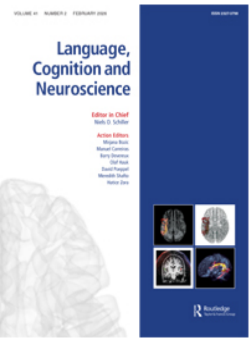


Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: www.tandfonline.com/journals/plcp21

Integrated versus independent processing of auditory features in speech sounds

Alex Chabot, Ellen Lau, Philip J. Monahan & William J. Idsardi

To cite this article: Alex Chabot, Ellen Lau, Philip J. Monahan & William J. Idsardi (04 Mar 2026): Integrated versus independent processing of auditory features in speech sounds, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2026.2630749

To link to this article: https://doi.org/10.1080/23273798.2026.2630749

Published online: 04 Mar 2026.

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2026.2630749

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Integrated versus independent processing of auditory features in speech sounds

Alex Chabot[a], Ellen Lau[a], Philip J. Monahan[b,c,d] and William J. Idsardi[a]

[a]Department of Linguistics, University of Maryland, College Park, MD, USA; [b]Department of Linguistics, University of Toronto, Toronto ON, Canada; [c]Department of Language Studies, University of Toronto Scarborough, Toronto ON, Canada; [d]Department of Psychology, University of Toronto Scarborough, Toronto ON, Canada

ABSTRACT
Two MMN experiments investigate integrated versus independent processing of complex auditory information in linguistic sound. Our study asks if electrocortical techniques can be used to find feature additivity in the perception of linguistic sound and tests the efficacy of a roving-standard design, which allows for the direct comparison of multiple deviant contexts within a single session. We hypothesise that neurophysiologically evoked responses to multiple distinctive cues are stronger than responses to single cues. Two magnetoencephalographic (MEG) protocols are deployed, which are identical except for the order of presentation of standard and deviant stimuli. We observed larger evoked-response fields to deviants that differed in two phonetic features relative to deviants that differed in a single feature from the standard. Our results are consistent with the independent processing of acoustic cues and indicate that this novel methodology may be useful in testing questions of feature additivity.

ARTICLE HISTORY
Received 18 July 2025
Accepted 23 January 2026

KEYWORDS
Phonological processing; phonological features; additivity; magneticencephalography (MEG); evoked-response fields (ERF); mismatch negativity (MMN)

1. Introduction

Perception requires extracting useful representations from a physical world that contains a potentially boundless range of sensory information, filtering out a great deal of redundant or irrelevant input. One method of improving processing efficiency is by sensory dimensionality reduction, as when distal stimuli with correlated cues are processed in an integrated way (Lerousseau et al., 2010; Parise & Ernst, 2016). For example, in vision, when participants organise decks of coloured stimulus cards which co-vary in both hue and chroma, they disregard redundant dimensions and treat both dimensions as a single feature, rapidly sorting cards into colour categories despite the variation (Garner & Felfoldy, 1970). This suggests that not every perceivable dimension is given equal attention, and during perception, hue and chroma are processed in an integrated, holistic fashion. However, some dimensions, such the size of the colour circle and the angle of an inscribed line, slow the organisation of the stimuli decks, suggesting that attention to non-correlated dimensions is required and that these visual features are processed independently. In another domain, musicians exhibit sub-additive neurophysiological responses to changes in frequency in melody perception when those changes are coupled with changes in intensity and perceived location, suggesting that frequency-related deviations are processed in an integrated manner with other acoustic changes, rather than as independent events (Hansen et al., 2022).

This study investigates the extent to which speech perception makes use of independent processing of linguistically relevant cues, rather than integrated processing alone. Language is an interesting empirical domain of related inquiry because there are reasonable hypotheses about tractable links between our theories of linguistic representation and our understanding of how the brain representationally encodes sensory objects (see Embick & Poeppel, 2015; Poeppel & Embick, 2005 for discussion). Additionally, speech sounds are well understood in terms of their production, perception, and neuro-cortical correlates (Grimaldi, 2018; Monahan, 2018; Poeppel & Monahan, 2008; Poeppel et al., 2007). However, because languages differ in which acoustic cues are significant, there is a possibility that the same co-occurring cues might be processed in an independent fashion in one language, but in an integrated fashion in another language. This means that the way the brain processes linguistically-significant sensory cues in perception may change depending on individual language experience.

The perception of speech sounds requires the simultaneous processing of multiple, co-occurring auditory cues, such as degree of occlusion or manner, voicing, and place of articulation. The conventional view in theoretical phonology holds that the representations

CONTACT Alex Chabot ✉ chabot@umd.edu Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742-7505, USA

Table 1. English obstruents at the labial place of articulation.

	Stops	Fricatives
Voiceless	[p]	[f]
Voiced	[b]	[v]

of speech sounds are not holistic but rather composed of atomic features which reoccur in various configurations across the entire phonological system (Jakobson, 1939; Jakobson et al., 1952). Features may be shared across sounds (e.g. [voice] in /b d g/), or present for some sounds but not others (e.g. [nasal] in /m n ŋ/ but not /b d g/). In this way, features organise the speech sounds of language and are the most basic representational units of phonological systems.

Representations are the mapping between internal mental states and external physical events (Gallistel, 1990), and like other kinds of representations, features must be instantiated in human brains (Mesgarani et al., 2014; Yi et al., 2019). Indeed, evidence from electrophysiological research suggests that speakers use features to parse continuous speech (Fu & Monahan, 2021; Monahan et al., 2022; Politzer-Ahles & Jap, 2024); however, it is not known if the set of relevant organisational features is universal to all languages, or if features are language specific (Mielke, 2008). In one view, all languages make use of the same universal features (Chomsky & Halle, 1968). In another, the only features in the acoustic signal that are relevant to speakers are those that are linguistically significant and form the basis for language-specific contrasts (Dresher, 2009). This study contributes experimental evidence in the form of a neuro-cortical correlate for acoustic cues that are present in the speech signal and linguistically significant in some languages, but not all.

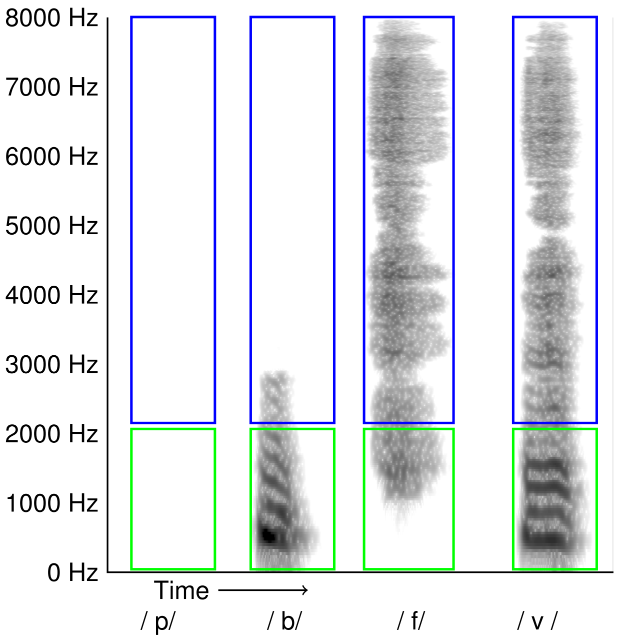


Figure 1. Spectrograms of stops and fricatives, showing the cross-classification of higher-frequency noise (blue boxes) and lower-frequency periodicity (green boxes). Dark bands are the harmonics of the fundamental frequency. As /p/ is a voiceless stop, it produces no noise or periodicity. Similarly, though there is some low-frequency noise for /f/, there is no periodicity, as the harmonic bands are absent.

For example, in English, voiceless stops, such as /p/, are typically realised with noticeable post-closure aspiration [p^{h}], but this aspiration is not required for voiceless stops to be categorised as such, in phonological terms, as aspiration is generally predictable from voicing and prosodic position and does not constitute an independent phonological contrast. Rather, aspiration functions as one of several phonetic cues that can realise the voicing distinction in English (Lisker, 1986), and in some contexts may reinforce lexical contrasts grounded in voicing rather than introduce new ones. Phonological patterns, such as plural allomorphy and voicing assimilation, operate over only two categories: voiced and voiceless. Thus, the typical realisation of *put* is [p^{h}ʊt̚], but the realisation [pʊt̚] does not change the meaning of this word for English speakers. In Hindi, in contrast, whether or not a stop is realised with aspiration can change the meaning of words, as for example in [p^{h}ɑːl] ’knife blade’ and [pɑːl] “nurture”, suggesting that in this language, aspiration is critical for categorisation.

Linguists typically organise speech-sound inventories according to these phonetic cues. For example, in Table 1, four speech sounds relevant to English are organised according to a cross-classification of their phonetic features. Both [p] and [b] are bilabial stops produced with total occlusion of the pulmonic air stream, but while [p] is typically realised without vocal fold vibration, [b] is realised with periodicity – the acoustic correlate of voicing in English (Lisker & Abramson, 1964). Similarly, both [f] and [v] are fricatives produced with partial occlusion, the acoustic correlate of which is broadband noise during articulation (see Figure 1). They contrast in voicing, as do [p] and [b], but differ from that pair in their manner of articulation.

Though there is variation in how noise is generated in stops and fricatives and the timing of laryngeal and supralaryngeal articulators (Haggard, 1978; Stevens et al., 1992), the essential dimensions are shared. The phonological organisation of manner and voicing can be understood as a two-by-two matrix where phonological categories line up in a linguistically significant way (Monahan et al., 2022; Politzer-Ahles & Jap, 2024; Schluter et al., 2017). Despite differences in their phonetic realisation, the set of stops and set of fricatives demonstrate a phonological parallelism in their contrastive relationships. In this way, the four sounds in Table 1 are decomposable into distinct configurations of the relevant acoustic cues, where discrete elements in the signal are distinctive in cognition precisely because they are the basis for linguistically-meaningful contrasts.

Figure 1 demonstrates how four speech sounds can be cross-classified in English according to the presence or absence of both noise and periodicity. The contrastive status of these cues in English suggests that cross-classification results in processing of these speech cues in an independent fashion, where the cues do not depend on each other and may variably co-occur or not. That is, their presence or absence is the basis by which speakers distinguish between the various sounds to make meaningful contrasts.

However, this full cross-classification configuration does not hold across all languages, since speech sounds may form complete or incomplete cross-classifications across various dimensions (see Table 2). The PHOIBLE cross-linguistic inventory of language sound systems is a catalog of 3,020 speech-sound inventories, of which 2,594 (86%) contain /p/, 1,906 (63%) contain /b/, 1,329 (44%) contain /f,/ and only 816 (27%) contain /v/. For example, the Acehnese language, spoken in Indonesia, has /p/, /b/, and /f/, but not /v/. Campidanese Sardinian, a Romance language spoken in Italy, has a voiced voiced labial fricative [ʱ], but only as a contextually predictable allophone with no significant contrastive status (Chabot, 2023).

This means that in some languages, the cross-classification, which holds in Table 2, is incomplete. In such cases, some cues that may be present in the acoustic signal are not phonologically significant, as they play no role in making a speech sound distinctive. Put another way, the role a sound plays in a phonological system is language-specific, and discrete elements of the acoustic signal may not be distinctive in cognition if they are not meaningfully contrastive. In such cases of incomplete cross-classification, instead of being independently processed, processing of acoustically complex speech sounds may proceed in an integrated manner.

This raises a question: Does language-specific experience determine whether perceptual processing of speech sounds proceeds in an independent or integrated manner? We propose a methodological approach in which independent processing of multiple auditory cues are distinguished from integrated processing through the identification of distinct neural evoked responses (Event-Related Fields, ERF) elicited by speech-sound stimuli (Janssen et al., 2020; Obleser et al., 2003). The hypothesis is that whether the multiple cues of speech sounds are processed in an integrated or in an independent fashion is determined by language-specific experience. Specifically, we predict that complete classification leads to independent processing of speech properties, whereas incomplete classification leads to integrated processing.

Table 2. Cross-linguistic comparison of stops and fricatives at the labial place of articulation in intervocalic contexts.

	English		Acehnese		Arabic		Campidanese Sardinian	
Voiceless	[p]	[f]	[p]	[f]	▨	[f]	[p]	[f]
Voiced	[b]	[v]	[b]	▨	[b]	▨	[b]	[ʱ]

Note: Gaps in the table correspond to sounds absent from that language's consonant inventory. Cells in gray are sounds which are absent or occur only as predictable allophones.

The goal of the present study is to determine if electro-cortical measures can be used to investigate whether fully cross-classified speech cues are processed independently or in an integrated fashion, providing a necessary first step toward testing predictions about how language-specific experience shapes perceptual processing. Specifically, we hypothesise that an observed additivity in the evoked responses implies independent processing of correlated cues, while a non-additive response suggests integrated processing of multiple cues.

In auditory and visual neuroscience, the Mismatch Negativity (MMN, Jääskeläinen et al., 2004; Näätänen et al., 2005) is a response in electro-cortical activity elicited by changes in multiple stimuli within the time span of sensory memory, detectable in both EEG and MEG. When two or more comparable stimuli are presented in the same context, aspects of those incoming stimuli are compared to traces stored in working memory where divergences from expected cues elicit robust MMN responses (see Näätänen et al., 2019 for an overview). As the perceptual system builds representations of environmental stimuli, it makes predictions about what kinds of cues will be most salient (Bendixen et al., 2009; Winkler, 2007). Divergences from those predictions result in a robust MMN evoked response. Thus, the MMN is a correlate for aspects of perceptual representations that diverge from a standard in a perceptible fashion; in audition, this includes cues such as duration, intensity, and pitch (Giard et al., 1995; Gomes et al., 1995). These cues are present in any given speech signal, and may vary independently of each other. For example, Wolff and Schröger (2001) show that with auditory stimuli which vary along multiple dimensions – including duration, frequency, and intensity – when one dimension is changed between stimuli, this is sufficient to elicit an MMN.

When partially different neural populations are involved in processing different but related stimuli

(Allen et al., 2017, 2022), the MMNs elicited by divergent features demonstrate neural additivity (Paavilainen et al., 2001), reflecting a neural correlate of perceptual differentiation across domains, including music (Hansen et al., 2022) and vision (Stefanics et al., 2014). When divergences along multiple dimensions of features in an auditory stimulus are relevant to perceivers, the resulting evoked responses are generated by multiple populations of neurons, producing multiple ERF components which can be summed together as a cumulative evoked response (as in Figure 2).

We follow Paavilainen et al. (2001) in using the term additivity to describe the phenomenon in which evoked responses to multi-feature deviants tend to be larger than responses to single-feature deviants. Importantly, our usage is descriptive rather than a specific measure of formal statistical additivity in the sense of no interaction (i.e. we do not evaluate superadditivity or subadditivity). Rather, we use additivity to indicate that the cumulative evoked response is larger when multiple independent features diverge from the standard, reflecting independent processing of these features in the brain. For example, Caclin et al. (2006) demonstrated that three dimensions of timbre can be independently varied above the threshold of discernability, and each dimension elicits distinct MMN responses in partially separate neuron populations. The cumulative effect of varying multiple dimensions simultaneously is an additive effect in the evoked responses themselves, demonstrating that the three dimensions of timbre are perceived in an independent fashion. Such additive effects in MMN responses have been shown in the processing of frequency and location (Schröger, 1995), pitch and location (Takegata et al., 2001), and vowels and pitch (Lidj et al., 2010).

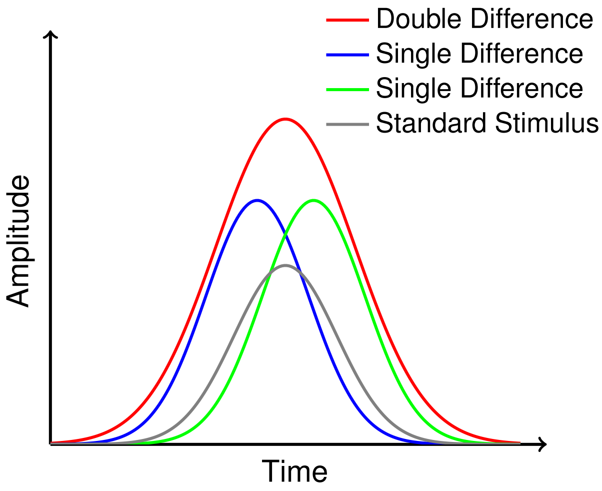


Figure 2. Schematic plot of evoked responses demonstrating neural additivity. Each waveform is an evoked response to an individual speech sound. A difference along one dimension (blue and green) produces a waveform with greater amplitude relative to the standard stimulus. A difference along two dimensions produces a waveform with an amplitude greater than either of the single-dimension deviants.

On the other hand, some aspects of the speech signal do not seem to elicit additive effects, suggesting that they are perceived in an integrated way, perhaps within a shared population of neurons rather than separate populations for each cue. Pavilainen et al. (2003) examined tone pairs in which frequency, intensity, or both could change from the first to the second tone. In each pair, the direction of change was constant: The second tone was always higher in frequency and/or louder in intensity than the first. When this direction of change was reversed, an MMN response was observed without an additive contribution from the individual frequency and intensity deviations. This pattern of non-additivity indicates the processing of frequency and intensity was integrated rather than independent. This suggests that the difference in feature changing direction was perceived as the relevant cue in the stimuli, while irrelevant differences in the physical features went unperceived. One possible interpretation of this result is that the different cues are not fully cross-classified. That is to say, since frequency and intensity co-varied in a redundant way, they were processed in an integrated manner.

This raises the possibility that phonological processing may not be sensitive to all aspects of the acoustic signal, even to cues which are significant in some languages. A necessary preliminary step towards answering this question in phonology is finding a correlate for independent processing of correlated speech sounds. Our hypothesis is that the additive property of evoked responses can be used as a reliable way of determining if elements of speech sounds are perceived in a holistic, integrated fashion or as independent percepts that are only parts of a whole (Han et al., 2023). For example, where linguistically-significant sounds are concerned, K. Yu et al. (2022) found that vowels, consonants, and tones are all processed independently by speakers of Cantonese. This means that tones and vowels are perceived as independent percepts even though they co-occur simultaneously in the speech signal. These experiments are a necessary first step towards determining if the processing of these features as independent components is a universal fact of human perception or if individual language experience plays a role in their processing as integrated or independent.

The experiments described here test the hypothesis that evoked responses can serve for a reliable correlate for the perception of linguistically significant, simultaneously occurring auditory cues in speech sounds. In particular, we are interested in whether these cues are processed independently or in an integrated manner, as reflected in the additivity of evoked-response amplitudes. Multiple cues will produce an additive effect visible in the amplitude of those evoked responses: when participants are exposed to a deviant stimulus after a series of standards, this will elicit an MMN, and that the degree of difference between standards and deviants will have a direct effect on the amplitude of evoked response. We also expect to find this response inside a window of about 100–400 ms post stimulus, the expected time-course of responses to phonological mismatches (Eulitz & Lahiri, 2004; Monahan et al., 2022; Näätänen et al., 1997; Rhodes et al., 2019; Sams et al., 1985; Scharinger et al., 2012; Sharma & Dorman, 1999). Finally, we expect the response to be most salient in the signal from channels that correspond to auditory cortex in the left-temporal region of the brain (Binder et al., 1997; Boatman et al., 1995; Crinion et al., 2003). We ran two related but different experiments to test this.

2. Experimental materials and methods

2.1. Participants

Thirty adult native English speaking participants with normal hearing and no history of auditory or neurological pathology were recruited from the University of Maryland community and paid for their time ($18/hr), with each session lasting approximately two hours. Two related protocols were designed (see Figure 3), and participants were split into two groups, with 15 participating in each protocol and no participant participating in both experiments. Exclusion criteria included the presence of ferromagnetic dental work, and a history of auditory, speech, or neurological pathology. One participant data-set was excluded from Experiment 2 due to excessive noise caused by a ferromagnetic filling, for a final count of 15 participants in Experiment 1 and 14 in Experiment 2. Participants self-reported handedness, and only right handed participants were recruited. Ethical approval for this research was granted by the University of Maryland Institutional Review Board, and all participants gave written informed consent prior to participating in the experiment session.

2.2. Design and materials

Stimulus material consisted of four syllable types (an obstruent followed by a low vowel [ɑ]: PA, BA, FA, VA) recorded by a native speaker of North American English at 44,100 Hz, presented as 16-bit WAV files. The mean stimulus duration was 712 ms (SD = 57 ms), with individual syllables being 618 ms (PA), 720 ms (FA), 743 ms (BA), and 766 ms (VA) in duration. Stimuli were normalised for intensity and presented to participants at 60 dB SPL during the experimental procedure. Experiment 1 employed a roving-standard oddball paradigm. Stimuli were presented in a single block, within which the standard syllable (BA, FA, or VA) varied randomly across trial sequences, each sequence ending with a fixed deviant (PA). Participants listened passively while neural responses were recorded (see §2.4). Experiment 2 used a single-standard oddball paradigm. The PA syllable served as the standard, and BA, FA, or VA syllables served as deviants (see §2.5).

Each experiment used a set of stimuli which can be fully cross-classified in terms of periodicity and noise, where stimuli differ from each other along one or both of these dimensions. It is this difference that we expect to elicit an MMN response in participants. All experimental stimuli were syllables which differed from each other in their initial consonant, which is the locus of acoustic differences across the experimental conditions. For example, while a PA syllable is characterised by the lack of periodicity and lack of noise in the consonantal onset, FA shares the lack of periodicity, but is characterised by noise. VA, in turn, is characterised both by periodicity and noise, meaning that while FA is different from PA along a single dimension (noise), VA is different along two simultaneously (periodicity and noise). If the MMN is additive, then we expect a larger MMN here compared to the other two cases, see Figure 1.

It is worth noting that, in the roving-standard design (§2.4), one syllable is not overrepresented in the number of stimulus presentations compared to the single-standard design (§2.5). In the latter design, the number of standard stimuli (PA) is predominant with respect to the other stimuli; for every presentation of VA, BA, or FA, participants heard approximately 18 PA syllables, while in the roving-standard paradigm, trains of stimuli alternate among the three possible comparison syllables (BA, FA, VA) as deviants, yielding a more balanced distribution of presentations across syllables. This stimuli balance is important for ensuring that observed effects are not due to a disproportionate number of standards. If there were an effect associated with the standard train preceding each deviant, this design would be more sensitive to detecting it, since each comparison stimulus occurred with equal frequency as the preceding standard.

Continuous neurophysiological activity was recorded using a 160 channel axial gradiometer whole-head magnetoencephalographic MEG system (KIT) at the University of Maryland Neuroimaging Centre. Stimuli were

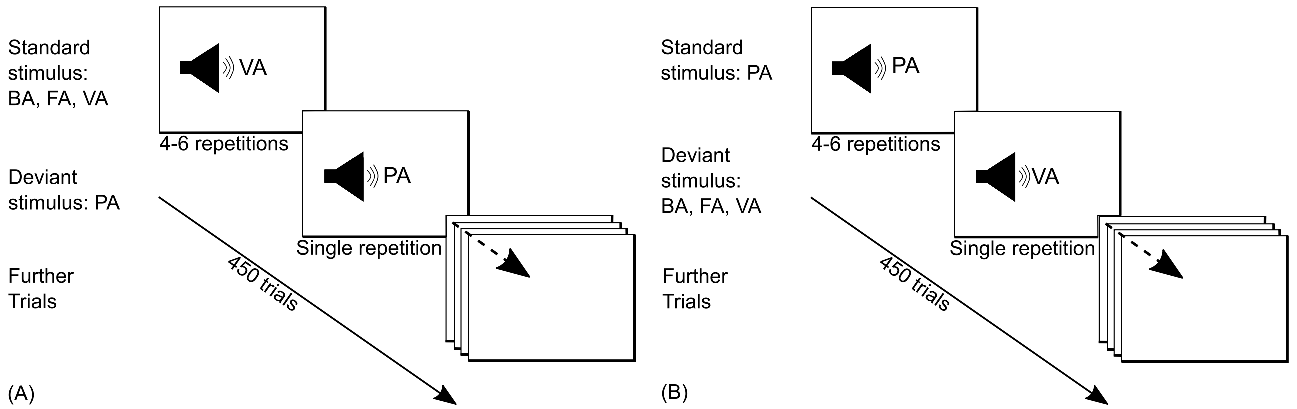


Figure 3. (A) Protocol for Experiment 1, the roving-standard paradigm. In each standard-deviant train, participants were auditorally presented with one of three repeating standard stimuli categories (e.g. [va], [ba], [fa]) followed by the deviant [pa]. The standard stimulus in each train was pseudorandomly sampled from the three categories and repeated between four and six times. (B) Protocol for Experiment 2, the fixed-standard paradigm. In each standard-deviant train, participants were auditorally presented with a single repeating standard stimulus category [pa] followed by the one of three deviant categories (i.e. [va], [ba], [fa]), and a 200 ms interval of silence. The standard stimulus in each train was repeated between four and six times and the deviant categories were pseudorandomly sampled from the three categories. The two protocols are identical in the stimuli used, they differ only in the order of presentation of standards compared to deviants. The standard [pa] differs from the deviant [va] in two features (i.e. manner, voicing), the standard [pa] differs from the deviant [ba] in one feature (i.e. voicing) and the standard [pa] differs from the deviant [fa] in one feature (i.e. manner).

delivered binaurally into the magnetically shielded room via Etymotic ER3A insert earphones that were calibrated and equalised to have a flat frequency response between 100 and 5000 Hz. Continuous recordings of cortical activity were made in DC (no high pass filter) at a sampling frequency of 1000 Hz. An online low pass filter of 200 Hz and a 60 Hz notch filter were applied during recording. For the duration of the experiment, participants laid in a supine position and quietly watched a silent movie in a dimly lit, magnetically shielded room (Tervaniemi et al., 1999).

2.3. MEG data preprocessing

The data was analysed using MNE-Python (Gramfort et al., 2013). Flat and noisy channels were identified for each data-set and interpolated using the MNE algorithm. A band-pass filter was applied at .01 and 30 Hz to the raw MEG data. Using independent component analysis (ICA; Picard algorithm) implemented in MNE-Python, fifteen components were computed for each participant, and those reflecting ocular (blinks and saccades) or cardiac artifacts were identified by visual inspection of the component topographies and time courses were subsequently removed. Across participants in both experiments, the number of components removed ranged from 3 to 6 (mean = 4.71, SD = 0.91). After preprocessing, epochs of 900 ms were extracted from the continuous data (−100 to 800 ms relative to stimulus onset). Epochs were baseline-corrected by subtracting the mean signal in the −100 to 0 ms pre-stimulus interval from each channel prior to statistical analysis. Epoch rejection was set to 3000 fT to exclude trials with abnormal peak-to-peak amplitudes caused by non-cortical signal, such as muscle movement or coughs, resulting in .92% of all epochs being dropped.

2.4. Experiment 1

2.4.1. Procedure

Despite the prominent role that the MMN has played in understanding the neurophysiology of our sensory systems, its exact mechanism is not completely understood (Garrido et al., 2009; May & Tiitinen, 2004, 2010); however it can be elicited after only a few repetitions of a standard stimulus (Garrido et al., 2008; Haenschel et al., 2005; Jääskeläinen et al., 2004; Näätänen et al., 2007). In this novel multi-deviant paradigm, participants were exposed to a fully random sequence of 4–6 standard stimuli selected from the set standards – BA, FA, and VA. Stimuli were deployed using PsychoPy (Peirce et al., 2019). The inter-stimulus interval (ISI) was 350–1000 ms. The ISI was randomly determined for each standard train and following deviant stimulus. After this train, a single deviant stimulus, PA, was presented. Each participant was exposed to 450 trials (Figure 3(a)). Assuming a mean of 5 repetitions per standard, participants were exposed to 2,250 standard and 450 deviant presentations (total 2,700 stimuli).[1] Participants listened passively while neural responses were recorded.

In each trial, one standard syllable (randomly selected from BA, FA, or VA) was repeated 4–6 times and was then followed by a single deviant (PA). Each participant completed 450 trial sequences in total. Using MNE-Python, PA events were sorted into categories according to which standard they followed (BA, FA, or VA). After excluding artifact-contaminated trials, we focussed on the subset of transitions in which the deviant followed six standards, yielding an average of 150 usable transitions per participant, or approximately 2,250 usable evoked responses across all participants.

2.4.2. Data analysis

Regions of interest were determined by inspection of butterfly plots of every channel at time intervals that fall within the expected time-course of MMN (see Figure 4). Sensor-level activity shows the overall spatiotemporal distribution of evoked responses, with the most active sensors are localised to the left and right posterior temporal channels, with possible peaks of interest at about 100, 200, and 350 ms, within the range of the expected time-course of an MMN response.

A topographic plot of the time windows confirms the presence of a dipole in each condition (Figure 5). Following this, we limited our data analysis to a set of 10 channels in the posterior temporal region of the left hemisphere, which showed the largest evoked responses over the expected time course, as shown in the schematic sensor array in Figure 6.

Plotting the ERFs by condition at these channels revealed peaks at 100 ms, at 200 ms, and at 350 ms (Figure 6). At each peak, the VA condition differs most from the standard. The plotted evoked responses in Figure 6 suggest that there is a response beginning just before 150 ms, with an approximately 100 ms duration, within the expected MMN response time window.

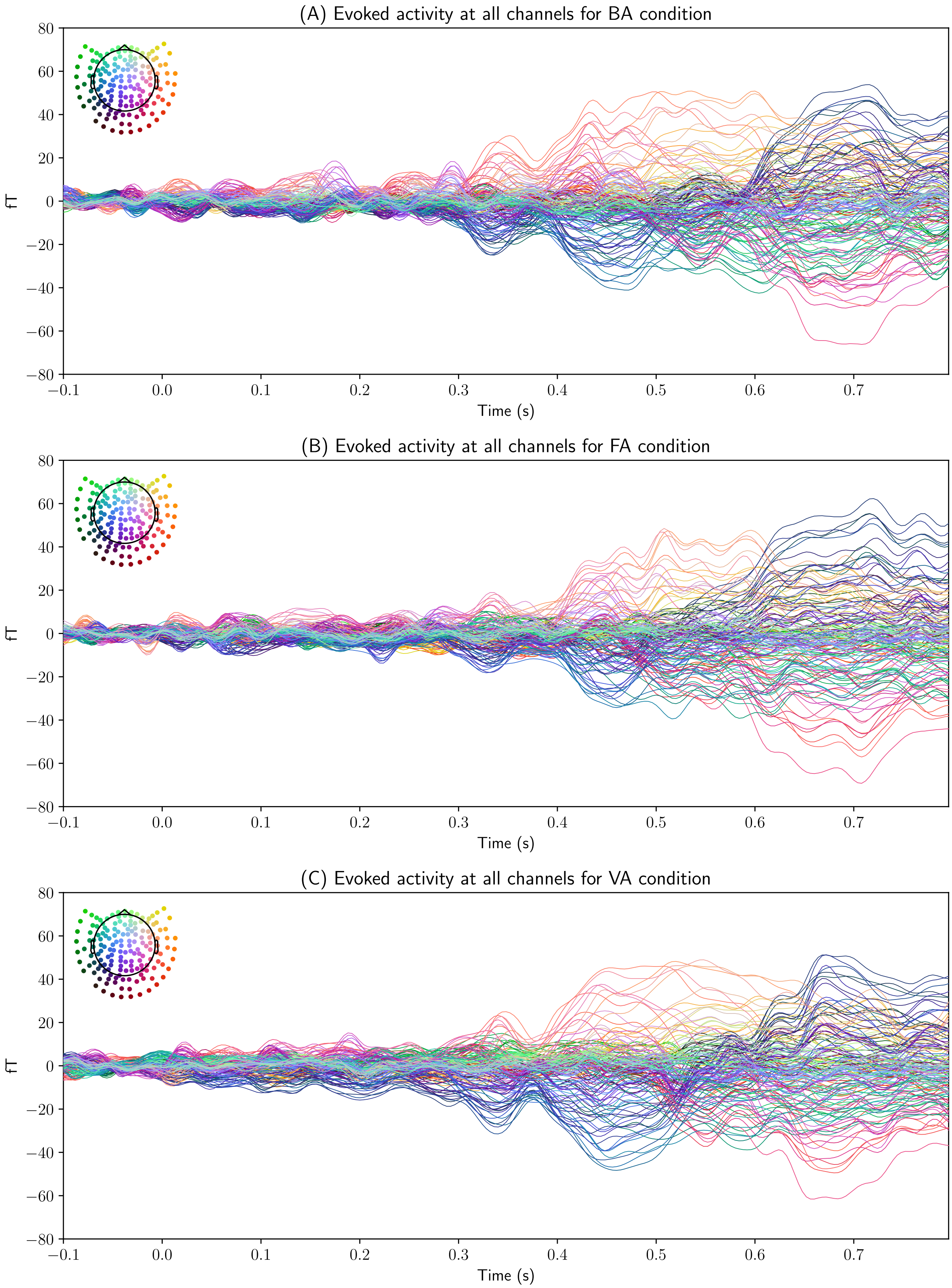
2.4.3. Results

The ERF results are consistent with our hypothesis: The VA condition is the "most" divergent, while the BA and FA conditions elicit similar evoked responses (see Figure 6). To quantify these differences, we fit a linear mixed-effects model (LMM) using the `statsmodels` package in Python (Seabold & Perktold, 2010). The dependent variable was the mean amplitude within each time window, and Condition (BA, FA, VA) was included as a fixed effect. Participant was included as a random intercept to account for subject-specific baseline differences. The overall temporal window of analysis (0–400 ms post-stimulus onset) was selected *a priori* based on the typical time course of the mismatch negativity (MMN). Time intervals divide the approximately 400 ms window in which the MMN is known to occur: 0–80 ms, 80–160 ms, 160–240 ms, 240–320 ms, and 320–400 ms (see Garrido et al., 2009 for discussion concerning the time course of MMN).

Channel selection included the 10 active channels in the posterior temporal region as described in §2.4.2, as well as selections of ten channels in three other quadrants: the posterior temporal region of the right hemisphere, and bilateral frontal regions. Condition, the primary fixed effect, reflected the standard-to-deviant transition: PA following BA, FA, or VA. Categorical predictors (Condition, TimeInterval, and ChannelSelection) were dummy-coded, with the first level of each variable serving as the reference category. Participant was included as a random intercept to account for variability in baseline mean amplitude across participants. The main effect of Condition tested whether mean amplitude differed across the three deviant response types, averaged across time intervals and channel selections.

The results of the model indicated a significant difference between conditions, with the reference condition being PA in the context of BA standards. The predicted mean amplitude for PA after BA was 2.798 fT (95% CI [1.082, 4.514]) fT. In contrast, PA in the context of the FA condition did not differ significantly from that of the BA condition ($\beta = -.846$, SE = .959, $z = -.882$, $p = .378$), with a predicted mean amplitude of 1.952 fT (95% CI [.236, 3.668] fT). PA in the context of the VA condition, however, showed a larger response ($\beta = 2.502$, SE = .959, $z = 2.608$, $p = .009$), with a predicted mean amplitude of 5.3 fT (95% CI [3.584, 7.01] fT). Because the same physical deviant (PA) was presented across all standard contexts, this difference can be attributed to the preceding standard rather than to the physical properties of the deviant itself. This effect was particularly prominent in the left temporal channels during the 160–240 ms time window, corresponding to the expected MMN response.

To further examine the differences between conditions within the 160–240 ms time interval at posterior left temporal channels, we performed a post-hoc Tukey HSD test using the `statsmodels` package (Seabold & Perktold, 2010) in Python. Tukey HSD tests were conducted across all time interval and channel selection combinations; only the 160–240 ms interval at posterior left temporal channels yielded pairwise differences. Further, the results indicated a difference between PA in the context of BA versus VA ($M = 15.36$, $p = .041$), as well as between the FA and VA contexts ($M = 17.82$, $p = .015$). These results support our hypothesis in that there was no difference between PA conditions in the BA and FA contexts ($M = -2.46$, $p = .915$), suggesting that while the BA and FA contexts did not differ significantly from each other, the VA context showed a stronger response (results summarised in Table 3).


(A) Evoked activity at all channels for BA condition

(B) Evoked activity at all channels for FA condition

(C) Evoked activity at all channels for VA condition

Figure 4. Experiment 1 (roving standards paradigm) sensor-level evoked response fields across all 156 channels for the deviant [pa] following each standard category: (A) standard [ba], deviant [pa], (B) standard [fa], deviant [pa], (C) standard [va], deviant [pa]. Channels are colour coded, and a head map illustrating sensor distribution is provided in the upper left corner of each panel. Data is from 15 human participants.

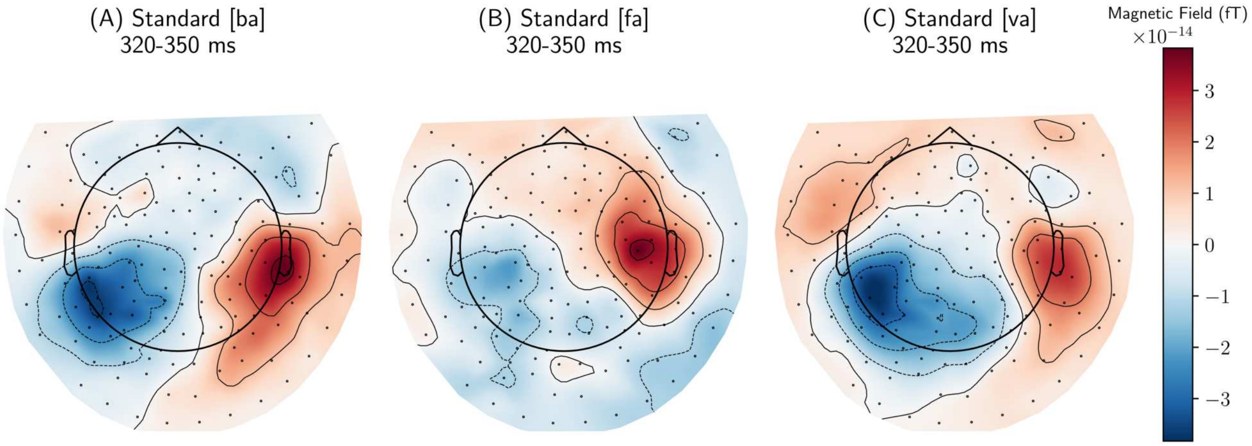


Figure 5. Experiment 1 (roving standards paradigm) sensor-level topography of the sink and source for the evoked response fields averaged over the 320–350 ms post-stimulus onset time-window to the (A) standard [ba], deviant [pa], (B) standard [fa], deviant [pa], (C) standard [va], deviant [pa] stimulus trains. Topographic regions in red indicate the source of the magnetic field (i.e. positive evoked response fields), while topographic regions in blue indicate the sink of the magnetic field (i.e. negative evoked response fields).

To complement the inferential statistics, we calculated Cohen's *d* for each pairwise comparison using the `pingouin` package in Python (Vallat, 2018). The largest effects were observed between the BA and VA contexts ($d = -.884$) and the FA and VA contexts ($d = -1.105$), indicating large differences, with VA

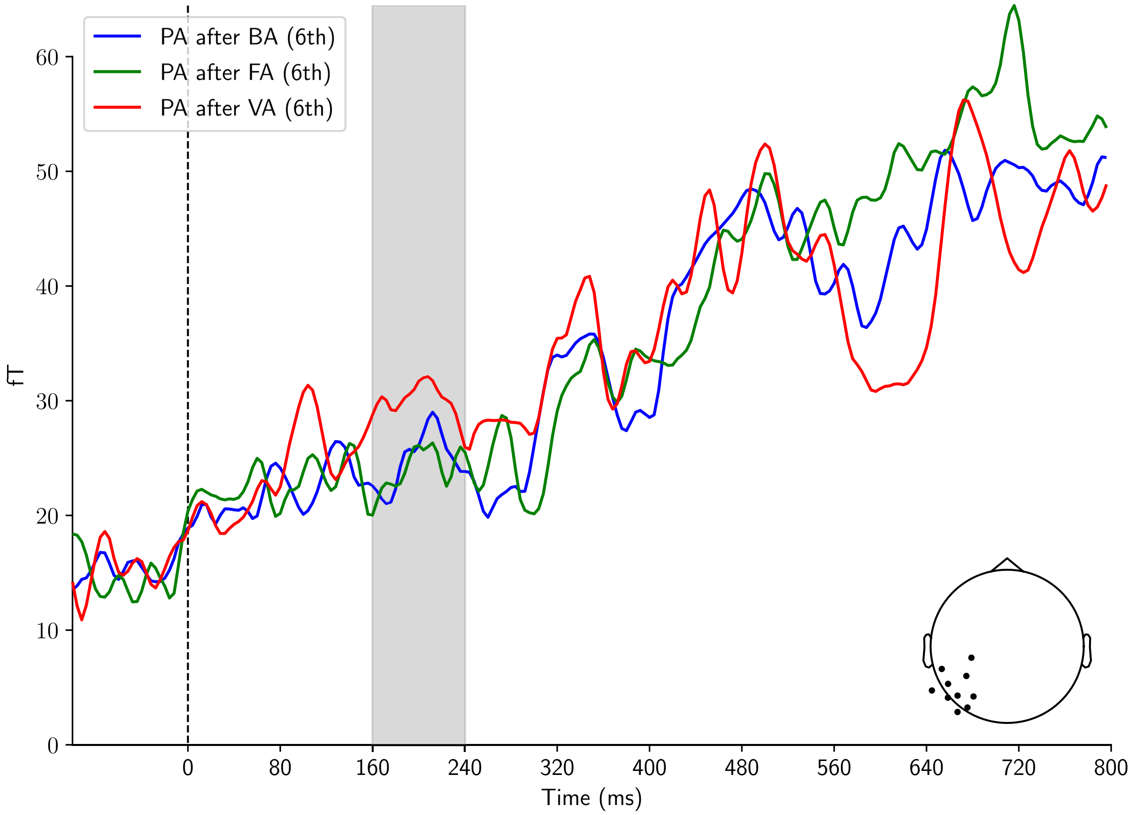


Figure 6. Root-mean square (RMS) of the evoked response fields in Experiment 1 (roving standards paradigm) to the deviant [pa] following one of the three possible standard stimuli (i.e. [ba] (blue), [fa] (green), [va] (red)). The RMS responses here are obtained from the mean of 10 left temporal sensors (see head map with channel selection in the lower right portion of the Figure) and are only for stimulus trains that contained six standard stimuli preceding the deviant. The dotted line at 0 ms indicates onset of the auditory stimulus. Time ticks on the x-axis are spaced at 80 ms intervals; the first five intervals were included in the statistical analysis. A significant peak of interest is highlighted in gray.

Table 3. Experiment 1: Post-hoc Tukey HSD results for 160–240 ms time interval and left-temporal channels.

Group 1	Group 2	Mean Diff.	p-adj	Lower	Upper
PA after 6 BAs	PA after 6 FAs	−2.46	.915	−17.30	12.39
PA after 6 BAs	PA after 6 VAs	15.36	.041	.51	30.21
PA after 6 FAs	PA after 6 VAs	17.82	.015	2.97	32.66

eliciting stronger responses in each comparison. In contrast, the effect size between the BA and FA contexts was negligible ($d = .147$), suggesting similar responses for those two conditions. These effect sizes further support the interpretation that the VA condition differed most strongly from both BA and FA, which themselves did not differ meaningfully.

These results support the additive hypothesis: the evoked responses to the deviant stimulus were different across all three standard conditions, but the response to the deviant following a standard from which it differed along two dimensions was larger. This contributes further evidence that speakers attend to contrastive speech cues in an independent manner.

2.5. Experiment 2

2.5.1. Procedure

Experiment 2 differs from Experiment 1 in that the standard and deviant conditions are reversed. Experiment 1 was a roving-standard paradigm, but Experiment 2 uses a fixed-standard paradigm: A PA context to evoke responses to a roving deviant FA, BA, or VA. Stimuli were presented using PsychoPy (Peirce et al., 2019). In each stimulus train, participants heard between four and six repetitions of the standard PA stimulus, each separated by a randomly sampled inter-stimulus interval (ISI) between 350 and 1000 ms. After the final PA, a single deviant stimulus from the set BA, FA, or VA was presented following another independently sampled delay within the same ISI range.

Each participant completed 450 trials (Figure 3(b)), producing approximately 2,100 evoked responses for each deviant condition. Since there are more standard presentations in the data (between four and six standards for every deviant in each trial), a random sample of PA stimuli was taken on a by-participant basis, equivalent in number to the average of deviant stimuli, selecting the fourth PA in the sequence as the “most standard” stimulus that a participant could expect on any trial.

2.5.2. Data analysis

Regions of interest were determined by inspection of butterfly plots of every channel used at time intervals which fall within the expected time-course of MMN (see Figure 7). Visual inspection of the channel activity shows that the most active channels are localised to the left and right posterior temporal channels, with possible peaks of interest at about 150, 250, and 325 ms, within the range of the expected time-course of an MMN response.

A topographic plot of the time windows confirms the presence of a dipole in each condition (Figure 8). Following this, we limited our data analysis which a set of 10 channels in the posterior temporal region of the left hemisphere which showed the strongest response over the expected time course. These channels are shown in the schematic sensor array in Figure 9.

A grand average was obtained across all participant data sets to yield the plot in Figure 9.

2.5.3. Results

Using the `statsmodels` package in Python (Seabold & Perktold, 2010), we ran a linear mixed-effects model (LMM) to examine the effects of Condition (BA, FA, VA), TimeInterval (0–80 ms, 80–160 ms, 160–240 ms, 240–320 ms, and 320–400 ms), and ChannelSelection (Left Temporal, Left Temporal Front, Right Temporal, Right Temporal Front) on the mean amplitude of evoked responses in Experiment 2. All categorical predictors were dummy-coded, with the first level of each variable serving as the reference category. Participant was included as a random intercept to account for baseline variability across subjects. The model revealed no significant three-way interaction, and post-hoc Tukey HSD tests likewise revealed no significant differences across deviant conditions.

One potential reason for the lack of difference in these results might be the different temporal aspects of each condition. For example, the PA stimulus is characterised by a period of closure during the articulation of the consonant, which is then followed by a burst and a period of aspiration before the vowel begins, for a total obstruent duration of 77 ms. The BA stimulus also begins with an obstruction characterised by a period of closure and a burst, but there is no period of aspiration and the total duration of the consonant is only 38 ms. The two fricative conditions, FA and VA, have no closure period and no burst, and have durations of 118 ms and 75 ms respectively. A further difficulty, beyond the question of duration, is that, since BA and PA are characterised by periods of articulatory closure and acoustic silence, while FA and VA are not, is that it is not clear when participants perceive the stimuli as distinctive speech sounds. These facts compound such that the stimuli being compared are in fact heterogeneous in a way that could impact their perception: the obstruents do not temporally align with respect to when the

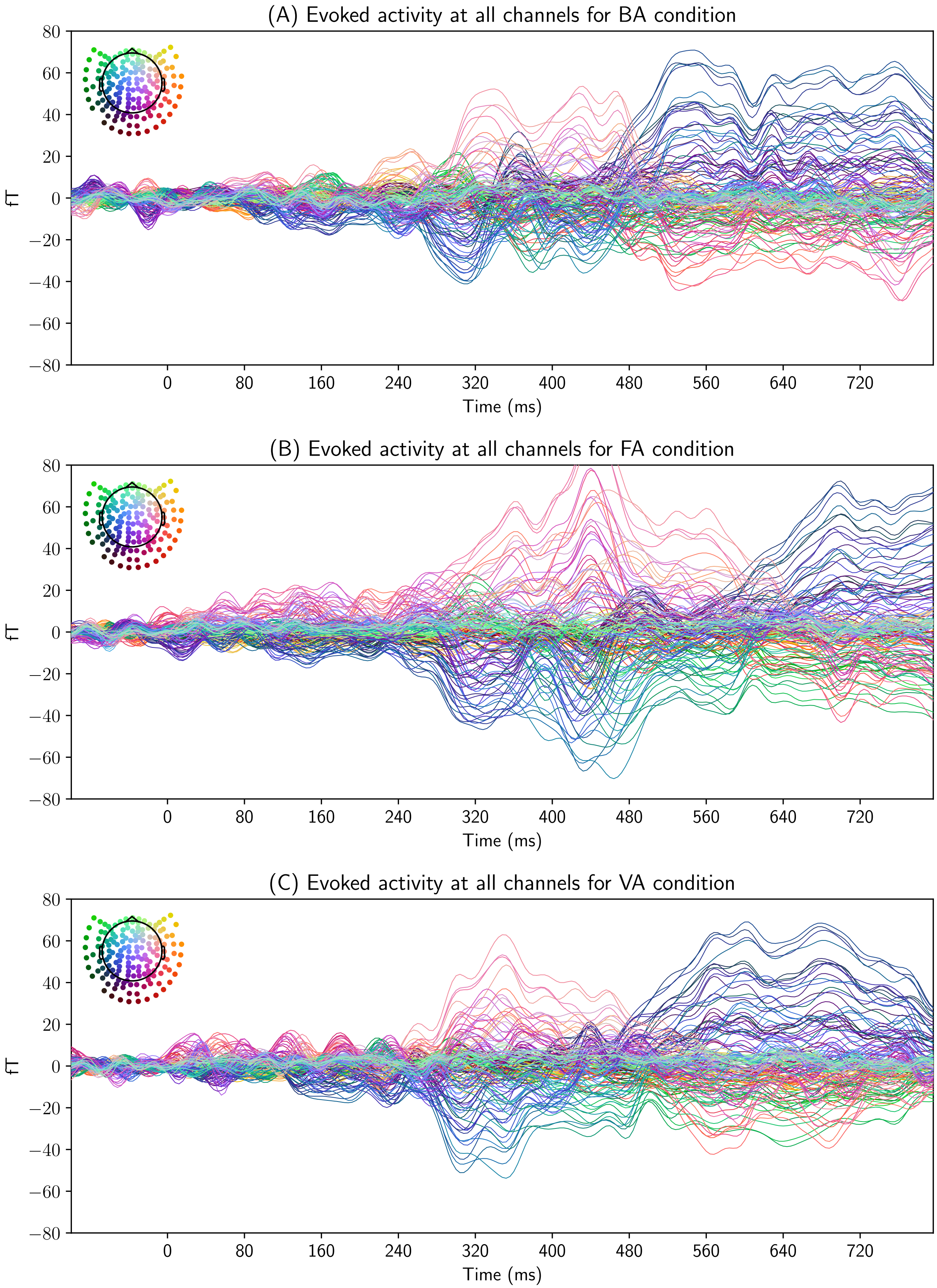


Figure 7. Experiment 2 (roving deviants paradigm) sensor-level evoked response fields across all 156 channels for the deviants [ba], [fa], and [va] following the standard [pa]: (A) standard [pa], deviant [ba], (B) standard [pa], deviant [fa], (C) standard [pa], deviant [va]. Channels are colour coded, and a head map illustrating sensor distribution is provided in the upper left corner of each panel. Data is from 14 human participants.

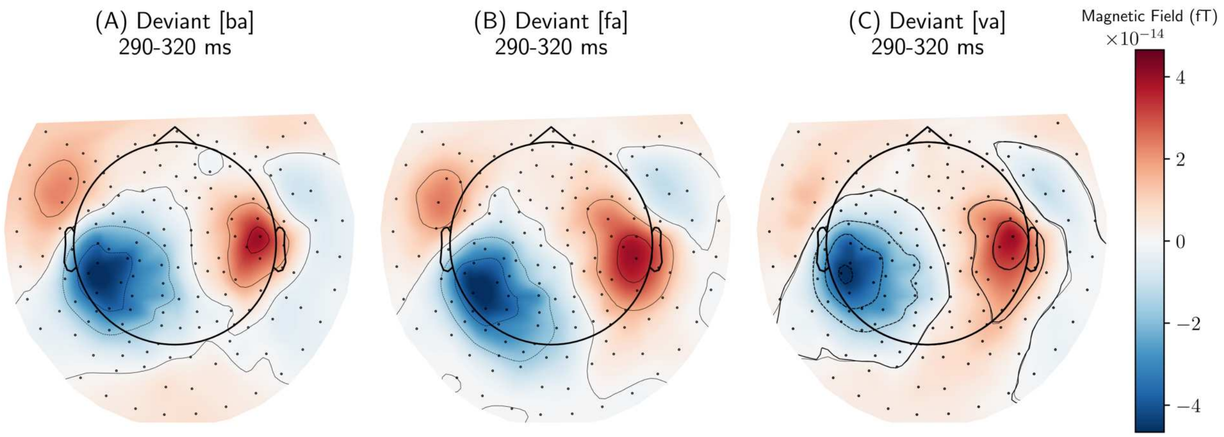


Figure 8. Experiment 2 (roving deviants paradigm) sensor-level topography of the sink and source for the evoked response fields averaged over the 290–320 ms post-stimulus onset time-window to the (A) standard [pa], deviant [ba], (B) standard [pa], deviant [fa], (C) standard [pa], deviant [va] stimulus trains. Topographic regions in red indicate the source of the magnetic field (i.e. positive evoked response fields), while topographic regions in blue indicate the sink of the magnetic field (i.e. negative evoked response fields).

acoustic-phonetic cues to voicing and manner of articulation are available to listeners. The roving-standard design in Experiment 1 obviates this issue, because every evoked-response is a reaction to the same physical token.

3. Discussion

The experiments presented here constitute an effort to link cognitive neuroscience and phonological theory (Embick & Poeppel, 2015; Poeppel & Embick, 2005). In

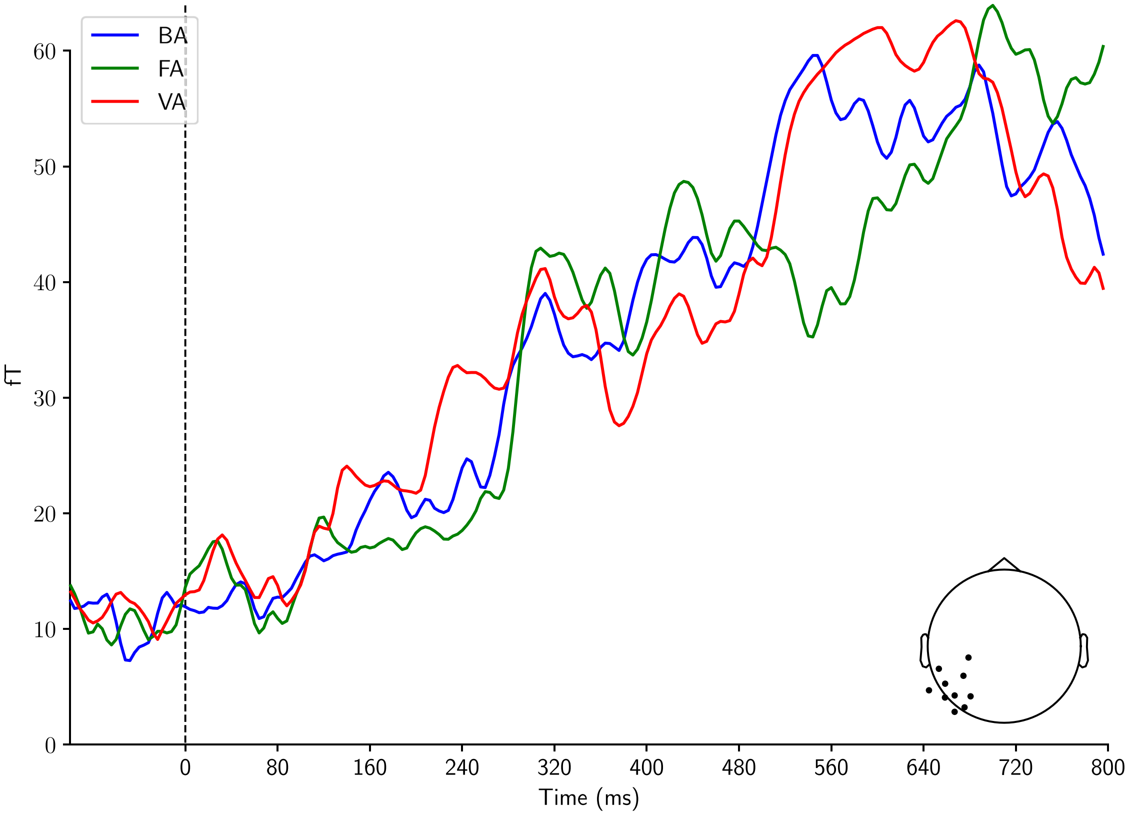


Figure 9. Root-mean square (RMS) of the evoked response fields in Experiment 2 (roving deviants paradigm) to one of the three possible deviant stimuli (i.e. [ba] (blue), [fa] (green), [va] (red)) following the standard [pa]. The RMS responses here are obtained from the mean of 10 left temporal sensors (see head map with channel selection in the lower right portion of the Figure) and are only for stimulus trains that contained six standard stimuli preceding the deviant. The dotted line at 0 ms indicates onset of the auditory stimulus. Time ticks on the x-axis are spaced at 80 ms intervals; the first five intervals were included in the statistical analysis.

phonology, there is a virtually consensual view which holds that the representations of speech sounds are complexes of features, basic representational units correlated to acoustic phenomena and/or articulatory configurations relevant to the act of speaking (see the contributions in Clements & Ridouane, 2011). However, there is extensive debate in the literature concerning the exact nature of these representations; in particular there has been recent debate concerning the origin of phonologically-relevant features. Early work in contemporary phonology assumed that features are universal, with a single, genetically-endowed set of features shared by all speakers of all languages (Chomsky & Halle, 1968; Jakobson et al., 1952). This assumption is still axiomatic in contemporary work in phonology, though recent work has questioned it, arguing for a theory of features that emerge on a language specific basis (Dresher, 2014; Mielke, 2008; Odden, 2022).

We used English to establish a full cross-classification of two features, which are contrastively meaningful in phonology (voice, manner), correlated to two phonetic features in the acoustic signal (lower-frequency periodic noise, higher frequency aperiodic noise). The larger MMN to PA after the standard VA suggests that we can use the novel roving-standards MMN paradigm to find neurophysiological changes to individual feature changes. Our results suggest that evoked additivity is a reliable correlate for independent processing of linguistically-significant acoustic features.

The present study is of both methodological and theoretical interest. Methodologically, it asks if techniques from cognitive neuroscience can to used to derive a reliable correlate for features – the most basic unit of phonological representation – by examining additivity in evoked responses to co-occurring acoustic cues in speech. It lays the groundwork for investigating whether phonological features are universal or emerge during acquisition from a more general capacity for categorising linguistic sound. While previous studies have explored additivity in correlated speech cues (Lidj et al., 2010; K. Yu et al., 2022), we extend this work by asking how language-specific experience shapes the processing of acoustically comparable, but linguistically variable, cues. Put another way, we ask if linguistic experience changes perception in speakers' brains.

Our linking hypothesis is additive: we predict that evoked responses to a speech sound which differed relative to another along two dimensions would be additively larger than those which differed to the standard along only a single dimension. Further, we predicted that those evoked responses to stimuli which differed along only one axis would be largely comparable to each other. We interpret additivity as a measure of independent processing of acoustic features, suggesting that a larger MMN response from a deviant to a standard correlates to more feature changes, broadly speaking.

This hypothesis is supported by the results of Experiment 1, which used a novel roving-standard task to explore the MMN to the same speech sound in different contexts. The results from Experiment 1 show that a change from VA to PA elicits a larger MMN than a change from BA or FA to PA, the latter of which are equivalent in terms of featural differences. These results contribute to the literature on additivity, showing that acoustic features are processed in an additive manner when they are linguistically significant. Moreover, this study makes a methodological contribution by developing a novel roving-standard paradigm which can be used on any four-way cross-classification system.

One possible interpretation of these results is as an asymmetric MMN, as reported in oddball paradigms that probe phonetic and phonological representations (Monahan, 2018). In traditional block designs, where each phonetic category alternates as standard and deviant, the resulting MMNs are typically not of equal amplitude, contrary to predictions based solely on acoustic-phonetic properties. Instead, studies consistently report larger MMN amplitudes when the standard stimulus corresponds to a putatively specified or marked category, and the deviant corresponds to an underspecified or unmarked category. For example, an MMN was elicited when German round and back vowels – specified with [labial] and [dorsal], respectively – served as the standard, and an underspecified coronal vowel served as the deviant; in contrast, no MMN was observed when the unmarked coronal vowel was the standard and the marked labial or dorsal vowel was the deviant (Eulitz & Lahiri, 2004).

Collectively, these findings suggest that voiced and voiceless stops are not representationally equivalent in the phonology: voiced stops are underspecified relative to voiceless stops, producing larger MMN responses to voiced deviants among voiceless standards, though the reverse does not hold. This is consistent with predictions of the Featurally Underspecified Lexicon framework (Lahiri & Reetz, 2002, 2010). Similar results have been observed in consonants (Cornell et al., 2013; Fu & Monahan, 2021; Hestvik & Durvasala, 2016; Hestvik et al., 2020; Højlund et al., 2019; Maiste et al., 1995; Monahan et al., 2022; Politzer-Ahles & Jap, 2024; Schluter et al., 2016, 2017), vowels (Cornell et al., 2011; de Rue et al., 2021; Scharinger et al., 2012, 2016; Y. H. Yu & Shafer, 2021) and lexical tone (Politzer-Ahles et al., 2016). Though consistent with privative feature systems, this research does not adjudicate hypotheses that regard the universality of the feature set. Further, Hestvik and Durvasala (2016) demonstrated that asymmetric mismatch responses emerge only when there is intra-category variation among the standard stimuli. Because the present study used a single, invariant standard token for each condition, no such asymmetry was expected; the current design thus minimises potential asymmetry effects.

A potential limitation of the present study is that we did not include a separate P-standard condition, which would have allowed for a more direct isolation of the MMN component. Adding such a condition would have substantially increased the duration of the experiment and risked introducing fatigue-related noise into the data. Nevertheless, the differential responses observed across standard contexts, together with the consistent left-temporal localisation of effects, support the interpretation that the observed responses index mismatch processing. While direct isolation of the MMN would be ideal, we believe that the current design provides a reliable measure of auditory feature processing in this paradigm.

Experiment 2 used a classic roving deviant design to elicit an MMN; unlike Experiment 1, the design of this paradigm presents participants with a fixed standard and roving deviant. What is compared are cross-category stimuli, each with their own temporal, articulatory, and acoustic differences. The results obtained are suggestive but difficult to interpret. The evoked responses for each condition cluster together but do exhibit apparent divergences, though our statistical analysis did not reveal any differences. One potential reason is the difficulty inherent in properly aligning temporally heterogeneous speech sounds. Another possibility is that the sample size does not provide adequate statistical power in this paradigm.

One remaining question concerns to which domain the additive effect in Experiment 1 is relevant. In other words, is the additive effect generated by auditory processing of phonetic features, or by a more abstract phonology? English is the empirical domain of our study, but this work is only a first step investigating a fundamental question about representations in human minds, not just individual languages. As such, it is important to extend the paradigm to a language where the phonetics and phonology diverge, to disentangle the two domains.

Speakers are sensitive to the phonological contrasts of their own language, but what about speech sounds with non-contrastive acoustic cues? As shown in Table 2, some languages have gaps or allophony resulting in a system of contrasts that does not fully use every feature available in the acoustic signal. The protocol developed in Experiment 1 could be appropriately adapted to test speakers of languages where the phonetics do not align with the phonology as they do in English. This has the potential to contribute to phonological theory, specifically with respect to the language-specific view of phonologically significant features and integrated processing of correlated speech cues.

Evidence for the emergent theory of language-specific featural representations would take the form of an absent or suppressed MMN to a speech sound that is phonetically, but not phonologically, distinct. For example, in Campidanese Sardinian, stops use voicing as a contrastive cue, but the distribution of voiced fricatives is restricted (Bolognesi, 1998; Lai, 2021; Virdis, 1978). The bilabial fricative appears as a predictable allophone of [p] in intervocalic contexts. Acoustically, it is distinct from [p], [b], and [f] in the same way as English, but phonologically it may lack a [voice] feature, since voicing has an incomplete contrastive function. An MMN response to this allophone that does not differ significantly from [p], relative to [b] and [f], would indicate that acoustic correlates for voice are not represented phonologically, and its acoustic features are perceived in an integrated manner – this would provide support to the language-specific view of emergent feature systems.

In conclusion, our study tested whether or not contrastive phonetic cues are independently processed. We designed two oddball paradigms – one of which is novel. We found MMN responses in this paradigm, consistent with independent processing of the acoustic cues. This work is a preliminary step towards further testing of speakers of a language with incomplete contrastive cross-classification of acoustic cues which could bring further theoretical insight.

Note

1. One participant session was terminated early due to an unexpected time constraint and yielded only 343 trials.

Acknowledgments

Noora Khattab was a great help in recruiting participants and providing technical assistance in the MEG lab. Ciaran Stone was an invaluable trouble-shooter during experimental sessions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Allen, E. J., P. C. Burton, Olman, C. A., & Oxenham, A. J. (2017). Representations of pitch and timbre variaiton in human auditory cortex. *The Journal of Neuroscience*, *37*(5), 1284–1293. https://doi.org/10.1523/JNEUROSCI.2336-16.2016

Allen, E. J., Mesik, J., Kay, K. N., & Oxenham, A. J. (2022). Distinct representations of tonotopy and pitch in human auditory cortex. *The Journal of Neuroscience*, *42*(3), 416–434. https://doi.org/10.1523/JNEUROSCI.0960-21.2021

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience*, *17*(1), 353–362. https://doi.org/10.1523/JNEUROSCI.17-01-00353.1997

Boatman, D., Lesser, R. P., & Gordon, B. (1995). Auditory speech processing in the left temporal lobe: An electrical interference study. *Brain and Language*, *51*(2), 269–290. https://doi.org/10.1006/brln.1995.1061

Bolognesi, R. (1998). *The phonology of campidanian sardinian*. HIL.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., & McAdams, S. (2006). Seperate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*(12), 1959–1972. https://doi.org/10.1162/jocn.2006.18.12.1959

Chabot, A. (2023). Prosodic strength in campidanese sardinian as substance-free phonology. *Phonology*, *40*(3–4), 197–228. https://doi.org/10.1017/S0952675724000137

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Clements, G. N., & Ridouane, R. (Eds.). (2011). *Where do phonological features comes from? Cognitive, physical and developmental bases of distinctive speech categories*. John Benjamins Publishing Company.

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

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

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*(5), 1193–1201. https://doi.org/10.1093/brain/awg104

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

Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. Cambridge University Press.

Dresher, B. E. (2014). The arch not the stones: Universal feature theory without universal features. *Nordlyd*, *41*(2), 165–181. https://doi.org/10.7557/12.3412

Embick, D., & Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: *Correlational, integrated* and *explanatory* neurolinguistics. *Language, Cognition and Neuroscience*, *30*(4), 357–366. https://doi.org/10.1080/23273798.2014.980750

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583. https://doi.org/10.1162/089892904323057308

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.

Gallistel, C. (1990). Representations in animal cognition. *Cognition*, *37*(1–2), 1–22. https://doi.org/10.1016/0010-0277(90)90016-D

Garner, W., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225–241. https://doi.org/10.1016/0010-0285(70)90016-2

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomoy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*(2), 133–143. https://doi.org/10.1162/jocn.1995.7.2.133

Gomes, H., Ritter, W., & Vaighan Jr, H. G. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neurocience*, *7*(1), 81–94.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*, *7*, 1–13. https://doi.org/10.3389/fnins.2013.00267

Grimaldi, M. (2018). The phonetics-phonology relationship in the neurobiology of language. In R. Petrosino, P. Cerrone & H. van der Hulst (Eds.), *From sounds to structures: Beyond the veil of Maya* (pp. 66–104). De Gruyter Mouton.

Haenschel, C., Vernon, D. J., Dwividi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, *6*(2), 95–102. https://doi.org/10.1016/S0095-4470(19)31101-5

Han, Z., Zhu, H., Shen, Y., & Tian, X. (2023). Segregation and integration of sensory features by flexible temporal characteristics of independent neural representations. *Cerebral Cortex*, *33*(16), 9542–9553. https://doi.org/10.1093/cercor/bhad225

Hansen, N. C., Højlund, A., Møller, C., Pearce, M., & Vuust, P. (2022). Musicians show more integrated neural processing of contextually relevant acoustic features. *Frontiers in Neuroscience*, *18*, 1–18.

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

Hestvik, A., Shinohara, Y., Durvasala, K., Verdonschot, R. G., & Sakai, H. (2020). Abstractness of human speech sound representations. *Brain Research*, *1732*, 1–14. https://doi.org/10.1016/j.brainres.2020.146664

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

Jääskeläinen, I., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Fa-Hsuan, L., May, P., Melcher, J., Shufflebeam, S., Tiitinen, H., & Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sceinces*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

Jakobson, R. (1939). Observations sur le classement phonologique des consonnes. In E. Blancquaert & W Pée (Eds.), *Proceedings of the 3rd International Congress of Phonetic Sciences* (pp. 273–279). Laboratory of Phonetics at the University of Ghent.

Jakobson, R., Fant, C. G. M., & Halle, M. (1952). *Preliminaries to speech anaysis: The distinctive features and their correlates*. MIT Press.

Janssen, N., van der Meij, M., López-Pérez, P. J., & Barber, H. A. (2020). Exploring the temporal dynamics of speech production with EEG and group ICA. *Scientific Reports*, *10*, 1–14.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 637–675). Mouton de Gruyter.

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

Lai, R. (2021). Sardinian. In C. Gabriel, R. Gess & T. Meisenburg (Eds.), *Manual of Romance phonetics and phonology* (pp. 597–627). De Gruyter Mouton.

Lerousseau, J. P., Parise, C. V., Ernst, M. O., & van Wassenhove, V. (2010). Multisensory correlation computations in the human brain identified by a time-resolved encoding model. *Nature Communications*, *13*, 1–12.

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

Lisker, L. (1986). Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 3–11. https://doi.org/10.1177/002383098602900102

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

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*(1), 68–89. https://doi.org/10.1097/00003446-199502000-00006

May, P. J., & Tiitinen, H. (2004). The MMN is a derivative of the auditory N100 response. *Neurology and Clinical Neurophysiology*, *20*, 1–5.

May, P. J., & Tiitinen, H. (2010). Mismatch negativity MMN, the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–112. https://doi.org/10.1111/psyp.2010.47.issue-1

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science (New York, N.Y.)*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Mielke, J. (2008). *The emergence of distinctive features*. Oxford University Press.

Monahan, P. J. (2018). Phonological knowledge and speech-comprehension. *Annual Review of Linguistics*, *4*(1), 21–47. https://doi.org/10.1146/linguistics.2018.4.issue-1

Monahan, P. J., Schertz, J., Fu, Z., & Pérez, A. (2022). Unified coding of spectral and temporal phonetic cues: Electrophysiological evidence for abstract phonological features. *Journal of Cognitive Neuroscience*, *34*(4), 618–638. https://doi.org/10.1162/jocn_a_01817

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/psyp.2005.42.issue-1

Näätänen, R., Kujala, T., & Light, G. (2019). *The mismatch negativity (MMN): A window to the brain*. Oxford University Press.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *285*(6615), 432–434. https://doi.org/10.1038/385432a0

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*, 544–2590.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post-syllable onset. *NeuroImage*, *20*(3), 1839–1847. https://doi.org/10.1016/j.neuroimage.2003.07.019

Odden, D. (2022). Radical substance-free phonology and feature learning. *The Canadian Journal of Linguistics*, *67*(4), 500–551. https://doi.org/10.1017/cnj.2022.10

Paavilainen, P., Valppu, S., & Näätänen, R. (2001). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: $1+1\asymp 2$ but $1+1+1 < 3$. *Neuroscience Letters*, *301*(3), 179–182. https://doi.org/10.1016/S0304-3940(01)01635-4

Parise, C., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*, 11543. https://doi.org/10.1038/ncomms11543

Pavilainen, P., Mikkonen, M., Kilpeläinen, M., Lehtinen, R., Saarela, M., & Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: A human auditory event-related potential study. *Neuroscience Letters*, *349*(2), 79–82. https://doi.org/10.1016/S0304-3940(03)00787-0

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Hochenberge, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poeppel, D., & Embick, D. (2005). The relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–120). Lawrence Erlbaum Associates.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2007). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transaction of the Royal Society B*, *363*(1493), 1071–1086. https://doi.org/10.1098/rstb.2007.2160

Poeppel, D., & Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, *7*(2), 80–85. https://doi.org/10.1111/j.1467-8721.2008.00553.x

Politzer-Ahles, S., & Jap, B. A. (2024). Can the mismatch negativity really be elicited by abstract linguistic contrasts? *Neurobiology of Language*, *5*(4), 818–843. https://doi.org/10.1162/nol_a_00147

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(10), 1547–1570.

Rhodes, R., Han, C., & Hestvik, A. (2019). Phonological memory traces do not contain phonetic information. *Attention, Perception, & Psychophysics*, *81*(4), 897–911. https://doi.org/10.3758/s13414-019-01728-1

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *62*(6), 437–448. https://doi.org/10.1016/0168-5597(85)90054-1

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

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

Schluter, K. T., Politzer-Ahles, S., Al Kaabi, M., & Almeida, D. (2017). Laryngreal features are phonetically abstract: Mismatch negativity evidence from Arabic, English, and Russian. *Frontiers in Psychology*, *8*(746), 1–19.

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*(6), 728–740. https://doi.org/10.1080/23273798.2016.1151058

Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, *32*(1), 55–65. https://doi.org/10.1111/psyp.1995.32.issue-1

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 92–96). SciPy.

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*(2), 1078–1083. https://doi.org/10.1121/1.428048

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neurosceince*, *8*, 55–65.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, *91*(5), 2979–3000. https://doi.org/10.1121/1.402933

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by seperate neuronal populations. *Neuroreport*, *12*(3), 525–529. https://doi.org/10.1097/00001756-200103050-00019

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *NeuroImage*, *9*(3), 330–336. https://doi.org/10.1006/nimg.1999.0405

Vallat, R. (2018). Pingouin: Statistics in python. *The Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss

Virdis, M. (1978). *Fonetica del dialetto sardo campidanese*. Della Torre.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163. https://doi.org/10.1027/0269-8803.21.34.147

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

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, *102*(6), 1096–1110. https://doi.org/10.1016/j.neuron.2019.04.023

Yu, K., Chen, Y., Wang, M., Wang, R., & Li, L. (2022). Distinct but integrated processing of lexical tones, vowels, and consonants in tonal language speech perception: Evidence from mismatch negativity. *Journal of Neurolinguistics*, *61*, Article 101039. https://doi.org/10.1016/j.jneuroling.2021.101039

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