Unified Coding of Spectral and Temporal Phonetic Cues: Electrophysiological Evidence for Abstract Phonological Features

Philip J. Monahan[1,2], **Jessamyn Schertz**[2,3], **Zhanao Fu**[2], **and Alejandro Pérez**[1,4]

Abstract

■ Spoken word recognition models and phonological theory propose that abstract features play a central role in speech processing. It remains unknown, however, whether auditory cortex encodes linguistic features in a manner beyond the phonetic properties of the speech sounds themselves. We took advantage of the fact that English phonology functionally codes stops and fricatives as voiced or voiceless with two distinct phonetic cues: Fricatives use a spectral cue, whereas stops use a temporal cue. Evidence that these cues can be grouped together would indicate the disjunctive coding of distinct phonetic cues into a functionally defined abstract phonological feature. In English, the voicing feature, which distinguishes the consonants [s] and [t] from [z] and [d], respectively, is hypothesized to be specified only for voiceless consonants (e.g., [s t]). Here, participants listened to syllables in a many-to-one oddball design, while their EEG was recorded. In one block, both voiceless stops and fricatives were the standards. In the other block, both voiced stops and fricatives were the standards. A critical design element was the presence of intercategory variation within the standards. Therefore, a many-to-one relationship, which is necessary to elicit an MMN, existed only if the stop and fricative standards were grouped together. In addition to the ERPs, event-related spectral power was also analyzed. Results showed an MMN effect in the voiceless standards block—an asymmetric MMN—in a time window consistent with processing in auditory cortex, as well as increased prestimulus beta-band oscillatory power to voiceless standards. These findings suggest that (i) there is an auditory memory trace of the standards based on the shared [voiceless] feature, which is only functionally defined; (ii) voiced consonants are underspecified; and (iii) features can serve as a basis for predictive processing. Taken together, these results point toward auditory cortex's ability to functionally code distinct phonetic cues together and suggest that abstract features can be used to parse the continuous acoustic signal. ■

INTRODUCTION

Spoken language comprehension requires listeners to recognize words embedded in the continuous acoustic speech signal, which in turn involves identifying the segmental units of words, phonemes (Kazanina, Bowers, & Idsardi, 2018). The identity of different phonemes primarily depends on specific spectral and temporal properties of the auditory signal (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). These are the frequency profile of the acoustic signal and the time intervals of major acoustic events (e.g., aspiration, vowel duration), respectively. In linguistics, these distinctive spectral and temporal characteristics are represented as features (Halle, 2002; Clements & Hume, 1995; Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1961). Ultimately, features are the units of speech sound representations that link articulatory and acoustic characteristics of a given phoneme (Baković, 2014; Halle, 1983). Some phonetic features code for a spectral property, whereas others code for a temporal property. Different features have been postulated to describe speech sounds in the field of linguistics, where the notation [feature] refers to a specific feature. As an example, sounds produced with primary airflow through the nasal cavity (i.e., [m n ŋ]), which causes the attenuation of acoustic energy in higher frequencies, all share the feature [nasal]. In addition to their importance in phonological theory, features have played a central role as perceptual units in various models of spoken word recognition (Hickok, 2014; Poeppel & Monahan, 2011; Gow, 2003; Stevens, 2002; McClelland & Elman, 1986; Halle & Stevens, 1962).

Neuroimaging and neurophysiological studies support the idea that features also play a central role in how auditory cortex represents speech sounds (see the work of Monahan, 2018, for a review). The spectral and temporal characteristics of speech sounds are encoded by different spectro-temporal receptive fields in superior temporal gyrus (STG; Fox, Leonard, Sjerps, & Chang, 2020; Yi, Leonard, & Chang, 2019; Hullett, Hamilton, Mesgarani, Schreiner, & Chang, 2016; Mesgarani, Cheung, Johnson, & Chang, 2014). Different populations in STG code distinct phonetic properties of speech sounds (Mesgarani et al., 2014), including voice onset time (VOT; Fox et al.,

[1]University of Toronto Scarborough, Canada, [2]University of Toronto, Canada, [3]University of Toronto Mississauga, Canada, [4]Cambridge University, United Kingdom

Journal of Cognitive Neuroscience 34:4, pp. 618–638
https://doi.org/10.1162/jocn_a_01817

2020). Moreover, Khalighinejad, da Silva, and Mesgarani (2017) and Mesgarani et al. (2014) showed that stops and fricatives are coded by distinct neuronal populations, independent of their voicing status. Collectively, phonetic classes are distributively coded by auditory cortex. It is important to note, however, that all stops share a common phonetic characteristic in both articulation and acoustics: the presence of a burst followed by a release. All fricatives also share a common phonetic characteristic: a narrow constriction in the vocal tract, resulting in turbulent acoustic energy. In other words, different stop consonants are phonetically linked, and different fricative consonants are phonetically linked. These studies demonstrate that auditory cortex encodes phonetic sound classes; however, it is still unknown if auditory cortex encodes features in a manner that goes beyond the phonetic properties of the speech sounds themselves.

Direct evidence of whether auditory cortex functionally links different phonetic sound classes of the same phonological class is key to determining whether auditory cortex can represent abstract phonological features. Here, we take advantage of the functionally defined English feature [voiceless] to determine whether two distinct phonetic classes of speech sounds, that is, stops and fricatives, can be grouped together when they share voicing. The feature [voiceless] in word-initial English stops (i.e., [p^h t^h k^h]) is primarily cued by a temporal dimension: a long duration VOT compared to a short duration VOT for voiced stops (Lisker & Abramson, 1964). Meanwhile, in fricatives (e.g., [f s]), the feature [voiceless] is primarily cued by a spectral dimension: the absence of periodic low-frequency energy that is present in their voiced counterparts because of glottal adduction (Smith, 1997). Overall, English provides a strong test case for our question of interest, as opposed to languages where voicing in fricatives and stops are cued by the same phonetic property. For example, Spanish uses periodic low-frequency energy to cue voicing in both stops and fricatives (Lisker & Abramson, 1964). Specifically, the distinction between voicing in English and Spanish indicates that whether a speech sound is phonologically voiced or voiceless is not tied directly to a single acoustic pattern or phonetic cue.

The question addressed here is whether spectral and temporal phonetic features can be disjunctively coded together into a larger phonological category. To address this question, we performed an auditory MMN experiment. The MMN is a component in the ERP, sensitive to language-specific properties of speech sounds in auditory memory (Winkler, Kujala, Alku, & Näätänen, 2003; Näätänen, 2001; Phillips et al., 2000; Sharma & Dorman, 1999, 2000; Winkler et al., 1999; Näätänen et al., 1997). It is usually evoked in an oddball paradigm, which consists of a series of repeating (standard) stimuli interrupted by an infrequent (deviant) stimulus. The auditory MMN typically peaks between 150 and 350 msec postonset of a deviant stimulus and is maximal over fronto-central electrode sites. The source of the auditory MMN response consistently localizes to the auditory cortex (Alho, 1995; Aulanko, Hari, Lounasmaa, Näätänen, & Sams, 1993; Sams, Kaukoranta, Hämäläinen, & Näätänen, 1991). The MMN is a useful tool to study whether auditory cortex perceptually groups over distinct cues. For example, Gomes, Ritter, and Vaughan (1995) studied the MMN response to pure-tone sinusoids and observed that the duration cue alone can be extracted from a series of standards that also varied in frequency and intensity. This suggests that auditory memory representations can be built based on a single cue, even when other cues are varying.

Here, English-speaking participants took part in a many-to-one oddball experiment. Their EEG was recorded as they listened to a series of frequent, standard consonant–vowel syllables, interrupted by an infrequent, deviant consonant–vowel syllable. The standards and deviants in each block were sampled from five different phones (i.e., voiceless: [p^h t^h k^h f s], voiced: [b d ɡ v z]). In the voiceless standards block, participants heard voiceless consonants as the standards, interrupted by occasional voiced deviants. In the voiced standards block, participants heard voiced consonants as the standards, interrupted by occasional voiceless deviants. As we used multiple phonetic categories with differing manners of articulation in the standards, a many-to-one oddball relationship among the standards only exists if the stops and fricatives that share voicing are grouped together along the basis of their abstract voicing representation. An MMN effect would indicate that distinct phonetic cues group together perceptually based on an abstract phonological feature.

In nearly all MMN studies of speech perception, the standard category in one block serves as the deviant in the other block. If both categories are native to the language of the participants and contrastive in the language, then equal-sized MMNs are predicted across the two blocks. This is not always the case, however. In fact, MMN evidence for features largely arises from findings that demonstrate that some speech sounds elicit larger MMNs than others. These asymmetric results have been observed for vowels (de Rue, Snijders, & Fikkert, 2021; Yu & Shafer, 2021; Scharinger, Monahan, & Idsardi, 2012, 2016; Cornell, Lahiri, & Eulitz, 2011; Eulitz & Lahiri, 2004), consonants (Fu & Monahan, 2021; Hestvik, Shinohara, Durvasula, Verdonschot, & Sakai, 2020; Schluter, Politzer-Ahles, Al Kaabi, & Almeida, 2017; Hestvik & Durvasula, 2016; Schluter, Politzer-Ahles, & Almeida, 2016; Cornell, Lahiri, & Eulitz, 2013; Maiste, Wiens, Hunt, Scherg, & Picton, 1995), and lexical tone (Politzer-Ahles, Schluter, Wu, & Almeida, 2016). These asymmetries are often—although not exclusively (see the work of Maiste et al., 1995)—taken to reflect the underlying featural content of the two categories consistent with underspecified representations (Lahiri & Reetz, 2002, 2010). Traditionally, features were conceptualized as binary. That is, the sound [n] was [+nasal], whereas the sound [d] was [−nasal]. Underspecified features posit that only contrastive aspects of a speech sound are represented.

Therefore, [n] is specified for nasality, possessing the feature [nasal], while [d] lacks any specification for [nasal] in memory and, thus, does not contain the feature [−nasal] (Archangeli, 1988; Steriade, 1995). 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. In studies that have tested voicing in English stops (Hestvik & Durvasula, 2016) and English fricatives (Schluter et al., 2017), larger MMNs are observed when voiceless consonants are the standard relative to when the voiced consonants are the standard, suggesting that voiceless consonants are specified with the feature [voiceless].

These neurophysiological results align with proposals from phonological theory, which posit that English voiceless stops and fricatives are specified for their voicing status and contain the feature [voiceless], whereas voiced consonants are unmarked and underspecified for voicing (Avery & Idsardi, 2001). Linguistic evidence for this claim arises from the findings that (i) voiced English obstruents are less consistent in their phonetic realization relative to voiceless obstruents (Smith, 1997; Docherty, 1992) and (ii) voiced English obstruents do not trigger assimilation, whereas voiceless English obstruents do (Iverson & Salmons, 1995). Both observations are consistent with voiceless English obstruents possessing a marked feature for voicing and voiced obstruents lacking such a marked feature. That voiceless obstruents are marked, whereas voiced obstruents are unmarked, however, is language-specific; languages with different voicing systems (e.g., Spanish, Japanese, Thai, Hindi) could have distinct voicing specifications. The above asymmetry in representation makes predictions for conditions in which we anticipate observing an MMN. We expect an MMN when the specified [voiceless] consonants are the standards but not when the underspecified voiced consonants are the standards.

It is important to note that standard–deviant designs entail a matching procedure of the upcoming physical stimulus with an auditory memory representation (Näätänen & Kreegipuu, 2012; Näätänen, Jacobsen, & Winkler, 2005; Näätänen, 2001). Thus, anticipating what sound comes next is inherent in oddball designs evoking the MMN component (Winkler, 2007). If we observe differences between specified standards and underspecified standards before hearing it, this suggests predictive processing that is at least partially based on that feature. In other words, different neural states preceding the onset of different types of repetitive sounds would indicate that the expectancy between these sounds varies depending on the segregating characteristic, that is, the feature. In fact, differences before the onset of different kinds of standards have been observed (Scharinger et al., 2016). Those differences were evident in the induced neural oscillatory activity, which is neurophysiological activity that—unlike ERPs—is time-locked but not phase-locked to endogenous sensory events. Neuronal oscillations play a key role in auditory perception of verbal input (Morillon & Schroeder, 2015; Doelling, Arnal, Ghitza, & Poeppel, 2014; Arnal & Giraud, 2012; Obleser & Weisz, 2012; Arnal, Wyart, & Giraud, 2011), with various rhythms ascribed to specific functional roles. Brain oscillations refer to rhythmic fluctuations in the excitability of neuronal populations, sometimes in response to sensory stimulation. Beta-frequencies (approximately 15–30 Hz) are dominant in propagating information flow from higher cortical areas and levels of representation (Riddle, Hwang, Cellier, Dhanani, & D'Esposito, 2019; Lewis, Schoffelen, Schriefers, & Bastiaansen, 2016; Fontolan, Morillon, Liegeois-Chauvel, & Giraud, 2014; Arnal et al., 2011; Wang, 2010). For example, descending information flow from higher-order brain areas in an individual listening to sentences is frequency specific (approximately 15–30 Hz), suggesting that beta-frequencies are dominant in top–down propagation (Fontolan et al., 2014). Scharinger et al. (2016) observed an increase in beta-power in a prestimulus time window when the standards were specified (i.e., the high-vowel [ɪ]) relative to when the standards were underspecified (i.e., the mid-vowel [ɛ]). This prestimulus increase in beta-power is taken to reflect predictive mechanisms associated with the MMN, as presumably, only specified features can be used to predict the upcoming stimulus. Here, we also investigated beta-band oscillatory activity before hearing the stimulus, to provide complementary evidence to the neural implementation of features. In the context of predictive processing, if features can be used as the basis for upcoming predictions, we expect beta-band power increases to the specified voiceless standards, consistent with Scharinger et al. (2016).

In short, this study aims to determine if auditory cortex constructs disjunctive groupings for linguistic features in auditory memory and whether features have predictive power in auditory processing. We employ an MMN design that adopted intercategory variation in the standards. In one block, the standards were voiced consonants, and the deviants were voiceless consonants. In the other block, the standards were voiceless consonants, and the deviants were voiced consonants. Importantly, the standards and deviants included both stops and fricatives. Eliciting the MMN suggests that auditory cortex disjunctively codes the temporal and spectral auditory cues for voicing into a coherent auditory memory trace. For the abstract phonological feature [voiceless], we predict an MMN in the voiceless standards block, as [voiceless] appears to be the specified feature, whereas voiced consonants are underspecified for voicing. Finally, given that increased beta-power has been observed in blocks whose standards are putatively specified for a given feature (see the work of Scharinger et al., 2016), we predict an increase in prestimulus beta-power in the voiceless standards block and a reduction in prestimulus beta-power in the voiced standards block.

METHODS

Participants

Thirty native speakers of English participated. All participants were recruited from the University of Toronto Scarborough. Data from three participants were excluded because of technical issues during data acquisition. This left 27 participants (mean age = 20.3 years, *SD* = 2.4 years, 17 women) included in the analysis. All participants were right-handed, as assessed by the Edinburgh Handedness Survey (Oldfield, 1971) and self-reported no known history of auditory, language, or neurological deficits. Finally, all participants provided written informed consent before the experiment and were remunerated for their time. This study was approved by the research ethics board at the University of Toronto.

Materials

The experimental stimuli included 14 consonant–vowel syllables in a [Cɑ] frame. There were five voiced ([b d g v z]) and five voiceless consonants ([p^h t^h k^h f s]), produced 3 times each. Each group of voiced and voiceless consonants included three stop consonants and two fricatives. In addition, the syllables [vi vu p^{h}o fu] were included as fillers. Stimuli were recorded by a female native speaker of North American English in a sound-attenuated cabin. Recordings were made with an Audio-Technica AT3035 microphone at a sampling rate of 44.1 kHz and 16-bit depth. The syllables were produced at a natural speaking rate (mean duration = 540 msec, *SD* = 58 msec). Stimuli intensity was normalized to a mean intensity of 75 dB SPL. Cosine^2 offset ramps were also applied to the offset of the auditory stimuli; this was not applied to the onset to preserve the burst properties of the stop consonants.

Figure 1A presents example acoustic waveforms for 10 representative experimental syllables, one for each category, used in the experiment. The voiced fricatives included periodic acoustic energy during the fricative, as is common in syllable-initial position in North American

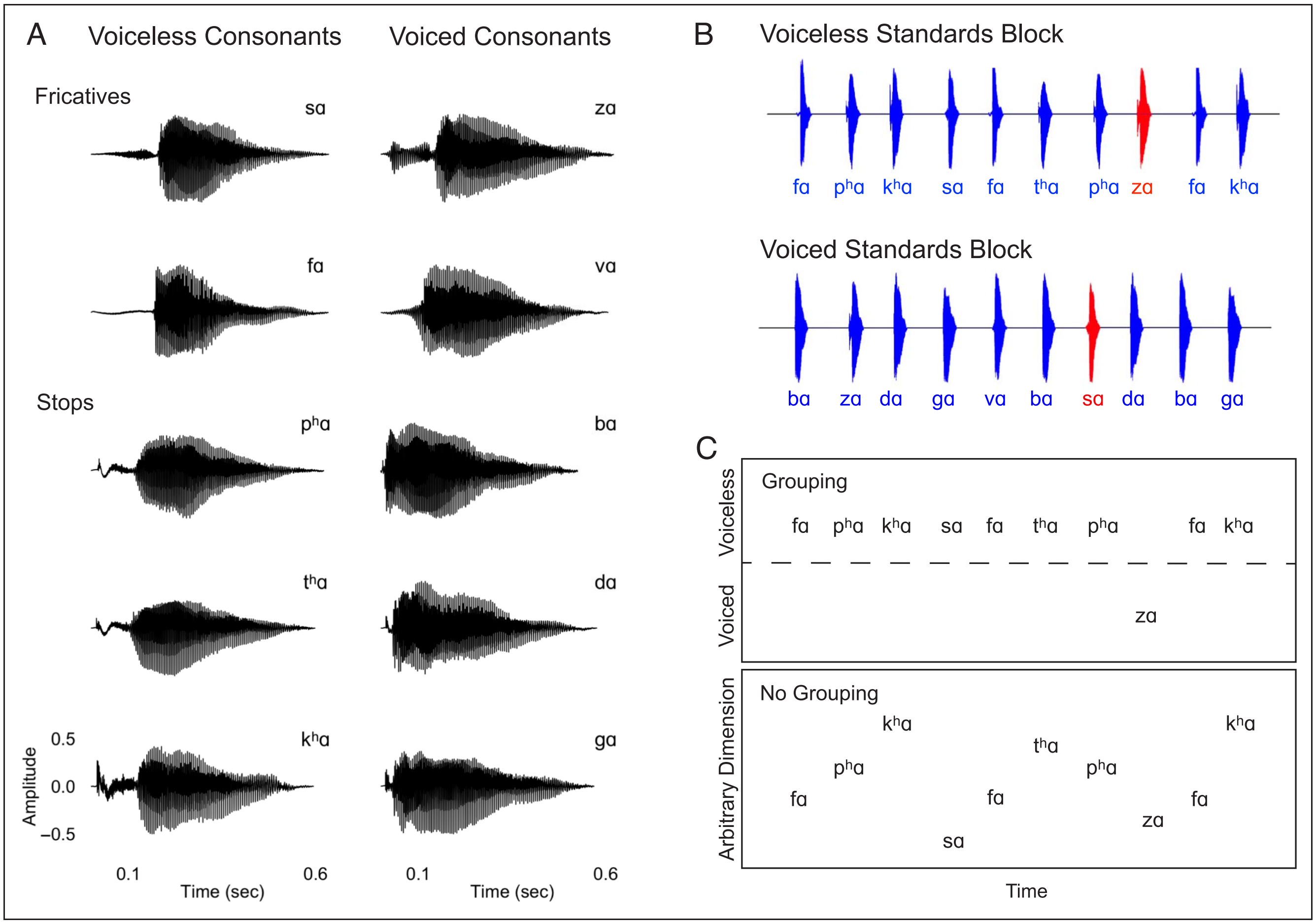

Figure 1. (A) Waveforms of sample stimuli for each phonetic category used in the experiment. Voiceless consonants are on the left, and voiced consonants are on the right. The top four waveforms are of fricatives used in the experiment, whereas the bottom six waveforms are of stops used in the experiment. The voiced fricatives contain periodic energy before the vowel, as is evident in both [zɑ] and [vɑ]. The voiceless fricatives lack this periodic energy. In the stop consonants, the voiceless stops have a longer interval between the release of the burst and the onset of the vowel; that is, they contain a longer VOT. The VOT for voiced stops is considerably shorter. (B) Schematic of the many-to-one oddball paradigm used in the experiment for both the voiceless (top) and voiced (bottom) standard blocks. The standards (blue) in each block included intercategory variation, interrupted by an infrequent deviant (red) sampled from the other voicing category. (C) Predictions for the MMN experiment for the voiceless block only. If the intercategory varying standards are grouped along a shared voicing feature (top), we predict a many-to-one relationship and, consequently, the presence of an MMN. If grouping is not possible (bottom), then no many-to-one relationship exists, and no MMN is predicted.

English (Smith, 1997). This was absent in the voiceless fricatives. The voiceless stops had a longer interval of time, that is, the VOT, between the release of the stop burst and the onset of the vowel as compared with the voiced stops (Lisker & Abramson, 1964). Note that there is no physical property that binds voiceless stops and voiceless fricatives together, only a linguistic property, namely, voicing.

As noted above, the primary phonetic cues to phonological voicing are distinct in stops and fricatives. There are, however, secondary acoustic–phonetic cues on which voiced and voiceless consonants systematically differ, although in a less reliable manner (Lisker, 1986). To determine whether such secondary acoustic–phonetic cues could robustly indicate voicing category membership in our stimuli, we measured several acoustic dimensions known to be secondary cues to the English voicing contrast. Specifically, we measured total syllable duration, consonant duration, vowel duration, fundamental (f0) and first formant (F1) frequency at vowel onset, and intensity of the following vowel. Table 1 provides the means and standard deviations, and Figure 2 provides the distributions for each of these cues in our stimuli. The average values showed expected cross-category patterns, but as is evident by the overlap in the distributions, a value from a single token is not informative about its voicing category membership. Specifically, the distributions for voiced stops and fricatives always overlap with the distributions for voiceless stops and fricatives. In short, these secondary cues do not reliably indicate phonological voicing category membership and, therefore, are unlikely to contribute to effects found in the current work.

Procedure

Participants were seated in a quiet room and passively listened to the stimuli, while watching a silent movie to maintain an awake state and reduce excessive ocular movements (Tervaniemi et al., 1999). Two blocks were presented with a break in between. One block contained voiceless standards and voiced deviants. The other block contained voiced standards and voiceless deviants. The order of blocks was counterbalanced across participants. Figure 1B represents the many-to-one oddball paradigm used in the voiceless (top) and the voiced block (bottom). The standards (blue) in each block included intercategory variation, interrupted by an infrequent deviant (red) sampled from the other voicing category.

In each block, participants heard approximately 785 stimuli, consisting primarily of standards (e.g., voiceless consonants in the voiceless standards block) interspersed with deviants (e.g., voiced consonants in the voiceless standards block) at semiregular intervals. There was, on average, a 7-to-1 standard-to-deviant ratio, with approximately 676 standard and exactly 105 deviant stimuli per block. In each block, three distinct acoustic tokens of the five standard syllables were presented to require participants to abstract over token-specific acoustic properties (Hestvik & Durvasula, 2016; Scharinger et al., 2012, 2016; Kazanina, Phillips, & Idsardi, 2006; Phillips et al., 2000; Aulanko et al., 1993).

For the deviants, only one acoustic token of each of the five deviant syllables was presented, and each deviant category was presented 15 times. This led to 75 deviants per block. Moreover, there were also two filler deviant syllables per block. These filler deviants matched the standards in their voicing but differed in their vowel. They were presented 15 times each, for an additional 30 deviant trials. Filler deviants were included to ensure that there would be detectable differences between tokens as the vowel changes are likely more salient than the consonant changes given the intercategory variation in the standard consonants. Each deviant stimulus was preceded by several standard stimuli (range: 4–10). As few as three to four consecutive standards allow for the elicitation of an MMN response (e.g., Dehaene-Lambertz, 2000; Cowan, Winkler, Teder, & Näätänen, 1993). The number of standards was randomly drawn from a uniform distribution, and the standard stimuli themselves were randomly sampled from the

Table 1. Means for the Measured Secondary Acoustic–Phonetic Cues

Voicing	*Manner*	*Total Duration (msec)*	*Consonant Duration (msec)*	*Vowel Duration (msec)*	*f0 (Hz)*	*F1 (Hz)*	*Intensity (dB SPL)*
Voiced		519 (62.5)	55 (56.5)	464 (34.6)	206 (14.3)	571 (64.0)	75.9 (0.35)
	Fricative	565 (75.1)	116 (38.7)	449 (43.3)	209 (11.2)	574 (81.2)	76.3 (0.26)
	Stop	488 (25.4)	14 (4.3)	474 (25.5)	203 (16.6)	569 (53.7)	75.7 (0.17)
Voiceless		562 (45.3)	121 (36.1)	441 (28.8)	224 (10.2)	743 (86.4)	76.5 (0.27)
	Fricative	590 (60.4)	157 (27.2)	433 (37.0)	220 (9.0)	661 (38.8)	76.7 (0.09)
	Stop	543 (17.4)	98 (15.1)	445 (23.1)	227 (10.5)	805 (50.6)	76.3 (0.21)

One standard deviation of the mean is presented in parentheses. Durations for stop consonants were measured from the offset of the burst to the onset of the vowel. Total duration and vowel duration were measured until the end of periodic voicing in the vowel. f0 and F1 frequency measurements were taken 10 msec after vowel onset. Intensity refers to the intensity of the vowel alone.

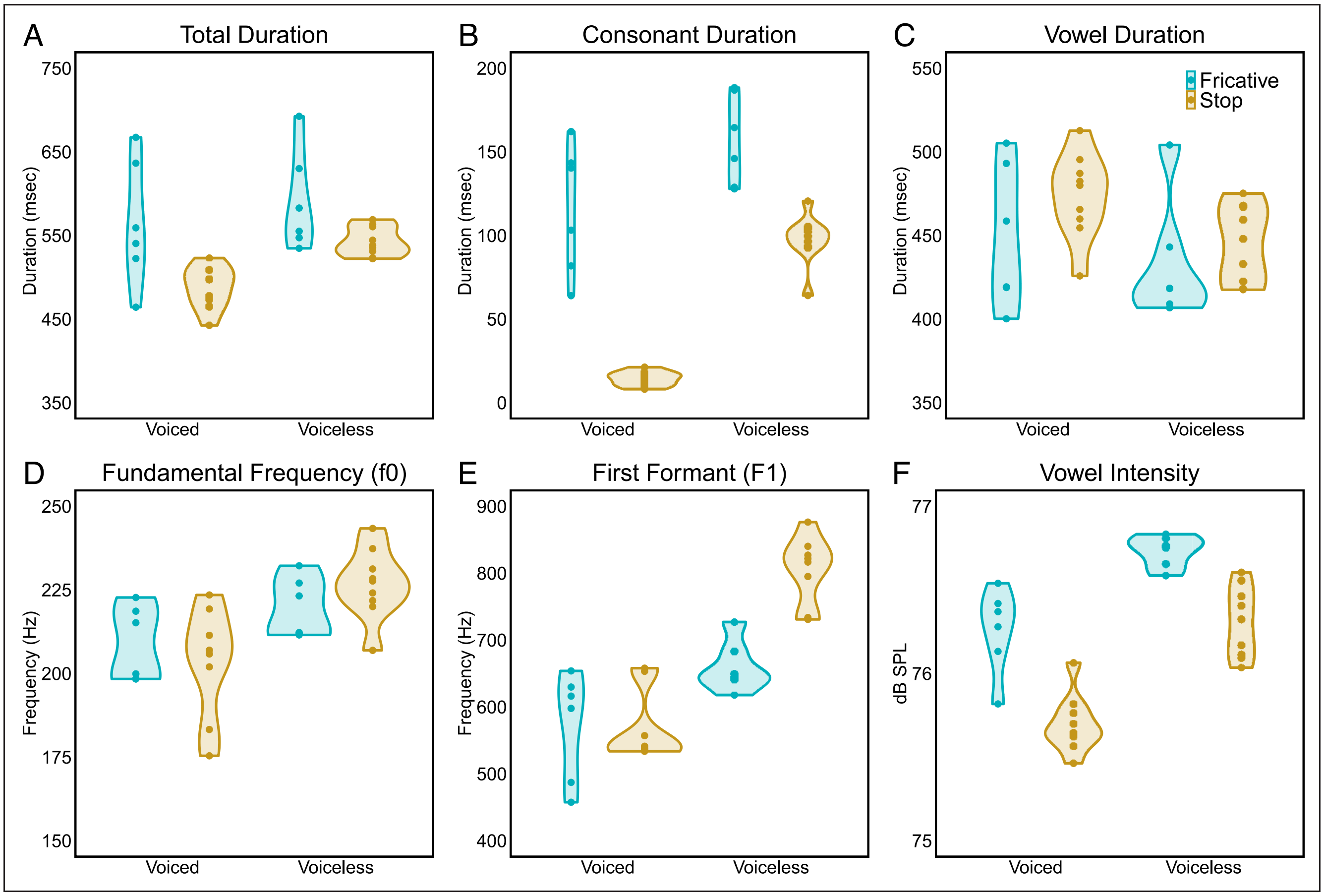

Figure 2. Distributions for each secondary acoustic–phonetic cue measured in the stimuli. Each panel represents one cue: (A) total stimulus duration, (B) consonant duration, (C) vowel duration, (D) f0, (E) F1 frequency, (F) vowel intensity. Points indicate each of the 30 stimuli.

five consonants of its voicing category. Given that the standard token categories were randomly sampled, the number of tokens of a given category in single standard train was not controlled. The duration of the ISI was randomly sampled from a uniform distribution between 1.0 and 1.5 sec. ISI durations in this time range have been shown to reinforce phonological-level processing (Werker & Logan, 1985).

EEG Recording and Analysis

Continuous EEG recordings were acquired from 32-channel actiCAP active electrodes connected to an actiCHamp amplifier (Brain Products GmbH). The EEG signal was digitized at a 500-Hz sampling frequency with a 200-Hz on-line low-pass filter. Electrodes were positioned on the scalp according to the International 10–20 system. Positions included 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 EEG signal was referenced to the right mastoid (TP10) on-line. Impedances were reduced below 20 kΩ at each electrode site before recording. Four additional bipolar electrodes were placed on the temples and above and below the right eye to monitor the EOG. The auditory stimuli were delivered to participants through RHA MA750i noise-isolating in-ear headphones. In addition to the EEG channels, the auditory signal was also sent to the amplifier. This provides the opportunity for off-line correction of any temporal delay between the delivery of the auditory stimulus and sending of the digital trigger marker (Pérez, Monahan, & Lambon Ralph, 2021).

Data analysis was carried out using the EEGLAB toolbox v14.1.1 (Delorme & Makeig, 2004) and custom programs running in MATLAB (Version 2017a, The MathWorks Inc.). First, offset delays between the trigger and the auditory stimulus presentation were corrected to ensure millisecond-precise stimulus-digital trigger synchrony. This was done by cross-correlating the original stimuli waveforms and the auditory track in the EEG recording. Second, only standards that followed at least two other standard trials were retained. This was done to ensure that the averaged standard responses included only trials that were preceded by multiple stimuli from the same voicing category. Next, the EEG signal was rereferenced to the linked mastoids, which provide the most robust MMN responses (Mahajan, Peter, & Sharma, 2017). Subsequently, a high-pass filter (finite impulse response; FIR) at 1 Hz was applied. The PREP toolbox (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015) was used to identify and remove bad channels (mean = 1.5, *SD* = 1.4). These bad channels were then interpolated. Because of the relatively small channel montage (i.e., 32 channels), the spherical spline interpolation method was utilized for bad channel interpolation, which is more accurate for low-density electrode montages (Perrin, Pernier, Bertrand, & Echallier, 1989). Next, an artifact subspace reconstruction (ASR) algorithm adapted from the EEGLAB software (Mullen et al., 2013) was implemented to remove high-amplitude artifacts. ASR transforms a sliding window of the EEG data with a principal component analysis to identify channels of high variance. This is accomplished via a statistical comparison of clean EEG data containing minimal artifacts. This clean data were automatically identified inside each individual EEG recording and subsequently used as the calibration data for the ASR. Corrupted channels (or subspaces of multiple channels) were reconstructed from neighboring channels using a mixing matrix, which is computed from the covariance matrix of the calibration data. In this study, a sliding window of 500 msec and a variance threshold of 3 *SD*s were used to identify corrupted subspaces. Portions of the data that were not possible to reconstruct because of the presence of multiple artifacts were removed.

Subsequently, an adaptive mixture-independent component analysis (AMICA) technique (Palmer, Kreutz-Delgado, & Makeig, 2012) was applied to the cleaned EEG data from each participant to separate the EEG recordings, a combination of individual brain and nonbrain sources mixed by volume conduction, into spatially static independent components of maximal temporal independence. AMICA maximizes mutual information reduction and the dipolarity of scalp projections following decomposition (Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012). AMICA was performed on 45 min of EEG data for each participant. Thus, the number of time points used to estimate the weighting matrix ranged from 161 to 177 times the number of channels squared, exceeding recommendations for satisfactory decomposition (Delorme et al., 2012). Next, an equivalent dipole current source was fit to each independent component from a three-shell boundary element model using the DIPFIT toolbox in EEGLAB (Oostenveld & Oostendorp, 2002). The EEG electrode positions were aligned to fit the standard Montreal Neurological Institute brain. Then, individual components accounting mainly for electrooculographic, electrocardiographic, electromyographic, or line noise were removed from the data. Component rejection was performed manually guided by the following criteria: (i) component's topography, (ii) component's time series, (iii) component's power spectrum properties, and (iv) properties of the dipole associated to each component: localization outside of the head or close to the eyes jointly with low variance (up to 5%). On average, we removed two components per participant (range: 1–5). From this step, the ERPs and event-related spectral perturbations (ERSPs) were computed.

After data preprocessing, in the voiceless standards block, there were, on average, 69 voiced deviants (*SD* = 9.2) and 440 voiceless standards (*SD* = 61.3). In the voiced standards block, there were, on average, 71 voiceless deviants (*SD* = 7.1) and 435 voiced standards (*SD* = 51.1). Overall, there were, on average, 438 standards (*SD* = 55.9) and 70 real deviants (*SD* = 8.2) per block, and the number of trials were approximately equivalent across blocks. Moreover, data preprocessing did not disproportionately affect the different real deviant categories (percentage of trials retained after data preprocessing; voiced deviant stop: 92.9%; voiced deviant fricative: 91.6%; voiceless deviant stop: 93.7%; voiceless deviant fricative: 95.4%).

ERP Analysis

First, the continuous EEG signal was down-sampled to 250 Hz. Next, epochs were extracted with a 100-msec prestimulus baseline and a 700-msec poststimulus onset time window. Epochs were baseline corrected using the 100-msec prestimulus baseline. Average ERPs were calculated for each condition. The difference ERP waveform was obtained for each block by subtracting the ERPs of the standard from the deviant. Because MMNs are largest over fronto-central scalp areas when referenced to linked mastoids (Näätänen, Paavilainen, Rinne, & Alho, 2007), the eight fronto-central electrode sites (i.e., Fz, FC1/2, Cz, C3/4, CP1/2) were collapsed. The resulting channel-averaged ERPs for each condition were statistically compared at each time point. Statistical analyses were conducted in the −100- to 700-msec time window with permutation tests on the *t* statistic and a false discovery rate (FDR) correction for multiple comparisons. The number of permutations was set to 2000. Differences with $p_{\mathrm{FDR}} < .05$ are reported as statistically significant.

To ensure that the observed ERP responses are driven by the standard–deviant relationship, we also included a control analysis. In this analysis, the ERP to the deviant is compared with the ERP to the standard version of the same stimulus (Hestvik & Durvasula, 2016; Peter, McArthur, & Thompson, 2010; Pulvermüller & Shtyrov, 2006; McGee et al., 2001; Deacon, Gomes, Nousak, Ritter, & Javitt, 2000). This type of analysis is referred as the identity MMN (iMMN). The iMMN for the voiced consonants was calculated by comparing the ERP responses when they were the standard (voiced standard block) compared to when they were the deviant (voiceless standard block). Similarly, the voiceless iMMN was calculated by comparing the ERP of the voiceless standards in the voiceless standard block versus the ERPs of the voiceless deviants in the voiced standard block. Despite substantial variation in the standards and deviants in the current design, it is possible that any differences in the within-block comparison may arise because of the ERPs being different to voiceless and voiced consonants and not because of the standard–deviant relationship. The iMMN comparison potentially eliminates this confound. The presence of an iMMN suggests that the differences observed in the within-block MMN cannot be solely attributed to intrinsic ERP response differences to voiceless versus voiced consonants, but that the MMN is also driven by the standard–deviant relationship.

ERSP Analysis

For the ERSP analysis, epochs were extracted from the continuous EEG signal with a 1-sec baseline and 2-sec poststimulus onset time window. The time–frequency decomposition of each epoch was computed using a wavelet window Morlet taper. The number of cycles in each Morlet increased linearly in 0.8 cycles beginning at 3 cycles at 3 Hz. Single-trial normalization at each time–frequency bin was performed by dividing frequency specific power averaged from −444 to −300 msec. Finally, ERSPs containing estimations from 3 to 30 Hz and −300 to 700 msec were averaged across all channels for each condition. Statistical analyses comparing the voiced and voiceless standards were conducted with bootstrap tests. The number of random samples used in the bootstrap was 2000. Differences with $p_{\mathrm{FDR}} < .05$ are reported as statistically significant.

RESULTS

ERPs

Average waveforms elicited over fronto-central electrode sites in each block are presented in Figure 3. Figure 3A shows the grand average for the voiceless standard block. The permutation test revealed significant differences in the time window from 116 to 196 msec and a later time window, from 316 to 440 msec. Figure 3B shows the grand average in the voiced standard block. The response to the voiceless deviant was relatively more negative in the 216- to 240-msec time window and at two very brief later time windows (i.e., 532–556 msec, 636–648 msec); however, the voiceless deviant elicited a larger positivity in the 292- to 364-msec time window relative to the voiced standard. Figure 3C provides a comparison of the two standards. The voiced standard stimuli elicited a larger negativity in the 68- to 138-msec time window and a larger positivity in the 208- to 248-msec time window. In turn, the voiceless standard stimuli elicited a larger positivity in the 292- to 436-msec time window. A comparison of ERP responses to the standards based on manner (i.e., stop, fricative) and voicing (i.e., voiced, voiceless) is provided in Appendix A. Finally, Figure 3D compares the difference waves between the two blocks (deviant minus standard). The difference between the deviants and standards across the two blocks is evident between 96 and 188 msec and between 288 and 412 msec.

Overall, in the voiceless standards block, the voiced deviant eliciting a larger negative deflection in the ERP relative to the voiceless standard is consistent with typical observations in an MMN paradigm. In the voiced standards block, the most robust difference is during the 292- to 364-msec time window. There, the deviant elicited a larger positive response relative to the standard, which is opposite of the characteristic MMN pattern. The presence of the early and late negative ERP deflections in the voiceless standard block suggests that auditory cortex was able to perceptually group the voiceless consonants together.

Next, we conducted an iMMN analysis to establish whether the MMN effect is driven by the standard–deviant relationship. Figure 4 shows that there was no difference when the voiceless consonants were the standards relative to when they were the deviant. The panels beneath the ERP waveforms show the difference waves (deviant minus standard) for each block. In the voiced consonants iMMN comparison, voiced consonants as the deviant elicited a larger negativity in the 196- to 212-msec time window compared to when they served as the standards. Results from iMMN analysis suggests that the differences in the within-block MMN comparison above are not solely because of intrinsic ERP responses to voiced and voiceless consonants alone but also reflect the standard–deviant relationship.

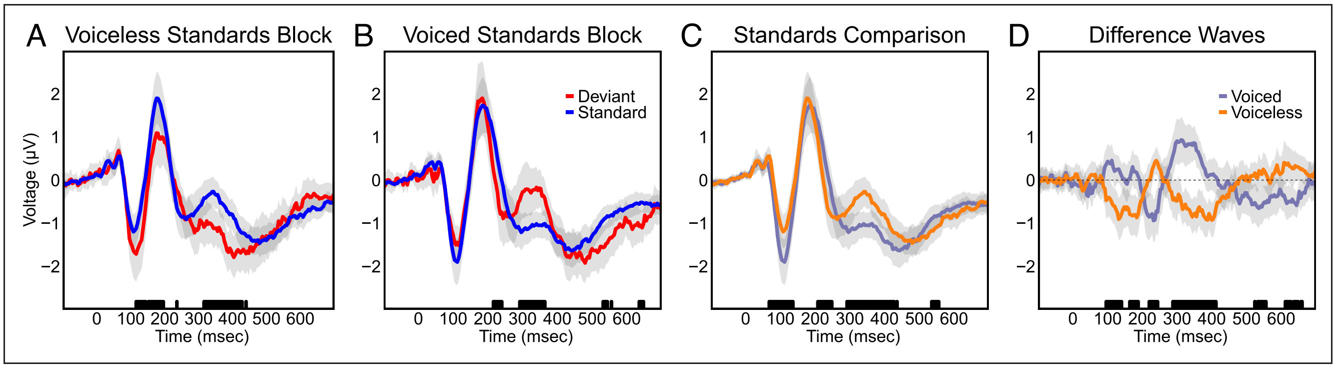

Figure 3. Average ERP responses for the within-block comparisons. ERPs are the average of eight fronto-central scalp electrodes (Fz, FC1/2, Cz, C3/4, CP1/2). (A) ERP responses in the voiceless standards block: voiceless standard and voiced deviant; (B) ERP responses in the voiced standards block: voiced standard and voiceless deviant; (C) ERP responses to the standards: voiceless standard and voiced standard; (D) comparison of ERP difference waves (deviant minus standard) in each block (voiced refers to the voiced standards block, voiceless refers to the voiceless standards block). Rug plots along the *x* axes indicate time points when the comparison was statistically significant (see text for analysis procedures). Shaded regions represent the 95% confidence interval of the mean.

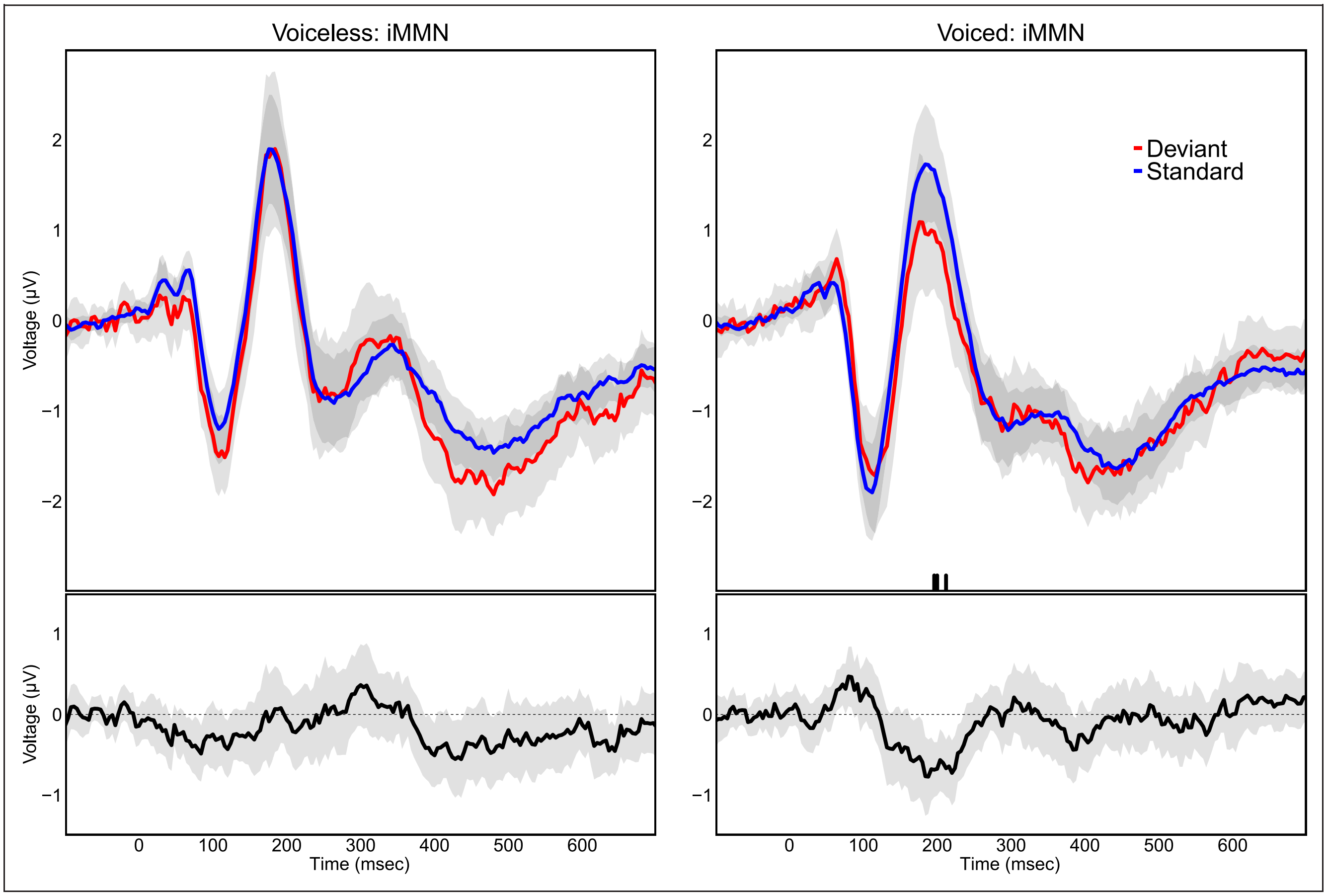

Figure 4. iMMN ERP responses averaged over eight fronto-central scalp electrodes (Fz, FC1/2, Cz, C3/4, CP1/2) for (left) voiceless standards and deviants and (right) voiced standards and deviants. The bottom panels are the difference ERP waves, computed as the response to the standard subtracted from the response to the deviant, for the (left) voiceless consonants and (right) voiced consonants. Rug plots along the *x* axes indicate time points when the comparison was statistically significant (see text for analysis procedures). Shaded regions represent the 95% confidence interval of the mean.

Thus far, the MMN comparisons have focused on fronto-central electrode sites. Analyzing the topographic distribution allows us to (i) identify the extent over the scalp in which these differences are observed and (ii) confirm that the overall distribution of our responses is consistent with the typical MMN topographical patterns. Given that such extensive intercategory variation in the standards is rarely tested, it is also important to provide a more complete picture of the nature of these responses. The two distinct time windows selected are within the typical MMN range: 125–225 msec and 300–400 msec. These time windows also align with the within-block MMN and iMMN results. Figure 5A presents each condition's topographical ERP distribution in the earlier 125- to 225-msec time window. In the within-block comparison (Figure 5C), the voiced deviant elicited a relatively larger negativity compared to the voiceless standard in the voiceless standard block. This difference was distributed over frontal, central, parietal, and occipital electrode sites. There were no differences between the voiceless deviant and the voiced standard in the voiced standards block and no difference between the iMMN difference waves. In the iMMN comparison (Figure 5B), the voiced deviant elicited a relatively larger negativity compared to the voiced standard over central and parietal electrode sites. No difference was observed in the voiceless comparison. The iMMN differences are distributed more squarely over central electrode sites in the voiced consonants comparison. In the within-block MMN comparison, the differences observed in the voiceless standards block extends beyond the fronto-central electrode sites in the ERP waveform analysis to include posterior and occipital sites.

To assess the nature of the ERP differences in the later time window, we also conducted an analysis in the 300- to 400-msec time window (Figure 6). In the within-block comparison (Figure 6C), we observed differences in both blocks. In the voiceless standards block, the voiced deviant elicited a larger negativity relative to the voiceless standard. In the voiced standards block, the voiceless deviant elicited a relatively larger positivity compared to the voiced standard. This larger positivity is inconsistent with typical MMN reports. Again, this effect was widely distributed over nearly all electrode sites. No iMMN (Figure 6B) was observed in the later time window. In summary, the extent of the topographic differences over the scalp in the later time window are even greater than those in the earlier time window; the voiceless standards block shows the relatively larger negativity consistent with the typical MMN distribution, whereas the voiced standards block shows the opposite pattern, inconsistent with the typical MMN distribution.

Summarizing, for the within-block comparisons, the voiceless standard block showed the canonical MMN pattern in two time windows, an earlier (125–225 msec) and later time window (300–400 msec). That is, voiced deviants elicited a larger negativity when preceded by a series of voiceless standards. Moreover, the iMMN response to voiced stimuli when they were a deviant resulted in a larger negativity compared to when they were the standard. On the other hand, voiceless deviants showed a larger positivity when preceded by a series of voiced standards in the later time window but not in the earlier time window. The timing and distribution of the larger negativity to the voiced deviants are consistent with previous MMN reports (Näätänen et al., 2007). The larger positivity to the voiceless deviants in the voiced standard block is not likely an MMN, but is also consistent with previous reports that have used an intercategory variation many-to-one oddball paradigm for the unmarked standards (Fu & Monahan, 2021).

ERSP

To assess whether neural activity preceding the voiced and voiceless standards is different, we analyzed the oscillatory power of the responses to the standards. Although we were particularly interested in the prestimulus period and beta-band frequencies, the analysis included a more complete exploration of the oscillatory activity, until 700-msec poststimulus onset and frequencies between 3 and 30 Hz. Figure 7A presents these results. First, consistent with our predictions, we observed decreased beta-band oscillatory power in the prestimulus baseline (−140 to −40 msec) for the putatively underspecified voiced standards relative to the specified voiceless standards (see the work of Scharinger et al., 2016). Next,

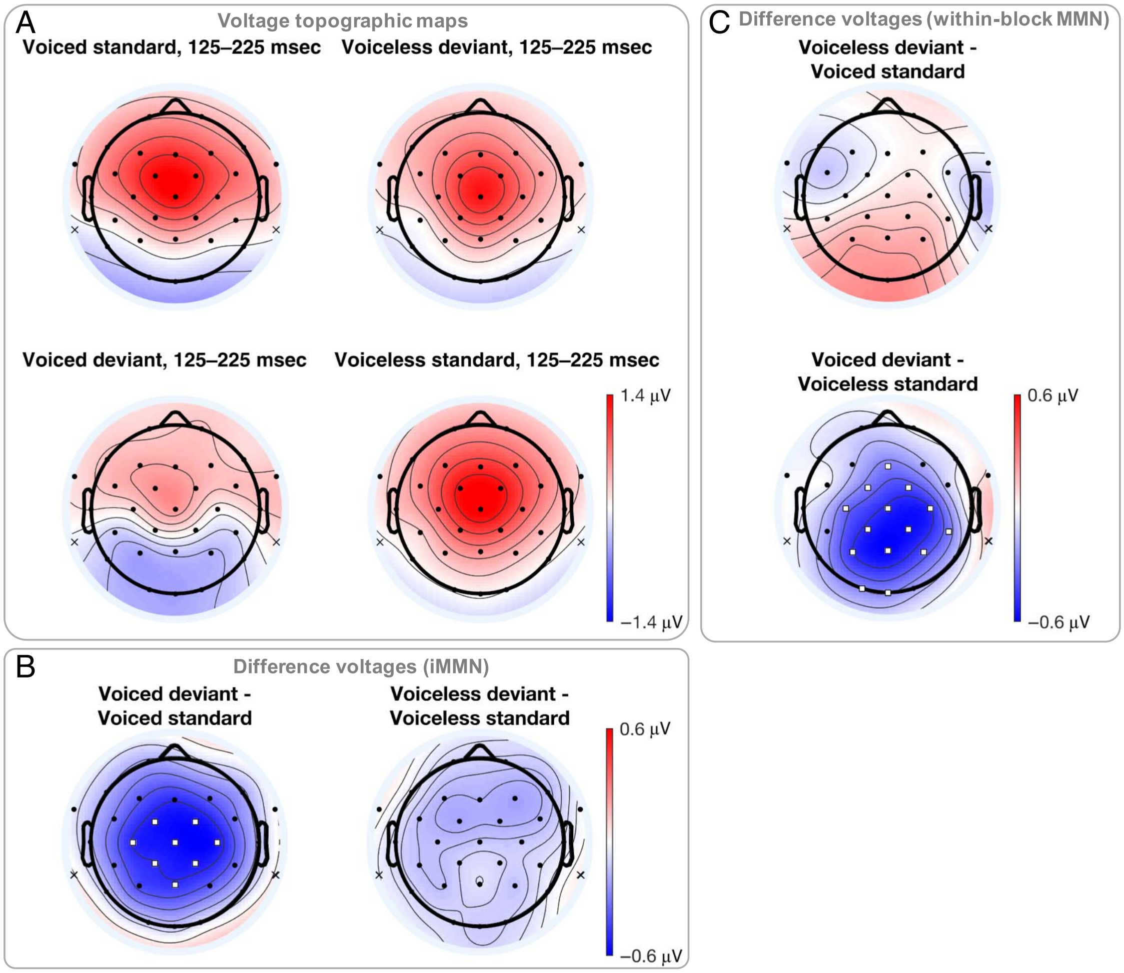
A
Voltage topographic maps
Voiced standard, 125–225 msec
Voiceless deviant, 125–225 msec
Voiced deviant, 125–225 msec
Voiceless standard, 125–225 msec
1.4 μV
−1.4 μV
B
Difference voltages (iMMN)
Voiced deviant - Voiced standard
Voiceless deviant - Voiceless standard
0.6 μV
−0.6 μV
C
Difference voltages (within-block MMN)
Voiceless deviant - Voiced standard
Voiced deviant - Voiceless standard
0.6 μV
−0.6 μV

Figure 5. (A) Voltage topographic maps for the 125- to 225-msec time window. (B) iMMN comparisons between standards and deviants across experimental blocks. (C) Within-block MMN comparisons between standards and deviants within experimental blocks. Electrodes highlighted in white squares denote electrode sites with significant differences using a permutation test ($p_{\mathrm{FDR}} < .05$). Topographic plots use the linked mastoid reference, whose locations are marked with an '*x*'. See the Methods section for analysis procedures.

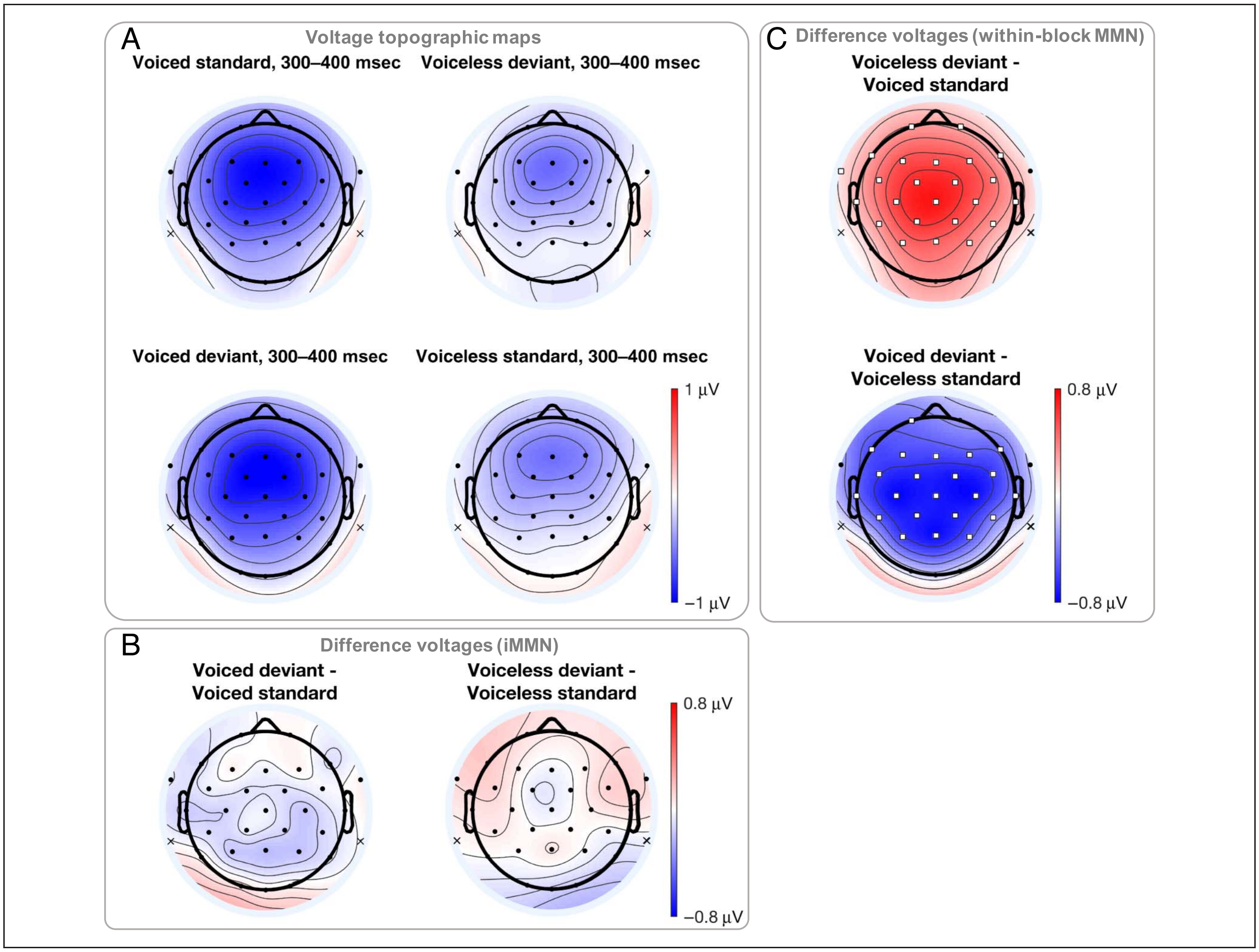
A
Voltage topographic maps
Voiced standard, 300–400 msec
Voiceless deviant, 300–400 msec
Voiced deviant, 300–400 msec
Voiceless standard, 300–400 msec
1 μV
−1 μV
C
Difference voltages (within-block MMN)
Voiceless deviant - Voiced standard
Voiced deviant - Voiceless standard
0.8 μV
−0.8 μV
B
Difference voltages (iMMN)
Voiced deviant - Voiced standard
Voiceless deviant - Voiceless standard
0.8 μV
−0.8 μV

Figure 6. (A) Voltage topographic maps for the 300- to 400-msec time window. (B) iMMN comparisons between standards and deviants across experimental blocks. (C) Within-block MMN comparisons between standards and deviants within experimental blocks. Electrodes highlighted in white squares denote electrode sites with significant differences using a permutation test ($p_{FDR} < 0.05$). Topographic plots use the linked mastoid reference, whose locations are marked with an "*x*." See the Methods section for analysis procedures.

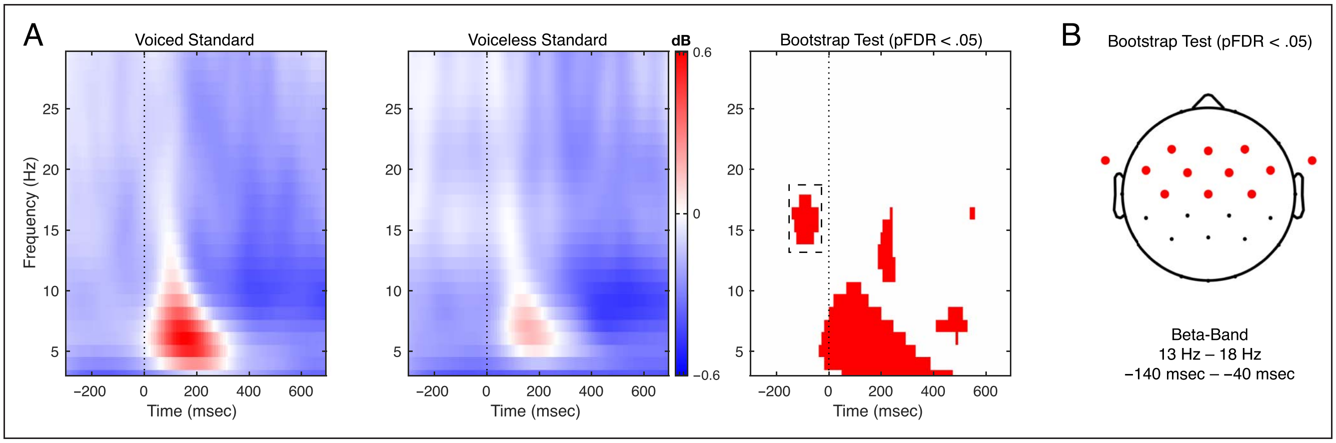

Figure 7. (A) ERSP for both the voiced standards (left) and voiceless standards (center). Third panel illustrates the results of a statistical comparison between the two ERSPs. The ERSP response to the voiced standards has higher oscillatory power compared to the ERSP response to the voiceless standards in theta-band (3–10 Hz) between 150 and 200 msec. In the prestimulus period, beta-band oscillatory power differences (13–18 Hz, marked with a dashed rectangle) are observed between the two standard conditions. The voiced standards show lower oscillatory power relative to voiceless standards. (B) Distribution of significant channel responses (marked in red) to beta-band (13–18 Hz) oscillatory power in the prestimulus period (−140 to −40 msec).

voiced standards showed increased oscillatory power on frequencies in the theta-band range (3–7 Hz) approximately from stimulus onset until 400-msec poststimulus onset. The opposite pattern, larger oscillatory power in the voiceless standards relative to the voiced standards, was observed between 11 and 16 Hz in the 200- to 250-msec time window. Figure 7B provides the topographical distribution of channels that show a difference between the two standard conditions in the beta-band between −140- and −40-msec prestimulus onset. This time window was chosen based on the bootstrap statistical test. The beta-power differences localized over fronto-central electrode sites. These differences in prestimulus beta-power are detected using a bootstrap test and $p_{\mathrm{FDR}} < .05$ but disappear with the more conservative permutation test. In general, these results suggest the voiceless segments provide predictive information.

DISCUSSION

The goal of the current experiment was to determine whether auditory cortex represents abstract phonological features. Specifically, we tested English voicing, which codes stops and fricatives as voiced or voiceless. We used a many-to-one oddball MMN design to test if listeners construct a unified auditory memory representation of distinct phonetic cues that are functionally linked to one phonological feature. The key feature of the design is that intercategory variation is introduced into the standards. A many-to-one relationship only exists if spectral and temporal phonetic cues to voicing are grouped together. This many-to-one relationship would lead to an MMN effect. The absence of an MMN would suggest no perceptual grouping. Below, we discuss the three primary findings from the current study.

First, in the context of voiceless standards, the neurophysiological response reflects a disjunctive encoding of a spectral and temporal phonetic cue that is functionally linked: periodic, low-frequency spectral energy in fricatives, and VOT in stops. A larger negative deflection over fronto-central electrode sites was observed in two distinct time windows to the voiced deviants in the context of a series of voiceless standards: 116–196 msec and 316–440 msec. The timing and topographic distribution of these responses indicate that these are the MMN (Näätänen & Kreegipuu, 2012; Näätänen et al., 2007; Näätänen, 2001). Observing the MMN suggests that the auditory memory trace for the standards was constructed based on the voicing feature, despite its distinct phonetic implementation in stops versus fricatives. That is, spectral and temporal phonetic cues can be disjunctively coded when they are functionally linked in the phonology.

Second, an MMN was observed only in the voiceless standards block. Asymmetric MMNs of this type are not uncommon. As noted above, previous MMN results have found that in English, voiced stop (Hestvik & Durvasula, 2016) and fricative (Schluter et al., 2017) deviants elicit a larger MMN in the context of a series of voiceless stops or fricatives, respectively. The conclusion drawn is that in English, voiceless consonants are specified for their voicing with the feature [voiceless], whereas voiced consonants are underspecified for voicing (Avery & Idsardi, 2001; Iverson & Salmons, 1995). Considering these previous findings, the observation in the current study that the voiced deviant elicited a negative deflection in the voiceless standards block was predicted. Differences were largest over fronto-central electrode sites, where the MMN is typically observed. The presence of an MMN in the 100- to 200-msec time window is consistent with the time course of a number of studies using the MMN to investigate speech sound categories (Cornell et al., 2011; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006; Phillips et al., 2000; Winkler et al., 1999; Näätänen et al., 1997). The later negative deflection has also been observed in oddball experiments that expose listeners to intercategory variation in the standards (Fu & Monahan, 2021), which might require listeners to tap into putatively phonemic levels of representation (Kazanina et al., 2006) or perceptually group experimental stimuli based on dialect (Lanwermeyer et al., 2016; Scharinger, Monahan, & Idsardi, 2011).

In a comparison of the standards, the voiced standards elicited a larger negativity in an early time window (~100-msec poststimulus onset) and a later time window (~300- to 430-msec poststimulus onset). It is possible that the differences in the standards comparison are because of differences in the obligatory auditory evoked potential responses to the voiced consonants as a class compared to the voiceless consonants as a class. First, however, MMN and repetition suppression (Larsson & Smith, 2012; Gruber, Malinowski, & Müller, 2004) experiments suggest that repetition fundamentally alters responses to exogenous stimuli. Moreover, it is possible, for example, that repetition intrinsically affects one category differently from the other. This is especially relevant in designs where asymmetric responses are predicted, as in the current experiment. Second, the responses in the late time window here are consistent with previous findings (Fu & Monahan, 2021). There, the deviant in the underspecified standard block elicited a positivity in the ~300- to 400-msec time window, and the deviant in the specified standard block elicited a negativity in a similar time window. That experiment and the current experiment used two distinct features (i.e., [retroflex] vs. [voiceless]) from two distinct languages (i.e., Mandarin Chinese vs. English). That said, we cannot entirely rule out an auditory evoked potential account of the standards in the within-block analysis.

To address this possibility, an iMMN analysis was conducted. While the within-block MMN contrast compares standards and deviants from within the same block (Scharinger et al., 2012, 2016; Kazanina et al., 2006; Näätänen et al., 1997; Kraus, McGee, Sharma, Carrell, & Nicol, 1992), it is possible that the MMN could be partially driven by intrinsic ERP responses to different acoustic–

phonetic (or even phonological) properties of the stimuli. An iMMN analysis, however, putatively eliminates the influence of stimulus intrinsic properties by comparing the responses to the same category (or class of speech categories) to itself when in the standard and deviation position. Similar experiments testing phonetic and phonological representations have also reported the iMMN (Fu & Monahan, 2021; Hestvik & Durvasula, 2016; Cornell et al., 2011, 2013). In the current iMMN analysis, the ERP to the voiced deviants was more negative relative to the ERP to the voiced standards around 200-msec poststimulus onset. No such differences were observed in the comparison of the voiceless consonants. The difference in the iMMN suggests that the results of the within-block MMN comparison cannot be solely because of intrinsic physical differences between voiced stops and fricatives as a class, relative to voiceless stops and fricatives.

As noted above, Fu and Monahan (2021) also report an asymmetric MMN with intercategory variation in the standards but tested retroflex consonants in Mandarin Chinese. There, Mandarin-speaking participants were exposed to intercategory variation in the standards. An MMN was observed only in the retroflex standard block, where retroflex consonants were the standards and nonretroflex consonants were the deviants. No MMN was observed when nonretroflex consonants were the standards and retroflex consonants were the deviants. It was concluded that the feature [retroflex] was extracted from the sequence of standards in the retroflex standards block and served as the basis for the auditory memory trace. Whereas intercategory variation was used in the standards, stops, affricates and approximants all acoustically code retroflex similarly, that is, lower spectral energy ranges. Like the current experiment, an MMN in the retroflex standard block was observed in a later time window (256–380 msec). There, it was argued that the later negativity might reflect the added task difficulty involved with having to group various phonetic categories with one another based on a single phonetic feature. Negative deflections in this study also occurred around 316- to 440-msec poststimulus onset, which is later than the normal MMN window (Strotseva-Feinschmidt, Cunitz, Friederici, & Gunter, 2015; Martynova, Kirjavainen, & Cheour, 2003; Čeponienė et al., 2002; Cheour, Korpilahti, Martynova, & Lang, 2001). Late negativities have been observed in both speech (Fu & Monahan, 2021; Datta et al., 2020; Hestvik & Durvasula, 2016) and nonspeech paradigms (Peter, McArthur, & Thompson, 2012; Zachau et al., 2005). These late negativities typically appear with the within-block MMN but can also appear independently (Strotseva-Feinschmidt et al., 2015). Bishop, Hardiman, and Barry (2011) suggest that late negativities might appear as a result of additional processing required by certain features of stimuli that are difficult to detect.

Gomes et al. (1995), who included variation across multiple cues in the standards, observed an MMN at approximately 250-msec poststimulus onset. This is later than other studies that used sinusoids as their stimuli (e.g., Sams, Paavilainen, Alho, & Näätänen, 1985). In another varying standards design, Hu, Gu, Wong, Tong, and Zhang (2020) observed a visual MMN in the 230- to 290-msec time window when participants could group distinct, visually presented lexical items belonging to the same semantic class (e.g., colors vs. tastes). The later time course of differences in the current experiment might reflect the more complex integration of multiple categories into a single memory trace. Alternatively, these late negativities might also reflect the increased time taken to accumulate sufficient information to identify the stimulus as deviating from the standard. It should also be noted that in the current experiment, we observed a positivity in the voiced standards block, which is putatively the block in which the standards could not be grouped together if voiced consonants lack a specification for voicing. In another study, we also observed a larger positivity to the deviant in the underspecified standard block in a similar time window, that is, 320–364 msec (Fu & Monahan, 2021). The topography and time course of the positivity response are similar across the two experiments. Previous experiments on speech sound categories have reported differences in the P3 time window (Friedman, Cycowicz, & Gaeta, 2001; Escera, Alho, Schröger, & Winkler, 2000), but these often include overt tasks that require attention even if they are irrelevant to the standard or deviant stimuli (Winkler et al., 2003). Given the passive nature of the current design, the P3a, which is often elicited as an orienting response in oddball tasks with an overt task (Polich, 1993, 2007), was not predicted and is also not present in the current iMMN comparisons. Moreover, the substantive phonetic variation in our standards and deviants likely reduced the saliency of a change from standard to deviant and, as such, was potentially too subtle to draw participants' attention. To summarize our second finding, we observed an asymmetric MMN in both the within-block MMN and iMMN analyses, consistent with previous findings on English voicing in stops and fricatives, independently, as well as previous findings that have used intercategory variation in the standards. Moreover, the effects observed in the later time windows may reflect the additional processing demands and complexity that result from the current design.

Third, linguistic features appear to be used to generate predictions about the incoming auditory stimulus. This predictiveness occurs despite physical variation in the repeating standard auditory stimuli (Winkler et al., 1990). Predictions are created even when the standards remain constant along one physical dimension while varying in others, for example, when remaining constant in duration but varying in intensity and frequency (Nousak, Deacon, Ritter, & Vaughan, 1996; Gomes et al., 1995). This result is consistent with previous findings of information propagation from higher-level regions of the functional hierarchy observable in prestimulus intervals (Bastos et al., 2015) and in the activation of higher-level linguistic representations during language comprehension (Molinaro, Monsalve, & Lizarazu, 2016). This prestimulus differential neural activity in beta-band could be considered a hallmark of predictive coding mechanisms. That is, abstract characteristics of the speech sound representation are being predicted, perhaps demonstrating an instance of "what" cortical hypothesis generation (Doelling et al., 2014). The MMN is argued to be based on predictive neurophysiological models, where sensory cues and abstract information combine to produce predictions about upcoming sounds (Winkler, 2007). If the MMN fundamentally reflects predictive mechanisms (Wacongne et al., 2011; Garrido, Kilner, Kiebel, & Friston, 2009; Winkler, 2007; Friston, 2005), the response to the deviant then indicates a predictive coding error with the constructed neural representation of the standard. As such, it is predicted that differences in beta-band oscillations should be observed within an MMN paradigm.

Beta-power in electrophysiological responses has been taken to indicate information propagation from higher-order brain areas and levels of representation (Riddle et al., 2019; Bidelman, 2015; Lewis, Wang, & Bastiaansen, 2015; Fontolan et al., 2014; Arnal & Giraud, 2012; Arnal et al., 2011; Buschman & Miller, 2007; Engel & Fries, 2010; Wang, 2010). In addition, beta-band oscillatory power increases if the system maintains the actual language stimuli in memory, and it will decrease if the current cognitive state is interrupted by novel and/or unexpected stimuli. Consequently, a beta-band oscillatory power decrease predicts the probability of new processing demands (Engel & Fries, 2010). Scharinger et al. (2016) observed increased prestimulus beta-power when the standard was specified (i.e., [ɪ]) and decreased prestimulus beta-power when the standard was underspecified (i.e., [ɛ]). Here, we also observed a similar increase in prestimulus beta-power in the block where the standards were the specified, that is, voiceless consonants, and a decrease in beta-power prestimulus onset when the standards were the underspecified, that is, voiced consonants. The specification for voiceless consonants likely forms the basis of the auditory memory trace and is used to predict the upcoming speech sound. Moreover, it also potentially reflects the added maintenance of linguistic stimuli in memory, as they are marked with the feature [voiceless]. This is compared to the voiced consonants, which are underspecified for voicing. Furthermore, we observed reduced theta-power in the voiceless standards relative to the voiced standards, which is consistent with the construction of a short-term auditory memory trace in the voiceless standards block. Although we did not have predictions regarding differences in theta-band power, this is in a similar time window and frequency band as the MMN.

Positing that voiceless consonants are marked is consistent with proposals from linguistic theory (Avery & Idsardi, 2001; Iverson & Salmons, 1995). Given the ERP findings, the voiced deviants elicited a larger MMN in the context of the series of voiceless standards. If the voiceless standards are marked with the feature [voiceless] and this feature results in greater habituation because of the many-to-one organization of the standards, we expected increased beta-band oscillatory power in the voiceless standards block relative to the voiced standards block. In the ERSP results, the voiceless standards did elicit an increase in prestimulus beta-power compared to the voiced standards. Consistent with the ERP results, where an MMN was elicited only in the voiceless standards block, we attribute this increase in prestimulus beta-power to the voiceless standards in the oscillatory response to the marked feature [voiceless] that could be extracted and grouped. Again, this is potentially because of the construction of a short-term auditory memory trace to the voiceless standards, considering that the voiceless standards had the marked feature [voiceless] that could be extracted and grouped.

Not all accounts of neurophysiological and perceptual asymmetries rely on featural underspecification. Previous neurophysiological asymmetries have been attributed to low-level feature detectors based on specific acoustic properties that differ between standards and deviants (see the work of Maiste et al., 1995, for spectral differences between [ba] and [da]; Hisagi, Shafer, Strange, & Sussman, 2010, 2015; see the work of Ylinen et al., 2006, for vowel duration differences). Given that we did not find reliable unified differences separating voiced and voiceless stimuli in any of the acoustic dimensions that we measured in our stimuli, this is unlikely to be able to provide a plausible explanation for the effects found in this study. Moreover, if voiceless sounds are indeed specified for their voicing and voiced sounds are underspecified, our results are opposite of the work of Bishop, O'Reilly, and McArthur (2005). There, an MMN was observed only when frequency-modulated tones—which are presumably marked for an additional acoustic property—served as deviants in the context of unmodulated sinusoid standards, whereas no MMN was observed in the opposite configuration.

Perceptual asymmetries have also been observed in vowel perception, with discrimination performance depending on the order in which the stimuli are presented. Several behavioral studies have shown that it is easier to discriminate contrasts that move from more central to more peripheral, and thus more acoustically focal, regions of the vowel space than when the same stimuli are presented in the opposite direction (Masapollo, Polka, Molnar, & Ménard, 2017; Polka & Bohn, 2003, 2011); however, neurophysiological evidence for these asymmetries is less robust (de Rue et al., 2021; Polka, Molnar, Zhao, & Masapollo, 2021; Riedinger, Nagels, Werth, & Scharinger, 2021). In addition to vowel peripherality and focality, prototypicality and frequency also result in perceptual asymmetries (Kuhl et al., 2008; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). It is unclear, however, how easily notions of peripherality, centrality, and prototypically in a two-dimensional vowel space translate into the consonant domain, where the space is less continuous and potentially more multidimensional.

Frequency might also be a factor in the current results. Both lexical frequency, that is, how common a phoneme or a word is in a language, and cross-linguistic frequency, that is, how common a phoneme is across the world's languages, have been shown to affect MMN results. Lexical frequency appears to result in asymmetric neurophysiological responses (Jacobsen et al., 2004, 2021; Alexandrov, Boricheva, Pulvermüller, & Shtyrov, 2011; Jacobsen, Schröger, Winkler, & Horváth, 2005). If larger MMNs are observed to higher-frequency deviants relative to lower-frequency deviants (Alexandrov et al., 2011), then we might predict a larger MMN to the voiceless deviants, as the voiceless consonants in our study are more frequent in English than the voiced consonants (Wang & Crawford, 1960). We, however, found the opposite pattern: Voiced deviants in the context of voiceless standards resulted in larger MMNs. It is also possible that the cross-linguistic frequency of a given phoneme or phoneme class might affect MMN responses. For example, Shafer, Schwartz, and Kurtzberg (2004) observed an earlier negativity in Hindi participants compared to English participants when retroflex [ɖ] served as the standard and [b] was the deviant, compared to the opposite configuration. They posit that this asymmetry might be because retroflex consonants are cross-linguistically less common. In the context of the current experiment, voiced consonants are cross-linguistically less common than voiceless consonants (Maddieson, 1984). Given the Shafer et al. (2004) results, this might predict earlier/larger MMNs to the higher-frequency voiceless deviants in the context of the lower-frequency voiced standards. Again, we observed the opposite pattern, suggesting that neither lexical frequency nor cross-linguistic frequency is principally responsible for the effects observed in the current experiment.

There is now emerging evidence favoring not only underspecified features being stored, as described above, but also cross-linguistic evidence that auditory cortex supports featural underspecification in distinct laryngeal systems. Using Japanese, where [voiced] is the putatively marked feature, Hestvik et al. (2020) observed a larger MMN when voiced stops were the standard compared to when voiceless stops were the standard. And using Danish, Højlund, Gebauer, McGregor, and Wallentin (2019) observed a larger MMN when [d] was the standard compared to when [t^{h}] was the standard. They posit that this is potential evidence for the claim that voiced consonants are marked in Danish, whereas voiceless consonants are unmarked.

Future research could include designs that synthetically minimize acoustic differences between different speech sound classes. The goal would be to further isolate the contribution of the abstract phonological feature [voiceless] in the elicitation of the MMN in intercategory oddball designs. For the items in the current experiment, individual secondary acoustic–phonetic cues to voicing do not reliably predict voicing category membership (see Table 1 and Figure 2); however, it is possible that multiple secondary cues (e.g., f0 and F1, duration and f0) might conspire to provide a reliable signal-driven cue to phonological voicing. If true, this would indicate that participants could rely on a combination of acoustic–phonetic cues to identify phonological category membership. This could be tested with a design systematically controlling all acoustic cues in the stimuli. Moreover, different languages could be brought to bear on the current research questions. For example, Spanish and English speakers could be compared on the Spanish and English voicing contrasts. Recall that Spanish voiced stops and voiced fricatives use the same phonetic cue to phonological voicing, whereas English stops and fricatives use distinct phonetic cues. Such cross-linguistic comparisons could further emphasize the role that (i) the native language voicing inventory plays in the construction of auditory memory traces that reflect abstract phonological features and (ii) the distribution of different phonetic cues plays in the construction of phonological features. If acoustic–phonetic features, rather than an abstract phonological feature, were the primary drivers of the MMN effect, we would expect to find similar performance by the two language groups when listening to the same stimuli. On the other hand, if phonology is the main driver, as we expect, we would posit different performance across the two groups.

A limitation of the current study is that given the number of deviants used in the current design, it is difficult to know whether stops or fricatives played distinct roles in the perceptual grouping of voicing or the elicitation of the MMN. Each block contained 75 deviants. Forty-five of these deviants were stops and 30 were fricatives. This results in lower than desired power to determine their independent effects. For example, given the distinct nature of how voicing is coded in stops versus fricatives, it might be the case that voiced stops resulted in a larger MMN than voiced fricatives in the context of voiceless standards; however, the ratios of stops to fricatives were similar in the standards and deviants across both blocks, and so this is unlikely to be the case. At this point, we do not have clear predictions about which manner of articulation might have an increased effect on the MMN, but the current design makes it difficult to ascertain. We leave this possibility for future research.

The current results support theories of language processing that posit that speech sounds are stored in terms of featural representations and that these features are utilized during on-line spoken word recognition (Hickok, 2014; Poeppel & Monahan, 2011; Gow, 2003; Stevens, 2002; McClelland & Elman, 1986; Halle & Stevens, 1962). Whether researchers should search for linguistic units of representation in perception is contentious. In particular, Goldinger and Azuma (2003) and Samuel (2020) have argued against positing the phoneme and, to a lesser extent, the feature as a unit of language processing and perception. Their position is that the search for identifying units as conceptualized in linguistics within perception is misguided. As reviewed above, however, there is a body of neurophysiological evidence that is consistent with feature-based speech sound representations. Moreover, various models of feature representations (e.g., underspecification) make specific predictions regarding neurophysiological response asymmetries, and these patterns have been replicated across several speech sound classes and in different languages (see the work of Monahan, 2018, for a review). In particular, the current results indicate a brain signature underlying abstract groupings of speech sounds and are consistent with predictive theories of processing that are at least partially based on phonological features, such as analysis by synthesis (Halle & Stevens, 1962; Poeppel & Monahan, 2011).

Conclusion

The current experiment provides neurophysiological evidence that distinct phonetic cues that are functionally linked in a language's phonology are grouped together. This is shown by the presence of an MMN when hearing deviants that differ from the standards in their phonological voicing. This is true even when the speech sounds encompass phonetically distinct realizations of a single phonological feature. The MMN was asymmetric, consistent with previous studies arguing for underspecified featural representations. Here, the results support the conclusion that voiceless consonants in English are specified as [voiceless], whereas voiced consonants are underspecified for voicing. Moreover, we observed a difference between the voiced and voiceless standards in the prestimulus baseline, suggesting that predictions regarding which sounds come next can be constructed from specified features. These results point toward auditory cortex's ability to disjunctively code spectral and temporal phonetic cues when they are functionally joined and, ultimately, to encode abstract phonological features. In summary, the current work demonstrates that linguistic units and, specifically, features do have a role to play in understanding the cognitive neuroscience of language.

APPENDIX A

We calculated the ERPs to the four standard conditions (i.e., voiced fricative standards, voiced stop standards, voiceless fricative standards, and voiceless stop standards). The ERPs are presented in Figure A1 (rug plot, top: voicing; rug plot, bottom: manner). As expected, there are differences in the responses between the four conditions. Most time points showed a main effect of manner. A main effect of voicing was observed in three principal time windows: 68–160 msec, 212–252 msec, and 292–448 msec. There is a question as to whether certain subcategories of speech sounds (e.g., voiced fricatives) could drive the MMN responses reported in the article. To address this, we computed the ratios of stops to fricatives in the standards and deviants after data preprocessing. If there were substantial imbalances between these ratios in the standards and deviants, then this could potentially result in some subcategories driving the MMN response more than others. The ratio of stops to fricatives in the voiced standards (1:1.44) is similar to the ratio of stops to fricatives in the voiceless deviants (1:1.47). The same is true for the voiceless standards (1:1.52) and voiced deviants (1:1.54). The differences in stop–fricative ratios in the iMMN comparisons are slightly larger (voiceless standards: 1:1.52, voiceless deviants: 1:1.47; voiced standards: 1:1.44, voiced deviants: 1:1.54), but again, quite similar. Given that the stop–fricative ratios are similar between standards and deviants, it is unlikely that any one subcategory is driving the MMN when comparing standards to deviants.

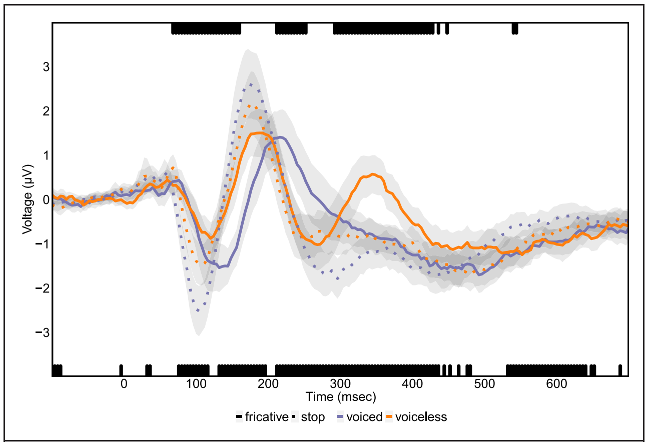

Figure A1. Average evoked ERP responses over fronto-central electrode sites (i.e., Fz, FC1/2, Cz, C3/4, CP1/2) to the voiced fricatives, voiced stops, voiceless fricatives, and voiceless stops. The ERP responses to fricatives are indicated with solid lines, and the ERP responses to stops are indicated with dotted lines. The ERP responses to voiceless consonants are indicated in orange, and the ERP responses to voiced consonants are indicated in purple. Rug plots along the *x* axes indicate time points when the comparison was statistically significant (see text for analysis procedures). Rug plots along the top of the figure indicate time points in which there is a main effect of voicing. Rug plots along the bottom of the figure indicate time points in which there is a main effect of manner. Shaded regions represent the 95% confidence interval of the mean.

Acknowledgments

We thank members of the Computation and Psycholinguistics Laboratory at the University of Toronto Scarborough for their assistance.

Reprint requests should be sent to Philip J. Monahan, Department of Language Studies, University of Toronto Scarborough, 1265 Military Trail, Toronto ON M1C 1A4, Canada, or via e-mail: philip.monahan@utoronto.ca.

Author Contributions

Philip J. Monahan: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology;

Resources; Supervision; Visualization; Writing—Original draft; Writing—review & editing. Jessamyn Schertz: Conceptualization; Investigation; Methodology; Writing—review & editing. Zhanao Fu: Formal analysis; Investigation; Methodology; Visualization; Writing—review & editing. Alejandro Pérez: Formal analysis; Investigation; Methodology; Visualization; Writing—Original draft; Writing—review & editing.

Funding Information

A. P. received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant number: 840885. This work was supported in part by the Social Sciences and Humanities Research Council (SSHRC) (http://dx.doi.org/10.13039/501100000155) of Canada, grant number: IDG 430-15-00647, and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, grant number: RGPIN-2017-06053.

Data Availability Statement

Stimuli, acoustic measurements, and EEG datasets are publicly available in the Open Science Framework: https://osf.io/jn5cr/.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .606, W/M = .213, M/W = .096, and W/W = .085.

REFERENCES

Alexandrov, A. A., Boricheva, D. O., Pulvermüller, F., & Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS One*, *6*, e22999. https://doi.org/10.1371/journal.pone.0022999, PubMed: 21853063

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, *16*, 38–51. https://doi.org/10.1097/00003446-199502000-00004, PubMed: 7774768

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

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*, 390–398. https://doi.org/10.1016/j.tics.2012.05.003, PubMed: 22682813

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*, 797–801. https://doi.org/10.1038/nn.2810, PubMed: 21552273

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R., & Sams, M. (1993). Phonetic invariance in the human auditory cortex. *NeuroReport*, *4*, 1356–1358. https://doi.org/10.1097/00001756-199309150-00018, PubMed: 8260620

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 41–70). Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110886672.41

Baković, E. (2014). Phonemes, segments and features. *Language, Cognition and Neuroscience*, *29*, 21–23. https://doi.org/10.1080/01690965.2013.848992

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, *85*, 390–401. https://doi.org/10.1016/j.neuron.2014.12.018, PubMed: 25556836

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language*, *141*, 62–69. https://doi.org/10.1016/j.bandl.2014.11.003, PubMed: 25540857

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*, 16. https://doi.org/10.3389/fninf.2015.00016, PubMed: 26150785

Bishop, D. V. M., Hardiman, M. J., & 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? *Developmental Science*, *14*, 402–416. https://doi.org/10.1111/j.1467-7687.2010.00990.x, PubMed: 22213909

Bishop, D. V. M., O'Reilly, J., & McArthur, G. M. (2005). Electrophysiological evidence implicates automatic low-level feature detectors in perceptual asymmetry. *Cognitive Brain Research*, *24*, 177–179. https://doi.org/10.1016/j.cogbrainres.2004.12.007, PubMed: 15922169

Buschman, T. J., & Miller, E. K. (2007). Top–down versus bottom–up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*, 1860–1862. https://doi.org/10.1126/science.1138071, PubMed: 17395832

Čeponienė, R., Yaguchi, K., Shestakova, A., Alku, P., Suominen, K., & Näätänen, R. (2002). Sound complexity and 'speechness' effects on pre-attentive auditory discrimination in children. *International Journal of Psychophysiology*, *43*, 199–211. https://doi.org/10.1016/S0167-8760(01)00172-6, PubMed: 11850086

Cheour, M., Korpilahti, P., Martynova, O., & Lang, A.-H. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. *Audiology and Neuro-Otology*, *6*, 2–11. https://doi.org/10.1159/000046804, PubMed: 11173771

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press.

Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 245–306). Oxford: Blackwell.

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, *1394*, 79–89. https://doi.org/10.1016/j.brainres.2011.04.001, PubMed: 21549357

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 757–772. https://doi.org/10.1037/a0030862, PubMed: 23276108

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 909–921. https://doi.org/10.1037//0278-7393.19.4.909, PubMed: 8345328

Datta, H., Hestvik, A., Vidal, N., Tessel, C., Hisagi, M., Wróblewski, M., et al. (2020). Automaticity of speech processing in early bilingual adults and children. *Bilingualism: Language and Cognition*, *23*, 429–445. https://doi.org/10.1017/S1366728919000099, PubMed: 32905492

Deacon, D., Gomes, H., Nousak, J. M., Ritter, W., & Javitt, D. (2000). Effect of frequency separation and stimulus rate on the mismatch negativity: An examination of the issue of refractoriness in humans. *Neuroscience Letters*, *287*, 167–170. https://doi.org/10.1016/S0304-3940(00)01175-7, PubMed: 10863021

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, *12*, 449–460. https://doi.org/10.1162/089892900562264, PubMed: 10931771

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, *7*, e30135. https://doi.org/10.1371/journal.pone.0030135, PubMed: 22355308

de Rue, N. P. W. D., Snijders, T. M., & Fikkert, P. (2021). Contrast and conflict in Dutch vowels. *Frontiers in Human Neuroscience*, *15*, 629648. https://doi.org/10.3389/fnhum.2021.629648, PubMed: 34163338

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin: Mouton de Gruyter.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035, PubMed: 23791839

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—Signalling the status quo? *Current Opinion in Neurobiology*, *20*, 156–165. https://doi.org/10.1016/j.conb.2010.02.015, PubMed: 20359884

Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, *5*, 151–166. https://doi.org/10.1159/000013877, PubMed: 10859410

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*, 577–583. https://doi.org/10.1162/089892904323057308, PubMed: 15185677

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications*, *5*, 4694. https://doi.org/10.1038/ncomms5694, PubMed: 25178489

Fox, N. P., Leonard, M., Sjerps, M. J., & Chang, E. F. (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *eLife*, *9*, e53051. https://doi.org/10.7554/eLife.53051, PubMed: 32840483

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*, 355–373. https://doi.org/10.1016/S0149-7634(01)00019-7, PubMed: 11445140

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622, PubMed: 15937014

Fu, Z., & Monahan, P. J. (2021). Extracting phonetic features from natural classes: A mismatch negativity study of Mandarin Chinese retroflex consonants. *Frontiers in Human Neuroscience*, *15*, 609898. https://doi.org/10.3389/fnhum.2021.609898, PubMed: 33841113

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2009). Dynamic causal modeling of the response to frequency deviants. *Journal of Neurophysiology*, *101*, 2620–2631. https://doi.org/10.1152/jn.90291.2008, PubMed: 19261714

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305–320. https://doi.org/10.1016/S0095-4470(03)00030-5, PubMed: 29093608

Gomes, H., Ritter, W., & Vaughan, H. G. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neuroscience*, *7*, 81–94. https://doi.org/10.1162/jocn.1995.7.1.81, PubMed: 23961755

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*, 575–590. https://doi.org/10.3758/BF03194584, PubMed: 12812280

Gruber, T., Malinowski, P., & Müller, M. M. (2004). Modulation of oscillatory brain activity and evoked potentials in a repetition priming task in the human EEG. *European Journal of Neuroscience*, *19*, 1073–1082. https://doi.org/10.1111/j.0953-816X.2004.03176.x, PubMed: 15009155

Halle, M. (1983). On distinctive features and their articulatory implementation. *Natural Language & Linguistic Theory*, *1*, 91–105. https://doi.org/10.1007/BF00210377

Halle, M. (2002). *From memory to speech and back: Papers on phonetics and phonology 1954–2002*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110871258

Halle, M., & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, *8*, 155–159. https://doi.org/10.1109/TIT.1962.1057686

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007, PubMed: 26705957

Hestvik, A., Shinohara, Y., Durvasula, K., Verdonschot, R. G., & Sakai, H. (2020). Abstractness of human speech sound representations. *Brain Research*, *1732*, 146664. https://doi.org/10.1016/j.brainres.2020.146664, PubMed: 31930995

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*, 2–20. https://doi.org/10.1080/01690965.2013.834370, PubMed: 24489420

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research*, *1360*, 89–105. https://doi.org/10.1016/j.brainres.2010.08.092, PubMed: 20816759

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2015). Neural measures of a Japanese consonant length discrimination by Japanese and American English listeners:

Effects of attention. *Brain Research*, *1626*, 218–231. https://doi.org/10.1016/j.brainres.2015.06.001, PubMed: 26119918

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

Hu, A., Gu, F., Wong, L. L. N., Tong, X., & Zhang, X. (2020). Visual mismatch negativity elicited by semantic violations in visual words. *Brain Research*, *1746*, 147010. https://doi.org/10.1016/j.brainres.2020.147010, PubMed: 32663455

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *Journal of Neuroscience*, *36*, 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016, PubMed: 26865624

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*, 369–396. https://doi.org/10.1017/S0952675700002566

Jacobsen, T., Bäß, P., Roye, A., Winkler, I., Schröger, E., & Horváth, J. (2021). Word class and word frequency in the MMN looking glass. *Brain and Language*, *218*, 104964. https://doi.org/10.1016/j.bandl.2021.104964, PubMed: 33964668

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., & Winkler, I. (2004). Pre-attentive auditory processing of lexicality. *Brain and Language*, *88*, 54–67. https://doi.org/10.1016/S0093-934X(03)00156-1, PubMed: 14698731

Jacobsen, T., Schröger, E., Winkler, I., & Horváth, J. (2005). Familiarity affects the processing of task-irrelevant auditory deviance. *Journal of Cognitive Neuroscience*, *17*, 1704–1713. https://doi.org/10.1162/089892905774589262, PubMed: 16269107

Jakobson, R., Fant, G., & Halle, M. (1961). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Kazanina, N., Bowers, J. S., & Idsardi, W. J. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*, *25*, 560–585. https://doi.org/10.3758/s13423-017-1362-0, PubMed: 28875456

Kazanina, N., Phillips, C., & Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 11381–11386. https://doi.org/10.1073/pnas.0604821103, PubMed: 16849423

Khalighinejad, B., da Silva, G. C., & Mesgarani, N. (2017). Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience*, *37*, 2176–2185. https://doi.org/10.1523/JNEUROSCI.2383-16.2017, PubMed: 28119400

Kraus, N., McGee, T., Sharma, A., Carrell, T., & Nicol, T. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear and Hearing*, *13*, 158–164. https://doi.org/10.1097/00003446-199206000-00004, PubMed: 1397755

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *363*, 979–1000. https://doi.org/10.1098/rstb.2007.2154, PubMed: 17846016

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

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 637–675). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197105.637

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

Lanwermeyer, M., Henrich, K., Rocholl, M. J., Schnell, H. T., Werth, A., Herrgen, J., et al. (2016). Dialect variation influences the phonological and lexical-semantic word processing in sentences. Electrophysiological evidence from a cross-dialectal comprehension study. *Frontiers in Psychology*, *7*, 739. https://doi.org/10.3389/fpsyg.2016.00739, PubMed: 27303320

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*, 567–576. https://doi.org/10.1093/cercor/bhr119, PubMed: 21690262

Lewis, A. G., Schoffelen, J.-M., Schriefers, H., & Bastiaansen, M. (2016). A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Frontiers in Human Neuroscience*, *10*, 85. https://doi.org/10.3389/fnhum.2016.00085, PubMed: 26973500

Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*, *148*, 51–63. https://doi.org/10.1016/j.bandl.2015.01.003, PubMed: 25666170

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461. https://doi.org/10.1037/h0020279, PubMed: 4170865

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*, 3–11. https://doi.org/10.1177/002383098602900102, PubMed: 3657346

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830, PubMed: 3657346

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511753459

Mahajan, Y., Peter, V., & Sharma, M. (2017). Effect of EEG referencing methods on auditory mismatch negativity. *Frontiers in Neuroscience*, *11*, 560. https://doi.org/10.3389/fnins.2017.00560, PubMed: 29066945

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, *16*, 68–89. https://doi.org/10.1097/00003446-199502000-00006, PubMed: 7774771

Martynova, O., Kirjavainen, J., & Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters*, *340*, 75–78. https://doi.org/10.1016/S0304-3940(02)01401-5, PubMed: 12668240

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *Journal of the Acoustical Society of America*, *141*, 2857–2869. https://doi.org/10.1121/1.4981006, PubMed: 28464636

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McGee, T., King, C., Tremblay, K., Nicol, T., Cunningham, J., & Kraus, N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, *38*, 653–658. https://doi.org/10.1111/1469-8986.3840653, PubMed: 11446578

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010. https://doi.org/10.1126/science.1245994, PubMed: 24482117

Molinaro, N., Monsalve, I. F., & Lizarazu, M. (2016). Is there a common oscillatory brain mechanism for producing and predicting language? *Language, Cognition and Neuroscience*, *31*, 145–158. https://doi.org/10.1080/23273798.2015.1077978

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

Morillon, B., & Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction: Neuronal oscillations and temporal predictions. *Annals of the New York Academy of Sciences*, *1337*, 26–31. https://doi.org/10.1111/nyas.12629, PubMed: 25773613

Mullen, T., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2184–2187. https://doi.org/10.1109/EMBC.2013.6609968

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. https://doi.org/10.1111/1469-8986.3810001, PubMed: 11321610

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*, 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x, PubMed: 15720578

Näätänen, R., & Kreegipuu, K. (2012). The mismatch negativity (MMN). In Emily S. Kappenman & Steven J. Luck (Eds.), *The Oxford handbook of event-related potential components* (pp. 143–157). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0081

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. https://doi.org/10.1038/385432a0, PubMed: 9009189

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026, PubMed: 17931964

Nousak, J. M. K., Deacon, D., Ritter, W., & Vaughan, H. G. (1996). Storage of information in transient auditory memory. *Cognitive Brain Research*, *4*, 305–317. https://doi.org/10.1016/S0926-6410(96)00068-7, PubMed: 8957572

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex*, *22*, 2466–2477. https://doi.org/10.1093/cercor/bhr325, PubMed: 22100354

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932, PubMed: 5146491

Oostenveld, R., & Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human Brain Mapping*, *17*, 179–192. https://doi.org/10.1002/hbm.10061, PubMed: 12391571

Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2012). *AMICA: An adaptive mixture of independent component analyzers with shared components* (p. 15, Technical Report). Swartz Center for Computational Neuroscience, University of California San Diego.

Pérez, A., Monahan, P. J., & Lambon Ralph, M. A. (2021). Joint recording of EEG and audio signals in hyperscanning and pseudo-hyperscanning experiments. *MethodsX*, *8*, 101347. https://doi.org/10.1016/j.mex.2021.101347, PubMed: 34430250

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*, 184–187. https://doi.org/10.1016/0013-4694(89)90180-6, PubMed: 2464490

Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, *482*, 71–75. https://doi.org/10.1016/j.neulet.2010.07.010, PubMed: 20630487

Peter, V., McArthur, G., & Thompson, W. F. (2012). Discrimination of stress in speech and music: A mismatch negativity (MMN) study. *Psychophysiology*, *49*, 1590–1600. https://doi.org/10.1111/j.1469-8986.2012.01472.x, PubMed: 23066846

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*, 1038–1055. https://doi.org/10.1162/08989290051137567, PubMed: 11177423

Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26*, 935–951. https://doi.org/10.1080/01690965.2010.493301

Polich, J. (1993). Cognitive brain potentials. *Current Directions in Psychological Science*, *2*, 175–179. https://doi.org/10.1111/1467-8721.ep10769728

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019, PubMed: 17573239

Politzer-Ahles, S., Schluter, K. T., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1547–1570. https://doi.org/10.1037/xhp0000242, PubMed: 27195767

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

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

Polka, L., Molnar, M., Zhao, T. C., & Masapollo, M. (2021). Neurophysiological correlates of asymmetries in vowel perception: An English–French cross-linguistic event-related potential study. *Frontiers in Human Neuroscience*, *15*, 274. https://doi.org/10.3389/fnhum.2021.607148, PubMed: 34149375

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*, 49–71. https://doi.org/10.1016/j.pneurobio.2006.04.004, PubMed: 16814448

Riddle, J., Hwang, K., Cellier, D., Dhanani, S., & D'Esposito, M. (2019). Causal evidence for the role of neuronal oscillations in top–down and bottom–up attention. *Journal of Cognitive Neuroscience*, *31*, 768–779. https://doi.org/10.1162/jocn_a_01376, PubMed: 30726180

Riedinger, M., Nagels, A., Werth, A., & Scharinger, M. (2021). Asymmetries in accessing vowel representations are driven by phonological and acoustic properties: Neural and behavioral evidence from natural German minimal pairs. *Frontiers in Human Neuroscience*, *15*, 612345. https://doi.org/10.3389/fnhum.2021.612345, PubMed: 33679344

Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: Different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, *28*, 21–29. https://doi.org/10.1111/j.1469-8986.1991.tb03382.x, PubMed: 1886961

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*, 437–448. https://doi.org/10.1016/0168-5597(85)90054-1

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, *111*, 104070. https://doi.org/10.1016/j.jml.2019.104070

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2011). You had me at "Hello": Rapid extraction of dialect information from spoken words. *Neuroimage*, *56*, 2329–2338. https://doi.org/10.1016/j.neuroimage.2011.04.007, PubMed: 21511041

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research*, *55*, 903–918. https://doi.org/10.1044/1092-4388(2011/11-0065), PubMed: 22232394

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *Neuroimage*, *128*, 293–301. https://doi.org/10.1016/j.neuroimage.2016.01.003, PubMed: 26780574

Schluter, K. T., Politzer-Ahles, S., Al Kaabi, M., & Almeida, D. (2017). Laryngeal features are phonetically abstract: Mismatch negativity evidence from Arabic, English, and Russian. *Frontiers in Psychology*, *8*, 746. https://doi.org/10.3389/fpsyg.2017.00746, PubMed: 28555118

Schluter, K. T., Politzer-Ahles, S., & Almeida, D. (2016). No place for /h/: An ERP investigation of English fricative place features. *Language, Cognition and Neuroscience*, *31*, 728–740. https://doi.org/10.1080/23273798.2016.1151058, PubMed: 27366758

Shafer, V. L., Schwartz, R. G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, *18*, 242–254. https://doi.org/10.1016/j.cogbrainres.2003.10.007, PubMed: 14741311

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, *106*, 1078–1083. https://doi.org/10.1121/1.428048, PubMed: 10462812

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, *107*, 2697–2703. https://doi.org/10.1121/1.428655, PubMed: 10830391

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, *25*, 471–500. https://doi.org/10.1006/jpho.1997.0053

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

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, *111*, 1872–1891. https://doi.org/10.1121/1.1458026, PubMed: 12002871

Strotseva-Feinschmidt, A., Cunitz, K., Friederici, A. D., & Gunter, T. C. (2015). Auditory discrimination between function words in children and adults: A mismatch negativity study. *Frontiers in Psychology*, *6*, 1930. https://doi.org/10.3389/fpsyg.2015.01930, PubMed: 26733918

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *Neuroimage*, *9*, 330–336. https://doi.org/10.1006/nimg.1999.0405, PubMed: 10075902

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 20754–20759. https://doi.org/10.1073/pnas.1117807108, PubMed: 22147913

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, *90*, 1195–1268. https://doi.org/10.1152/physrev.00035.2008, PubMed: 20664082

Wang, W. S.-Y., & Crawford, J. (1960). Frequency studies of English consonants. *Language and Speech*, *3*, 131–139. https://doi.org/10.1177/002383096000300302

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*, 35–44. https://doi.org/10.3758/BF03207136, PubMed: 3991316

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*, 147–163. https://doi.org/10.1027/0269-8803.21.34.147

Winkler, I., Kujala, T., Alku, P., & Näätänen, R. (2003). Language context and phonetic change detection. *Cognitive Brain Research*, *17*, 833–844. https://doi.org/10.1016/S0926-6410(03)00205-2, PubMed: 14561466

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. *Cognitive Brain Research*, *7*, 357–369. https://doi.org/10.1016/S0926-6410(98)00039-1, PubMed: 9838192

Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätanen, R. (1990). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology*, *27*, 228–235. https://doi.org/10.1111/j.1469-8986.1990.tb00374.x, PubMed: 2247552

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, *102*, 1096–1110. https://doi.org/10.1016/j.neuron.2019.04.023, PubMed: 31220442

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, *1072*, 175–185. https://doi.org/10.1016/j.brainres.2005.12.004, PubMed: 16426584

Yu, Y. H., & Shafer, V. L. (2021). Neural representation of the English vowel feature [high]: Evidence from /ε/ vs. /ɪ/. *Frontiers in Human Neuroscience*, *15*, 629517. https://doi.org/10.3389/fnhum.2021.629517, PubMed: 33897394

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. https://doi.org/10.1097/00001756-200512190-00009, PubMed: 16317345