
University of Delaware, United States

Reviewed by:
Philip J. Monahan,
University of Toronto, Canada
William James Idsardi,
University of Maryland, College Park, United States

****Correspondence:***
Natalia Nudga
nat.nudga@gmail.com

† These authors share first authorship

Specialty section:
This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience

Received: *18 December 2020*
Accepted: *29 June 2021*
Published: *09 August 2021*

Citation:
Nudga N, Urbanec J, Oceláková Z, Kremláček J and Chládková K (2021) Neural Processing of Spectral and Durational Changes in Speech and Non-speech Stimuli: An MMN Study With Czech Adults.
Front. Hum. Neurosci. 15:643655.
doi: 10.3389/fnhum.2021.643655

Neural discrimination of auditory contrasts is usually studied via the mismatch negativity (MMN) component of the event-related potentials (ERPs). In the processing of speech contrasts, the magnitude of MMN is determined by both the acoustic as well as the phonological distance between stimuli. Also, the MMN can be modulated by the order in which the stimuli are presented, thus indexing perceptual asymmetries in speech sound processing. Here we assessed the MMN elicited by two types of phonological contrasts, namely vowel quality and vowel length, assuming that both will elicit a comparably strong MMN as both are phonemic in the listeners' native language (Czech) and perceptually salient. Furthermore, we tested whether these phonemic contrasts are processed asymmetrically, and whether the asymmetries are acoustically or linguistically conditioned. The MMN elicited by the spectral change between /a/ and /ɛ/ was comparable to the MMN elicited by the durational change between /ɛ/ and /ɛː/, suggesting that both types of contrasts are perceptually important for Czech listeners. The spectral change in vowels yielded an asymmetrical pattern manifested by a larger MMN response to the change from /ɛ/ to /a/ than from /a/ to /ɛ/. The lack of such an asymmetry in the MMN to the same spectral change in comparable non-speech stimuli spoke against an acoustically-based explanation, indicating that it may instead have been the phonological properties of the vowels that triggered the asymmetry. The potential phonological origins of the asymmetry are discussed within the featurally underspecified lexicon (FUL) framework, and conclusions are drawn about the perceptual relevance of the place and height features for the Czech /ɛ/-/a/ contrast.

Keywords: mismatch negativity, auditory processing, vowels, phonology, perceptual asymmetries

INTRODUCTION

Speech perception is a cognitive process which transforms the acoustic signal into respective neural representations in the human brain. One of the most fundamental properties of human speech perception is the ability to detect phonetic and phonological contrasts. Sensitivity to such contrasts has been examined by the means of behavioral tests (discrimination or categorization tasks) (Repp and Crowder, 1990; Polka and Bohn, 2003; Johnson, 2015) as well as via techniques that monitor brain activity, such as event-related potentials (ERPs) measured with electroencephalography (EEG; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015) or their magnetic equivalents measured with magnetoencephalography (Scharinger et al., 2016; Højlund et al., 2019). The most common ERP component used to study the brain response to an auditory contrast is the mismatch negativity (MMN). The MMN response is elicited by an irregularity, typically when a series of frequently presented stimuli, standards, is interrupted by a different infrequent stimulus, deviant. ERP studies show that the magnitude of the MMN reflects the extent of the perceived difference between the standard and deviant, whereby not only the acoustic distance but also the category membership of the stimuli modulate the strength of the response (Näätänen et al., 1997). The MMN can thus be used to estimate the linguistic importance and relevance of phonetic differences between stimuli for speech perception.

The auditory ERP component MMN and its magnetic correlate MMNm have been used to assess the neural processing of both vowels and consonants, and to study the relevance of qualitative, or less commonly, quantitative phonemic contrasts. Ylinen et al. (2005) studied the processing of consonant quality and quantity via MMN, focusing on stop consonants /p/, /p:/, /t/, and /t:/. In their experiment, the plosive [t:] served as the standard, [t] as a quantity deviant, [p:] as a quality deviant, and [p] as a double deviant (all embedded in the same [i_i] frame). The MMN elicited by the double deviant was approximately equal to the sum of the quantity- and quality-deviant MMNs and the authors concluded that consonant quality and quantity are processed independently. Their results also show that the quantitative change of the consonant elicited greater and earlier MMN response than the qualitative change. This finding of differential strength of processing of phoneme quality and quantity could be specific to plosive consonants. In vowels, for instance, a change in quality is much more salient than a change in plosive consonant place of articulation. The question thus remains how robustly quality versus quantity changes are processed in vowels.

Previous studies focusing on vowels show that changes in vowel spectral quality elicit a larger MMN in listeners for whom these changes represent a linguistic, i.e., phonemic change, than in listeners for whom these changes are not phonemic (Näätänen et al., 1997). Similarly, changes in the duration of vowels elicit a stronger MMN response in listeners in whose native language vowel length is phonemic than in listeners for whom it is not (Kirmse et al., 2008; Hisagi et al., 2010; Chládková et al., 2013). The effect of native phoneme inventory on both vowel quality and vowel length processing is indisputable, however, it has not yet been shown how the neural processing of vowel length and vowel spectral quality compare to one another. The present study therefore aims to investigate and compare the neural processing of vowel duration and vowel quality of adult speakers of a language in which both vowel quality and vowel length have a contrastive role. Obtained results will also show if MMNs evoked by changes in vowel quality and quantity match with the pattern obtained by Ylinen et al. (2005) for plosive consonants, in which greater average MMN was observed in case of a quantity change.

A number of studies exploring the sensitivity to phonemic contrasts have encountered a phenomenon called perceptual asymmetry. Perceptual asymmetries can be observed when participants more readily process or respond to a change when category A is presented before category B than vice versa. Such findings imply that the perceptual space differs from the physical space and that due to its asymmetric nature its properties cannot be captured by Euclidean geometry (e.g., distances in the vowel formant space). Asymmetry in perception has been investigated for various types of stimuli including color, line orientation, numbers (Rosch, 1975), geometric figures (Tversky and Gati, 1978) as well as vowels (Polka and Bohn, 2003, 2011; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015), and consonants (Schluter et al., 2016; Cummings et al., 2017; Højlund et al., 2019). Vowel perception asymmetry has been studied by means of reaction time or accuracy in discrimination tasks, where a reversed order of stimuli led to the significant difference in the measured parameters. Asymmetrical perception of vowels has also been attested in neurolinguistic MMN studies, when the roles of standard and deviant stimuli were switched (Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015). For instance, De Jonge and Boersma (2015) found asymmetrical patterns in vowel perception when comparing MMN responses of French listeners to contrasts among four French vowels [y, u, ø, o]. Their results showed that the MMN evoked by a change from a high vowel such as [u] toward a high-mid vowel such as [o], and by a change from a back vowel such as [u] to a front vowel such as [y] was significantly larger (i.e., more negative) than vice versa. In addition to the asymmetry, they found that the average MMN resulting from a change in vowel place (backness or frontness) was significantly larger compared to the MMN resulting from a change in vowel height.

There are several hypotheses and theories that offer explanation to the perceptual asymmetry phenomena. According to Repp and Crowder (1990), perceptual asymmetries are caused by different rates of memory decay, which, as the authors argued, is slower for more prototypical (or less ambiguous) vowels. They concluded that at either point of a vowel continuum the difference between stimuli is more detectable when the more salient vowel comes second in a pair, and thus serves as the subject of comparison.

Polka and Bohn (2003, 2011) proposed the natural referent vowel (NRV) framework which operates with the concept of peripheral vowels and aims to explain language-general, i.e., auditorily-based, patterns in infant speech perception. Peripherality acoustically coincides with formant focalization, that is the convergence of two formant frequencies in a vowel (Schwartz et al., 2005). In a focal vowel, the proximity of two formants strengthens their respective amplitudes and results in a perceptually prominent frequency band. According to the NRV framework, a difference is more detectable for a change from a less peripheral, or non-focal, to a more peripheral, or focal, vowel than vice versa. Along those lines, the difference between two vowels such as [u] and [y] should be more readily detectable, i.e., perceived as greater, when [y] is presented before [u] than vice versa. Note that such NRV-based asymmetry is opposite to the asymmetries obtained by De Jonge and Boersma (2015) who tested adults (and it is opposite also to the asymmetries obtained by Wanrooij et al., 2014 for infants). Although not originally proposed as an explanation for asymmetries in the *neural* processing of vowels, it seems viable that a more detectable difference between stimuli leads to a stronger MMN response (as shown by e.g., Näätänen et al., 1997). Therefore, the NRV can be used to formulate acoustically-based predictions for MMN such that a focal (i.e., perceptually more salient) deviant should elicit a stronger MMN than a non-focal deviant.

Repp and Crowder as well as Polka and Bohn have based their theories of vowel perception asymmetry on the acoustic properties of vowels, while other authors, namely, Lahiri and Reetz (2002) have approached this phenomenon from the phonological point of view and formulated the featurally underspecified lexicon (FUL) theory. Their theory explains the perceptual asymmetries through reference to phonological representations, postulating that a change from a stimulus specified for a particular phonological feature to a stimulus underspecified for that feature is processed more strongly than a change in the reversed order. The predictions of the FUL theory have been borne out by a number of studies (Eulitz and Lahiri, 2004; Lipski et al., 2007; Scharinger et al., 2012, 2016; De Jonge and Boersma, 2015; Schluter et al., 2016).

Considering a vowel contrast such as one between a focal and phonologically specified /a/ and a non-focal and underspecified /ɛ/, one can see that an NRV-like asymmetry predicted by acoustics (i.e., a stronger response to a change from /ɛ/ to /a/) does not necessarily coincide with an asymmetry predicted by the phonological FUL framework (i.e., a stronger response to a change from /a/ to /ɛ/). Crucially, predictions based on phonological representations can also differ depending on the adopted phonological theory. If we again consider the vowels /a/ and /ɛ/, then according to the FUL theory, /ɛ/ is underspecified for feature [LOW]. However, in Element theory (Harris and Lindsey, 1995) which describes vowels in terms of elements |A|, |I|, and |U|, it is /a/ that contains 1 element and is thus underspecified in comparison to /ɛ/ which contains 2 elements. Consequently, one could hypothesize that it is /a/ and not /ɛ/ that should evoke greater MMN response when presented as a deviant. Although the predicted perceptual (MMN) asymmetries differ across phonological frameworks, they have been mainly tested within the FUL framework. An exception is De Jonge and Boersma (2015) who contrasted FUL and Element theory and whose MMN data from French adults supported FUL. Because it is the most widely researched phonological framework in the MMN literature, the present study adopts FUL as the basis for phonological predictions and contrasts it with NRV-like acoustic predictions.

As introduced above, the present experiment focuses on the MMN to vowel quality and vowel length contrasts which are both phonemic in the listeners' native language, Czech. The specific contrasts are /ɛ/-/a/ and /ɛ/-/ɛː/, for vowel quality and vowel length, respectively. Since spectrum can be a secondary perceptual cue to vowel length, we have selected the /ɛ/-/ɛː/ pair out of the five short-long pairs in Czech because it entails the smallest spectral difference, both in perception (Podlipský et al., 2019) and production (Paillereau and Chládková, 2019). Besides comparing the strength of the MMN elicited by the two distinct types of phonemic changes, the present experiment tests whether any MMN asymmetries exist for those vowel contrasts and if yes, whether they are phonologically or acoustically motivated.

In order to provide a further test of whether any potential asymmetries are more likely attributable to the phonology or to the acoustics, we compare Czech listeners' processing of the two vowel contrasts /ɛ/-/a/ and /ɛ/-/ɛː/ to their processing of identical acoustic differences in non-speech stimuli. The non-speech stimuli are inharmonic tone complexes with the first three formant frequencies and duration identical to those of the vowels /a/, /ɛ/, and /ɛː/; they are thus comparably complex as the vowels but not confusable with speech. If the potential asymmetries are acoustically conditioned, they should be found in both the non-speech and the speech conditions in the present study. If, on the contrary, the asymmetries are (at least to some extent) phonologically based the pattern of results should differ across speech and non-speech.

According to Polka and Bohn (2003, 2011), the acoustic properties of our stimuli predict a greater MMN when a focal vowel (or tone complex) is the deviant and a non-focal vowel (or tone complex) is the standard. In that respect, the vowel /a/ and the /a/-like tone are focal because their first and second formants are close to one another, concentrating (focalized) energy in the F1–F2 frequency band. In contrast, the first and second formants of the vowel /ɛ/ and the /ɛ/-like tone are relatively far apart and thus contain non-focalized energy. Acoustically, the change from the non-focal /ɛ/ (-like tone) to the focal /a/ (-like tone) should elicit a stronger MMN response than a reverse change. As for the durational dimension, for which focalization has not been formally defined, intuitively a longer stimulus is more prominent than a shorter stimulus. The acoustically-motivated prediction then is that a change from the short /ɛ/ (-like tone) to the long /ɛ/ (-like tone) will elicit a greater MMN than vice versa. This direction of predicted asymmetry is further in line with previous findings that the addition of information is more detectable than its deletion (Timm et al., 2011).

The alternative, phonologically-based predictions for asymmetries are made in accordance with the featural (under)specification framework (Lahiri and Reetz, 2010), which states that the magnitude of the MMN will be greater in case of change from a fully specified vowel to an underspecified vowel than vice versa. Czech central low vowel /a/ and front mid vowel /ɛ/ differ both in the horizontal plane and in height, nevertheless from the phonological point of view there are distinguished only by means of the feature [LOW] (which is specified for /a/ but not for /ɛ/) as they are both underspecified with respect to the feature [BACK]. Therefore, in conformity

TABLE 1 | Acoustically- and phonologically-based predictions of relative magnitude of the MMN response to the experimental stimuli.

	Direction of the MMN asymmetry		
Acoustics (NRV)	[ɛ] → [a] [ɛ] → [ɛː]	>	[a] → [ɛ] [ɛː] → [ɛ]
Phonology (FUL)		<	

with the FUL theory, we expect a greater MMN response when underspecified /ɛ/ is a deviant. Regarding the quantity contrast, according to some authors the difference between Czech /ɛ/ and /ɛː/ lies in the feature [LONG], which is specified for /ɛː/ (Palková, 1994, p. 206, Skarnitzl et al., 2016, p. 101). This means that in the vowel quantity condition, /ɛ/ is again underspecified, and the MMN should be larger when /ɛ/ is a deviant and /ɛː/ is a standard.

Predictions of the vowel perception asymmetry in terms of relative magnitude of the MMN response are summarized in **Table 1**. For the complex tone stimuli, the asymmetrical behavior is expected based solely on the acoustical approach, and thus coincides with the first row of **Table 1**.

To sum up, the present study has two goals. Firstly, it compares the neural processing of vowel length and vowel quality in a language that uses both types of contrasts phonemically [similarly to the comparison of consonantal quality and consonantal length reported by Ylinen et al. (2005)]. Secondly, it tests whether there are any directional asymmetries in the perception of vowel length and/or vowel quality and whether they can be explained by the vowels' acoustic properties or phonological specification.

MATERIALS AND METHODS

Stimuli

We created two sets of stimuli, one set for the speech condition and one set for the non-speech condition. The speech stimuli were naturally produced, edited consonant-vowel (CV) syllables [fɛ] and [fa]. The formants were stable throughout the vowels and corresponded to the Czech low-mid front /ɛ/ and low /a/, respectively. The first three formants of [ɛ] in [fɛ] were 755 Hz, 1646 Hz, and 2710 Hz, and the first three formants of [a] in [fa] were 864, 1287, and 2831 Hz; these values are in line with the formants of Czech vowels produced by women reported by Skarnitzl and Volín (2012). The duration of the vowels [ɛ] and [a] (extracted from the CV frames) was modified using PSOLA in Praat (Boersma and Weenink, 1992–2020). The vowel [a] had a duration of 220 ms, and [ɛ] was resynthesized with three durations, namely, 220, 180, and 360 ms, which met the following conditions: 220 ms was judged (by three expert phoneticians) as a typical duration of the mid and low short vowels in an isolated CV syllable, 360 ms represented a long vowel in a CV syllable that was not perceived as unnaturally exaggerated, and short /ɛ/ with the duration of 180 ms was considered to be sufficiently distinct from the long /ɛː/.[1] In order to create the stimuli, we cut out the initial fricative consonant [f] from one recorded syllable and combined it with the target [a] and [ɛ] vowels, such that the fricative [f] was identical across all four speech stimuli and had a duration of 150 ms. None of the created [f] + V syllables carries lexical or morphological content in Czech. The speech stimuli had been used in a behavioral study on vowel perception with Czech-exposed infants (Paillereau et al., 2021), and recently, along with the non-speech stimuli described below, in an ERP study with Czech newborns (Chládková et al., under review).

To test the discrimination of a spectral contrast, the non-focal [fɛ] and the focal [fa] lasting for 220 ms each were used. The vowel [a] is considered focal because the distance between its first and second formant is d_a = 2.07 Bark, while the vowel [ɛ] in [fɛ] is non-focal because its first two formants are spread apart by d_ε = 4.08 Bark. The difference between [a] and [ɛ] thus lies in their perceptual prominence, where [a] is the more prominent one. The discrimination of a durational contrast was tested by the short 180-ms [fɛ] and long 360-ms [fɛ]. Similarly as for the spectral dimension, the short and the long vowel differ in their perceptual prominence, where the short one contains energy over a shorter time interval (i.e., less energy in total) as can thus be seen as perceptually less prominent stimulus than a long vowel represented by energy in a longer time interval. The intensity of the stimuli was scaled by peak to be matched across all the 4 different syllables.

The non-speech stimuli were inharmonic tone complexes with spectral and durational properties mimicking those of the vowels described above. Inharmonic tone complexes are comparably complex as vowels in that their source signal contains a series of fundamental frequency harmonics and is filtered with vocal-tract like formants. At the same time, the inharmonic tone complexes are not confusable with vowels because their source signal frequencies are spaced inharmonically (Goudbeek et al., 2009; Scharinger et al., 2014). The tone complexes in the present experiment had 15 inharmonically spaced frequency components, the first one at 500 Hz and every following being 1.15 times higher. The inharmonic source signal was filtered with three formants, namely, for the focal spectral condition with the formants of [a], for the non-focal spectral condition and the short and long durational condition with the formants of [ɛ]. Durations of the non-speech stimuli were identical to the durations of the vowels from the speech condition. The amplitude was ramped linearly over 5 ms at stimulus onset and offset. Sound intensity was scaled to be identical across all the four stimuli. As in the speech condition, the [a]-like focal tone (prominent) and the [ɛ]-like non-focal 220-ms tone (non-prominent) were used to test discrimination of spectral differences, and the 180-ms [ɛ]-like tone (non-prominent) and the 360-ms [ɛ]- like tone (prominent) were used to test discrimination of duration differences.

Presentation Paradigm

The stimuli, i.e., the individual syllables or the individual tone complexes, were presented in a roving-standard paradigm (Haenschel et al., 2005; Garrido et al., 2008; Cooper et al., 2013). Four presentation blocks were created, one for each domain

[1]We did not adopt the 220-ms stimulus as a short counterpart of the 360-ms /ɛ/ because the resulting long/short ratio 1.6 is more typical of the high front Czech vowel pair while for mid-low vowels the ratio is closer to 2 (Paillereau and Chládková, 2019).

(speech and non-speech) and dimension (spectrum and duration) combination. For speech spectrum, the paradigm started with 8 tokens of [fɛ] and continued with 100 trains of [fɛ] and [fa] each, alternating in series of 4–8 identical stimuli. The count of 4–8 was pseudorandom, fulfilling the condition that each count eventually occurred 20 times. The number of presented tokens was 608 for [fɛ], and 600 for [fa]; summing up to a total of 1208 stimuli in each block. Stimulus onset asynchrony was 1.09 s. Total presentation time per block was 22 min. The blocks for speech duration were created in an identical way, alternating series of short [fɛ]s and the long [fɛ:]s. Analogous presentations were made for non-speech spectrum and non-speech duration. Each participant was tested with either the two speech blocks, or the two non-speech blocks. Stimulus domain thus varied between participants and dimension within participants, with the order of durational and spectral presentation counterbalanced.

Participants and Procedure

A total of 32 adult volunteers participated in the experiment. They were monolingually-raised native speakers of Czech, ages 18–28 years (mean age 24 years, 19 women, 13 men). They did not have any history of neurological or hearing disorders and reported to be right-handed.

Participants were tested in a quiet room at the Faculty of Medicine in Hradec Králové. Prior to the experiment, they filled in a demographic background questionnaire and signed an informed consent form. Half of the participants was randomly assigned to the speech condition and the other half to the non-speech condition. Within each condition, a participant received two blocks, one presenting changes in stimulus duration and the other with changes in stimulus spectral quality; the order of the blocks was counterbalanced across participants. Between the two blocks, there was a 5-min break. During auditory stimulation, participants watched a muted movie with Czech subtitles. Participants were instructed to focus on the movie and ignore the sounds. The experiment followed the standards for research with humans and was approved by the ethics committee of the Faculty of Medicine in Hradec Králové.

Electroencephalography and ERP Processing

The EEG was recorded from thirty one Ag/AgCl electrodes Fp1, Fp2, F7, F3, Fz, F4, F8, CP4, C3, Cz, C4, TP8, FT7, P3, Pz, P4, FC3, FC4, FT8, M1, M2, OPz, AFz, P7, P8, T7, T8, CPz, FCz, TP7, CP3 referenced to an electrode placed on the nose. The EEG was recorded at a 3000-Hz sampling rate with a bandwidth of 0.3–100Hz (DEYMED Diagnostic s.r.o., Czechia). After band-pass filtering 0.2–40 Hz using EEGLab (Delorme and Makeig, 2004), the data were down-sampled to 300 Hz and epoched with MATLAB release 2020a (MathWorks, United States). The epoch started 100 ms before and ended 800 ms after the onset of the vowel or the onset of the complex tone; mean voltage of the prestimulus part (from −100 to 0 ms) was subtracted from every epoch.

Deviant waveforms were derived from every first stimulus in the row of 4–8 repeated tokens, standard waveforms were derived from the last two stimuli in the row of 4–8 repeated tokens. Standard and deviant grand-average waveforms at central channels and the MMN topographies are shown in **Figure 1**. The individual ERPs were calculated as an average of epochs with absolute amplitude under 50 μV. The ERPs were additionally digitally filtered off-line by a smoothing Savitzky-Golay filter (first polynomial order, window of 21 samples).

Difference waves were computed by subtracting the averaged standard ERP from the averaged deviant ERP elicited by physically identical stimuli, e.g., the difference waveform for the [a]-deviant was computed by subtracting the [a]-deviant ERP from the [a]-standard ERP. From the difference waves, the MMN was quantified as area under curve in a pre-defined 100-ms window that started 150 ms after change onset. The window of analysis was determined based on previously published results (Näätänen et al., 1997, 2004; Eulitz and Lahiri, 2004; De Jonge and Boersma, 2015) and visual inspection of the curves, and thus has been set 150–250 ms after vowel or tone onset for the spectral condition and 330–430 ms after vowel or tone onset for the durational condition (where the onset of change was determined as the duration of the short vowel/tone, i.e., 180 ms).

Statistical Analyses

The calculated AUC were analyzed with a linear mixed-effects model (packages lme4, lmerTest in R, Bates et al., 2015; R Core Team, 2016; Kuznetsova et al., 2017). We modeled the main effects and all two- and three-way interactions of Domain (−speech, +non-speech), Dimension (−duration, +spectrum), and Deviant (−prominent, +non-prominent), as well as the main effects of Laterality (2 contrasts: −left +right, −lateral +midline) and Anteriority (2 contrasts: −central +frontal, −central +parietal). The random effects structure modeled a per-participant intercept and slopes for Dimension and Deviant.

RESULTS

The summary of the modeled fixed effects is presented in **Table 2**. As indicated by the significant intercept, overall there was a reliable MMN, estimated as -48 ± 15 μV × ms ($p = 0.003$). The two main effects for Anteriority suggest that the MMN was stronger (more negative) at frontal than at central sites, where it in turn was stronger than at parietal sites, thus following the expected frontally-localized distribution of the auditory and linguistic MMN response.

Regarding the predictors relevant for our research questions, there was a three-way interaction of Deviant, Dimension, and Domain[2]. To unpack the triple interaction, **Figure 2** visualizes

[2]A reviewer expressed concerns about a potentially low power of our experiment. We therefore simulated the power curves associated with an effect equal to the one we obtained, as well as a smaller effect, using the simr package in R (Green and MacLeod, 2016). For the simulations, we created a new model using the parameters of the initial model and calculated its power for various number of respondents for the effect of three-way interaction of Deviant, Dimension, and Domain. The

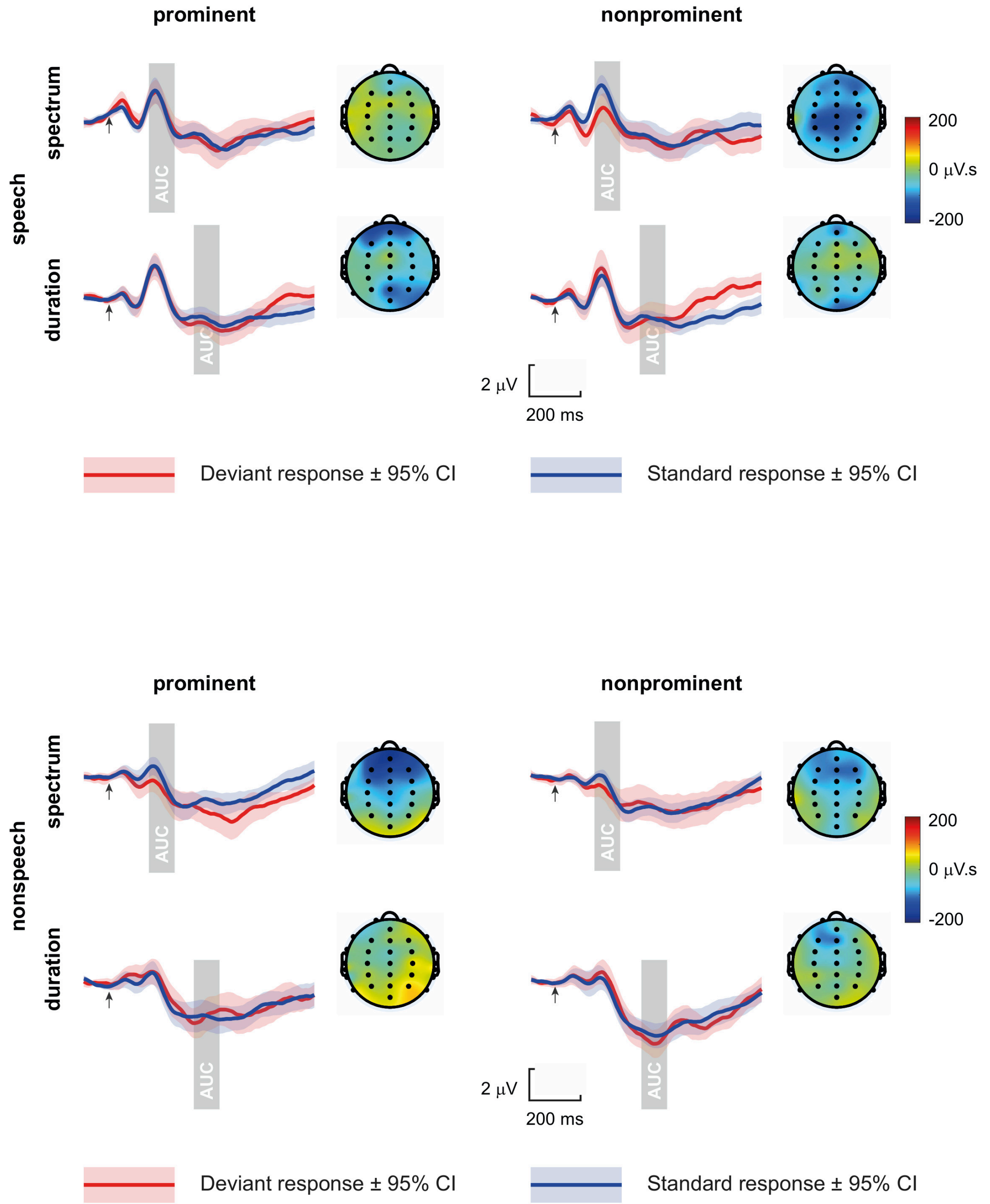


FIGURE 1 | Standard and deviant grand-average waveforms at central channels (averaged across C3, Cz, and C4), and the MMN topographies (displaying the area under curve, AUC, measured in the shaded time windows from deviant-standard differences), per Domain, Dimension, and Deviant type (arrows mark tones/vowels onset).

the estimated means and confidence intervals [modeled using the R package ggeffects, Lüdecke (2018)]. Pairwise comparisons of the two deviant types on each dimension and in each domain reveal that an asymmetry between the two deviants was found in speech for the spectral contrast: [fa] elicited a stronger MMN than [fɛ] {[fa] mean = −95 µV × ms, CI = (−164; −27), [fɛ] mean = −17 µV × ms, CI = (−84; 49)}; in all other conditions the MMNs elicited by the two deviant types overlapped (i.e., the 95% CI's of one deviant contained the mean of the other deviant, which implies that the difference is not significant at alpha 0.05).

obtained power curve indicated that to reach power of 80%, even with the smaller effect size (i.e., the lower bound of 95% CI of the mean estimated effect in our study) for the critical three-way interaction, a total of 20 participants (i.e., 10 per group) would be sufficient. From this we conclude that our experiment with 32 participants, i.e., 16 per group, is not underpowered.

TABLE 2 | Fixed-effects summary of the model outcomes.

Predictor	Estimate	SE	df	t	p
Intercept	**−47.999**	**15.150**	**31.738**	**−3.168**	**0.003**
Deviant (−prominent +non-prominent)	12.534	27.848	31.802	0.450	0.656
Dimension (−duration +spectrum)	−31.456	26.836	31.281	−1.172	0.250
Domain (−speech +tone)	4.757	30.299	31.738	0.157	0.876
Laterality (−left +right)	8.084	10.354	1057.792	0.781	0.435
Laterality (−lateral +midline)	−19.745	11.956	1057.792	−1.652	0.099
Anteriority (−central +frontal)	**−46.064**	**11.956**	**1057.792**	**−3.853**	**<0.001**
Anteriority (−central +parietal)	**30.782**	**11.956**	**1057.792**	**2.575**	**0.010**
Deviant × Dimension	17.550	17.138	1068.149	1.024	0.306
Deviant × Domain	−18.349	55.695	31.802	−0.329	0.744
Dimension × Domain	−38.804	53.672	31.281	−0.723	0.475
Deviant × Dimension × Domain	**−189.978**	**34.275**	**1068.149**	**−5.543**	**<0.001**

Rows marked in bold indicate the effects with $p < 0.05$.

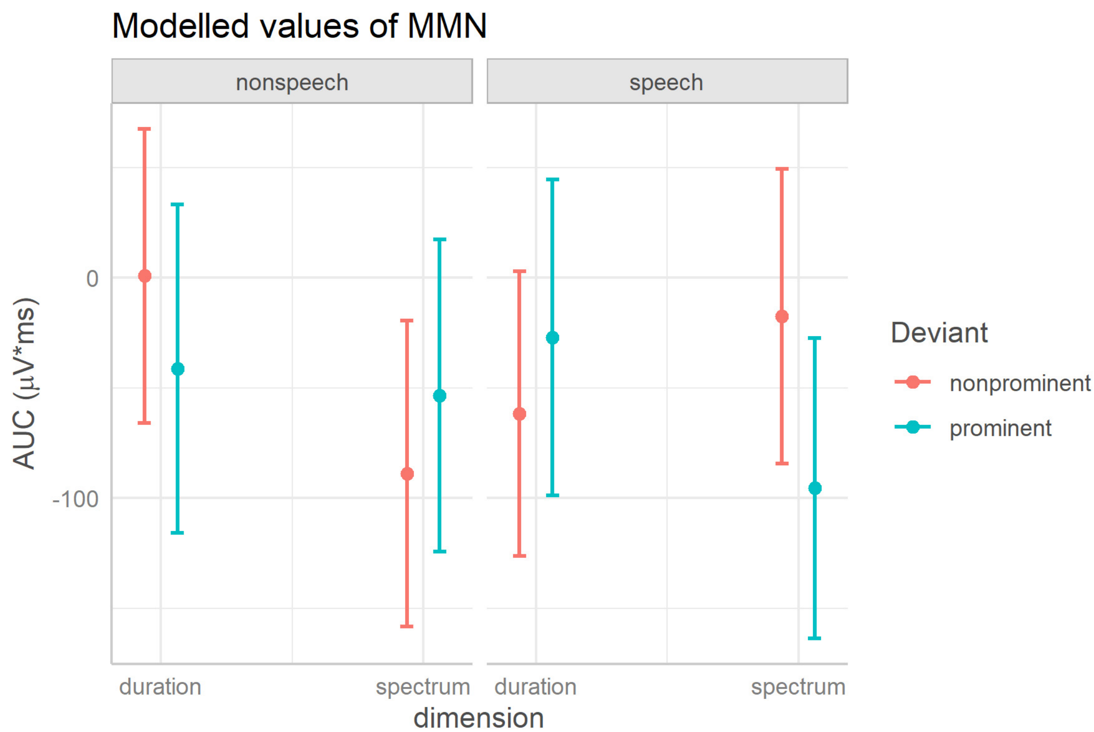


FIGURE 2 | Unpacking the significant three-way interaction of Deviant, Dimension, and Domain. The figure shows model-estimated means and 95% confidence intervals for the MMN elicited by acoustically prominent and non-prominent deviants on each dimension, separately in speech and non-speech stimuli.

DISCUSSION

The first question addressed by this experiment was whether the neural processing of phonemic vowel quality differs from the neural processing of phonemic vowel length. To that end, we assessed the neural mismatch response (MMN) in adult speakers of Czech listening to changes between [fɛ] and [fa] and to changes between [fɛ] and [fɛː] syllables, where both types of change represent a phonological vowel contrast. Our statistical analysis failed to detect a main effect of Dimension (or a two-way interaction of Dimension and Domain). A planned comparison of the MMN elicited by vowel quality (mean = −56 µV × ms, CI = [−111, −2]) and the MMN elicited by vowel length (mean = −44 µV × ms, CI = [−99, 11]) suggests a large overlap across the two types of vowel change, lending support to the conclusion that vowel length and vowel quality changes evoke comparable neural response in Czech adult listeners. Our results for vowels are thus different than the MMN patterns observed by Ylinen et al. (2005) for length and quality changes in plosive consonants.

If we consider the spectral and durational difference between the stimuli in just-noticeable difference units (JND), the Euclidean distance between the first three formants of the [a] and [ɛ] stimuli is equal to 5.1 JND, whereas the durational difference between the [ɛ] and [ɛ:] stimuli equals 12.8 JND [JNDs computed assuming the discrimination threshold of 0.3 bark for vowel formants, Kewley-Port (2001) and a 5 ms discrimination threshold to the reference value of 90 ms for vowel duration, Nooteboom and Doodeman (1980)]. Even though the JND in duration is more than 2 times greater than the JND in spectrum, the average MMNs elicited by each of the changes were not found to differ. Speculatively, this could be taken as an indication that the contrasts have been processed based on their phonological difference rather than the acoustic distance.

The second aim of the experiment was to test whether the vowel contrasts are processed asymmetrically, and if yes, whether the asymmetries are attributable to the acoustic or the phonological properties of the vowels. To that end, we compared the MMN elicited by changes in vowels to the MMN elicited by identical changes in non-speech stimuli. Regarding the spectral contrast, an acoustically-based approach formulated under the NRV framework (Polka and Bohn, 2003, 2011) predicted a larger MMN in case of vowel change from /ɛ/ to /a/ than vice versa. When comparing vowels /ɛ/ and /a/, the latter one is auditorily focal, or perceptually more salient, since its first and second formants are close to each other such that they merge into one prominent frequency band. In contrast, the first two formants of /ɛ/ are farther apart, resulting in vowel /ɛ/ assigned to the non-focal, perceptually less prominent, element of the comparison. Thus, under the acoustically-based approach, we expected a larger MMN when /a/ was the deviant, and smaller MMN is expected when /ɛ/ was the deviant in the present experiment. Concerning the durational difference in vowels, a long vowel, here /ɛ:/, contains acoustic energy over a longer time interval, and is thus inherently more auditorily prominent than a short vowel of the same quality, here /ɛ/. Therefore, for the change between /ɛ/ and /ɛ:/, the acoustically-based approach predicted greater MMN when the long /ɛ:/ was the deviant than when the short /ɛ/ was the deviant. Crucially, if perceptual asymmetries in vowels were acoustically conditioned, the same asymmetries were expected to be observed in the non-speech condition, which compared MMN to the changes between /ɛ/-like and /a/-like complex tones, as well as between /ɛ/-like and /ɛ:/-like complex tones. Alternatively, if any detected asymmetries did not conform to the acoustically-motivated predictions, or were not detectable in the non-speech stimuli, they could be attributable to the linguistic status of the vowels. The specific phonologically-based predictions were formulated in line with the FUL (Lahiri and Reetz, 2002, 2010), and predicted an opposite direction of asymmetry due to the phonological feature specification in vowel height. Since /a/ is specified for feature [LOW] and /ɛ/ is fully underspecified, greater MMN response was expected when /ɛ/ served as deviant than vice versa. As for the durational contrast, asymmetry would be caused by feature [LONG], which is specified for /ɛ:/ but not for /ɛ/, therefore predicting greater MMN response for the short vowel /ɛ/ deviant.

The statistical model revealed a significant triple interaction of Deviant, Domain and Dimension. Pairwise comparisons of the MMN across the two directions of change (i.e., the two deviants) within each condition (i.e., for each dimension and each domain) revealed an MMN asymmetry for the spectral contrast in speech. A change from [fɛ] to [fa] elicited a stronger MMN than a change from [fa] to [fɛ] (no other asymmetries were detected). On the one hand, this result shows that a change from a non-prominent to a prominent vowel is better detectable than a reverse change, which is in line with the acoustically-motivated predictions within the NRV framework and would favor an acoustically-based explanation for the asymmetry. On the other hand, however, this asymmetry was not detected in the non-speech condition where the stimuli differed in identical acoustic parameters as did the stimuli in the speech condition. Due to its lack in the non-speech condition, we conclude that the asymmetry that we found in the processing of the spectral vowel contrast between /a/ and /ɛ/ is specific to speech and cannot be entirely acoustically based.

Another factor suggesting that the phonologically-motivated explanation for the present MMN asymmetry is more plausible is the duration of stimulus-onset asynchrony (SOA) in our experimental paradigm. SOA was fixed at 1.09 s, which is relatively long, and therefore was more likely to tap into phonological rather than purely acoustic processing (Werker and Logan, 1985). Johnson (2015) addressed the predictions for perceptual vowel asymmetries made by the acoustic and phonological frameworks and has shown that the pattern of vowel perception asymmetry is modulated by the experimental setting. He explored perceptual asymmetries in vowels via reaction time in two discrimination tasks differing in the inter-stimulus interval (ISI), where short ISI (100 ms) implied lower-level auditory listening conditions and long ISI (700 ms) induced higher-level phonemic listening conditions. The results of Johnson's experiments indicated that the phonological underspecification model of Lahiri and Reetz (2002, FUL) accurately predicted the direction of vowel perception asymmetry in the phonemic conditions, and that in the auditory listening task this direction was reversed, and instead could be explained by the hypotheses employing acoustic characteristics of sounds. Here, we uncover an asymmetry in the processing of vowel quality but did not to detect it in a comparable non-speech condition, with a same, in Johnson's terms relatively long, ISI across the two conditions (the ISI being 730 or 910 ms depending on vowel/tone duration). It therefore appears that the asymmetry we detected for a spectral contrast in vowels is likely, at least in part, phonologically based.

However, the present asymmetry with a change from [fɛ] to [fa] eliciting a stronger MMN than vice versa, is opposite to what FUL would predict. Yet it is still possible that an underspecification account be compatible with such a finding if one considers not only the backness feature (as done in most previous MMN studies testing the FUL theory) or if one sees feature specifications as language specific.

The Czech vowels /a/ and /ɛ/ do not differ only in their featural specification of height as we considered (in line with previous studies on similar vowel contrasts in other languages, e.g., /ae/ vs. /ɛ/ in Scharinger et al., 2012), but also in their featural specification of place. One could thus argue that it was the (under)specification of vowel place rather than vowel height that caused the present perceptual MMN asymmetry. The feature [FRONT] is likely specified for Czech /ɛ/ but not necessarily for Czech /a/ because in the vowel system of Czech, /a/ (along with its long counterpart) is the only low vowel does not need to be contrasted by the feature place with another low vowel quality (unlike for the mid front vowel /ɛ/ which contrasts with the mid back vowel /ɛ/). The explanation that Czech listeners responded more strongly to a mismatch in the phonological specification of vowel place than to a mismatch in the phonological specification of vowel height would also be partially in line with the results of De Jonge and Boersma (2015) who examined MMN asymmetries in French listeners. Those authors found out that the changes between French front rounded and back vowels evoked greater MMN than did the changes between high and mid-high vowels, which indicates that the horizontal difference (in place) between vowels is more salient than the vertical difference (in height).

It is possible that for the Czech /a/-/ɛ/ contrast a place mismatch is more relevant than a height mismatch, or, that both are relevant phonologically but in the case of the stimuli used here, the place mismatch overrode the height mismatch. Comparing the F1 and F2 of the vowels used in the present experiment, it can be seen that the relative distance between the first formants of [a] and [ɛ] is less (namely, 2.07 bark) than the relative distance between the second formants of [a] and [ɛ] (namely, 4.08 bark). Although phonological specification operates on discretized entities, which means that the raw acoustic distance should not matter for whether or not a phonological category contrast is perceived, MMN amplitude is modulated both by linguistic and acoustic differences between standard and deviant stimuli (e.g., Näätänen et al., 1997; Phillips et al., 2000). Therefore, the apparent prime role of underspecification of vowel *place* (rather than vowel height) might as well be, at least partially, driven by the fact that the change in phonological place between the /a/ and the /ɛ/ was acoustically almost twice as large as the change in phonological height (i.e., 4.08 bark versus 2.07 bark). All in all, if phonological underspecification is extended to vowel place, the present results are explainable as phonologically conditioned asymmetries.

CONCLUSION

Pre-attentive processing of changes in phonemic vowel length and vowel quality by adult Czech speakers was assessed in an ERP experiment. The neural mismatch response (MMN) elicited by a change in vowel length between /ɛ/ and /ɛ:/ was comparable to the MMN elicited by a change in vowel quality between /ɛ/ and /a/, suggesting that both types of phonemic changes are equally salient to Czech speakers. For the vowel quality contrast, a perceptual asymmetry was detected where a larger MMN response was found to a change from /ɛ/ to /a/ than vice versa. No such asymmetrical pattern was observed in non-speech stimuli differing in the same acoustic parameters as the vowels, which indicated that the vowel asymmetry is more likely attributable to the vowels' linguistic status, namely phonological feature specification, than (purely) to the vowel acoustics. A stronger MMN for the vowel spectral change was elicited by a switch from /ɛ/ to /a/ than vice versa, from which we have inferred that for this Czech vowel contrast it is the feature specification for place which is primarily exploited by language users. We argued that it might have been a (language-specific) underspecification in terms of place for /a/ (rather than universal underspecification in terms of height for /ɛ/, assumed by the FUL, Lahiri and Reetz, 2002, 2010) which caused that listeners more readily detected a change from a FRONT /ɛ/ to an underspecified /a/ than vice versa.

DATA AVAILABILITY STATEMENT

The data supporting the conclusions of this article and the associated analysis scripts are available from the OSF website at https://osf.io/2849m/. The raw EEG data will be made available by the authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University Hospital Hradec Králové Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JU, JK, and KC designed and implemented the experiment. JU and ZO performed the data collection. NN, JK, and KC processed and analyzed the data. NN wrote the manuscript with contributions and edits from KC, JK, JU, and ZO. All authors contributed to the article and approved the submitted version.

FUNDING

This project was funded by the Charles University grant Primus/17/HUM/19 and project Progres Q40/7. KC and NN were also funded by the Czech Science Foundation grant 18-01799S.

ACKNOWLEDGMENTS

The authors thank Petr Voda for his technical support.

REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Statist. Software* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (1992–2020). *Praat: Doing Phonetics by Computer*. Available online at: http://www.praat.org (accessed date 22 November, 2018).

Chládková, K., Escudero, P., and Lipski, S. C. (2013). Preattentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain Lang.* 126, 243–252. doi: 10.1016/j.bandl.2013.05.020

Chládková, K., Urbanec, J., Skálová, S., and Kremláček, J. (under review). Newborns' neural processing of native vowels. *Manuscript Under Rev*.

Cooper, R. W., Atkinson, R. A., Clark, R. A., and Michie, P. T. (2013). Event-related potentials reveal modelling of auditory repetition in the brain. *Internat. J. Psychophysiol.* 88, 74–81. doi: 10.1016/j.ijpsycho.2013.02.003

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

De Jonge, M. J., and Boersma, P. (2015). French high-mid vowels are underspecified for height. in Proceedings of the 18th International Congress of Phonetic Sciences (Glasgow: The University of Glasgow).

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–12. doi: 10.1016/j.jneumeth.2003.10.009

Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* 42, 936–944. doi: 10.1016/j.neuroimage.2008.05.018

Goudbeek, M., Swingley, D., and Smits, R. (2009). Supervised and Unsupervised Learning of Multidimensional Acoustic Categories. *Journal of Experimental Psychology: Human Perception and Performance* 35, 1913–1933. doi: 10.1037/a0015781

Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., and Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *J. Neurosci.* 25, 10494–10501. doi: 10.1523/jneurosci.1227-05.2005

Harris, J., and Lindsey, G. (1995). "The elements of phonological representation," in *Frontiers of phonology: atoms, structures, derivations*, eds J. Durand and F. Katamba (Harlow: Longman), 34–79.

Hisagi, M., Shafer, V. L., Strange, W., and Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Res.* 1360, 89–105. doi: 10.1016/j.brainres.2010.08.092

Højlund, A., Line Gebauer, L., McGregor, W. B., and Wallentin, M. (2019). Context and perceptual asymmetry effects on the mismatch negativity (MMNm) to speech sounds: an MEG study. *Lang. Cogn. Neurosci.* 34, 1–16. doi: 10.1080/23273798.2019.1572204

Johnson, K. (2015). Vowel Perception Asymmetry in Auditory and Phonemic Listening. *UC Berk. PhonLab Ann. Rep.* 2015:11.

Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *J. Acoust. Soc. Am.* 110, 2141–2155. doi: 10.1121/1.1400737

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., and Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *Internat. J. Psychophysiol.* 67, 131–143. doi: 10.1016/j.ijpsycho.2007.10.012

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. *J. Statist. Software* 82, 1–26. doi: 10.18637/jss.v082.i13

Lahiri, A., and Reetz, H. (2002). Underspecified recognition. *Lab. Phonol.* 7, 637–675. doi: 10.1515/9783110197105.2.637

Lahiri, A., and Reetz, H. (2010). Distinctive features: Phonological underspecification in processing. *J. Phonet.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lipski, S. C., Lahiri, A., and Eulitz, C. (2007). Differential height specification in front vowels for German speakers and Turkish-German bilinguals: an electroencephalographic study. in *Proceedings of the International Congress of Phonetic Sciences XVI*, 809–812 (Saarbrücken).

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *J. Open Sour. Software* 3:772. doi: 10.21105/joss.00772

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115, 140–144. doi: 10.1016/j.clinph.2003.04.001

Nooteboom, S. G., and Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *J. Acoust. Soc. Am.* 67, 276–287. doi: 10.1121/1.383737

Paillereau, N., and Chládková, K. (2019). Spectral and temporal characteristics of Czech vowels in spontaneous speech. *AUC Philologica* 2019:19. doi: 10.14712/24646830.2019.19

Paillereau, N., Podlipský, V. J., Šimáčková, Š, Smolík, F., Oceláková, Z., and Chládková, K. (2021). Perceptual sensitivity to vowel quality and vowel length in the first year of life. *JASA Exp. Lett.* 1:025202. doi: 10.1121/10.0003369

Palková, Z. (1994). *Fonetika a fonologie češtiny*. Prague: Karolinum.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567

Podlipský, V. J., Chládková, K., and Šimáčková, Š (2019). Spectrum as a perceptual cue to vowel length in Czech, a quantity language. *J. Acoust. Soc. Am.* 146, EL352–EL357.

Polka, L., and Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Comm.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *J. Phonet.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

R Core Team. (2016). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna: R Core Team.

Repp, B. H., and Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *J. Acoust. Soc. Am.* 88, 2080–2090. doi: 10.1121/1.400105

Rosch, E. (1975). Cognitive reference points. *Cogn. Psychol.* 7, 532–547. doi: 10.1016/0010-0285(75)90021-3

Scharinger, M., Herrmann, B., Nierhaus, T., and Obleser, J. (2014). Simultaneous EEG-fMRI brain signatures of auditory cue utilization. *Front. Neurosci.* 8:137. doi: 10.3389/fnins.2014.00137

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2012). Asymmetries in the Processing of Vowel Height. *J. Speech Lang. Hear. Res.* 55, 903–918. doi: 10.1044/1092-4388(2011/11-0065)

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

Schluter, K., Politzer-Ahles, S., and Almeida, D. (2016). No place for /h/: an ERP investigation of English fricative place features. *Lang. Cogn. Neurosci.* 31, 728–740. doi: 10.1080/23273798.2016.1151058

Schwartz, J.-L., Abry, C., Boë, L.-J., and Vallée, N. (2005). The dispersion-focalization theory of sound systems. *J. Acoust. Soc. Am.* 117:4. doi: 10.1121/1.4786487

Skarnitzl, R., and Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. (Reference values of vowel formants of young adult speakers of standard Czech.). *Akustické listy* 18, 7–11.

Skarnitzl, R., Šturm, P., and Volín, J. (2016). *Zvuková báze øečové komunikace*. Prague: Karolinum.

Timm, J., Weise, A., Grimm, S., and Schröger, E. (2011). An asymmetry in the automatic detection of the presence or absence of a frequency modulation within a tone: a mismatch negativity study. *Front. Psychol.* 2:189. doi: 10.3389/fpsyg.2011.00189

Tversky, A., and Gati, I. (1978). "Studies of similarity," in *Cognition and Categorization*, eds E. Rosch and B. B. Lloyd (Lawrence: Erlbaum).

Wanrooij, K., Boersma, P., and van Zuijen, T. L. (2014). Distributional Vowel Training Is Less Effective for Adults than for Infants. A Study Using the Mismatch Response. *PLoS One* 9:e109806. doi: 10.1371/journal.pone.0109806

Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percept. Psychophy.* 37, 35–44. doi: 10.3758/bf03207136

Ylinen, S., Huotilainen, M., and Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *Neuroreport* 16, 1857–1860. doi: 10.1097/01.wnr.0000185959.11465.9b

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership