



Phonological memory traces do not contain phonetic information

Ryan Rhodes¹ · Chao Han¹ · Arild Hestvik¹

© The Psychonomic Society, Inc. 2019

Abstract

We use a “varying standards” oddball paradigm and compare two phonetically differing conditions to find evidence that the auditory cortex has access to discrete phonological representations when making predictions about incoming speech sounds. Brain responses were recorded with a 128-electrode EEG system as subjects passively listened to synthetic speech sounds from a /dæ/-/tæ/ continuum. Deviant stimuli were compared to a control condition to obtain a mismatch negativity (MMN) response, indicative of a “surprise” at a stimulus that deviates from the memory trace. We tested two conditions with phonetically different varying standards – a “low-T” condition with VOT values of 60, 65, and 70 ms, and a “high-T” condition with VOT values of 75, 80, and 85 ms. The amplitude of the mismatch response is generally sensitive to acoustic distance. If the memory trace generated by the auditory cortex contains phonetic information about the standards, then a difference in mismatch amplitude is expected. However, if a phonological memory trace is used, every standard will be represented identically and no difference in mismatch amplitude will be observed. We found no distance effect in our results, indicating an absence of fine-grained phonetic information. This supports the claim that the auditory cortex has access to phonological category representations.

Keywords MMN · Phonological representation · VOT · Perceptual boundary

Introduction

In natural language, speech sounds can be represented at several levels of abstraction. While the sound is by its nature analog and non-discrete, language is discrete and abstract. Representations used in language processing must bridge this gap between continuous sound signal and discrete linguistic representation. At the lowest level, sound can be modeled with explicit *acoustic* representations with information obtained from the auditory sensory system. The highest level uses abstract *phonological* representations using language-specific information, coded in features that lack explicit information about the physical properties of the sound such as frequency,

amplitude, or duration. Between these two extremes is a mediating *phonetic* level that uses language-specific parameters applied to acoustic information, such as voice-onset time (VOT) thresholds.

We are interested in the role that representations at the phonetic and phonological levels play in the early stages of auditory processing and speech perception. The ultimate linguistic goal of the auditory perceptual system is to assign sounds to phoneme categories in order to parse words from the speech stream. Once a sound has been assigned to a category by the perceptual system (e.g. /t/), *phonetic* information about the sound (its particular VOT or formant values) is no longer strictly useful to this goal. Our interest in this research is to test whether the auditory cortex has access to discrete phonological category representations when making predictions about incoming speech sounds.

We report the results of a study using electroencephalographic (EEG) recordings in a varying standards oddball paradigm to elicit auditory cortical responses based on phonological representations. We compared conditions differing in phonetic quality to determine whether phonetic information is

✉ Ryan Rhodes
robot@udel.edu

¹ Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, USA

present in the memory traces used by the auditory cortex to predict incoming sounds, or whether these memory traces are purely phonological.

Phonetic and phonological representation

The feature that distinguishes phonetic categories from phonological categories is the presence of category-internal structure on acoustic parameters. Phonetic categories group sounds according to their acoustic properties, such as VOT or formant values. When listeners categorize sounds on some acoustic continuum (such as VOT), an S-shaped probability curve emerges. Sounds on this continuum are grouped into distinct phonetic categories with only a small area of uncertainty between them. For the VOT continuum for alveolar stops in English, this area lies between 30 and 40 ms (Lisker & Abramson, 1964, 1967; Stevens & Klatt, 1974). This type of categorical perception involves a simple sorting of sounds into categories based on their phonetic properties. This is evidenced by the fact that the effect has been found in pre-linguistic infants (Bertoncini & Bijeljac-Babic, 1987; Eimas, 1974) as well as animals such as the chinchilla (Kuhl & Miller, 1975). Categorical perception has also been found in response to non-linguistic stimuli (J. D. Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore et al., 1977).

Because phonetic categories are defined in terms of these acoustic properties, a phonetic representation of each token would retain this detailed acoustic information. This gradient information is not contained in phonological representations, which consist of discrete units used for representing distinct words in long-term memory. To access lexical items, particular VOT values are not necessary – we only need the identity of the phonemes that compose the word. For example, the word “cat” can be stored in the lexicon with the three phonemes /k/, /æ/, and /t/, without any details about the phonetic realization of each sound.

Phonological categories are defined simply by features, such as [-voice] or [+coronal], and have no category-internal structure. A phonological representation does not distinguish between a /t/ of 60 ms and a /t/ of 90 ms. Every token of /t/ is treated in exactly the same way: simply as a member of the category /t/. We use the term “phonological categories” because they are the target for phonological processes, such as assimilation, syllabification, and stress assignment (Kenstowicz, 1994). These processes simply target members of an entire phonological category, and gradient phonetic information of a particular token is ignored. Phonemes also function as bundles of “bits” for the purpose of storage of information, ensuring the encoding of distinct words in long-term memory.

Evidence of phonetic and phonological categories

Early research in speech perception found that listeners can discriminate members of different categories more easily than members of the same category (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961). However, listeners have been shown to still be able to discriminate members of the same phonetic category, particularly vowels (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973). Carney, Widin, and Viemeister (1977) found that with training, good within-category discrimination of stop consonants was possible. Pisoni and Tash (1974) found faster reaction times when discriminating stimulus pairs with larger formant differences, even with all differences being across-category. This means that the relative acoustic distance between sounds was faithfully represented, in addition to their phonological category membership. This is strong evidence that phonetic categories have gradient structure, and that this structure is represented by listeners when discriminating across category.

The internal structure of phonetic categories has been demonstrated by studies in which listeners judged category goodness of exemplars (J. L. Miller & Volaitis, 1989; J. L. Miller, 1994; Volaitis & Miller, 1992). These results indicate that phonetic categories are defined around prototypes that serve as good examples of those categories. These prototypes are important for defining the space of the phonetic category, and are important anchors for auditory processing.

Samuel (1982) used adaptation to show prototype structure, finding that adaptation was more effective if the adaptor was near the subject’s prototype value. Iverson and Kuhl (1995) and Kuhl (1991) showed a “perceptual magnet effect” in adults and infants, where good examples of phonetic categories (i.e., examples close to the prototype of that category) make neighboring examples seem more similar, which makes them more difficult to discriminate. This is all evidence that phonetic information is available at all stages of auditory processing, and is widely used to make behavioral category judgments (as we should assume given the problem of invariance and a many-to-one mapping between phonetic exemplars and phonological categories). Our question, then, given extensive evidence of graded phonetic representation, is whether discrete phonological category representations are similarly available at all stages of auditory processing.

Electrophysiological evidence

There is also electrophysiological evidence of both phonetic and phonological category discrimination. Many studies have found categorical brain responses to stimuli on an acoustic continuum (e.g., a VOT continuum). Steinschneider et al. (1999) used multiunit invasive EEG recordings to show

categorical neural responses to VOT continua. Their results show a different response for stop consonants with 0–20 ms VOT than for consonants with 40–80 ms VOT. Simos et al. (1998) obtained similar categorical responses using MEG.

Several studies have investigated auditory perception using the mismatch negativity (MMN) response, an automatic response to any change in auditory stimulation generated by the auditory cortex (Alho, 1995). The MMN is typically elicited in an oddball paradigm, in which a repeated sequence of a standard stimulus is interrupted by an oddball, or deviant, stimulus. The appearance of this infrequent deviant causes a “surprise” response, which manifests as a negative deflection relative to the response to the standard stimulus.

In line with the findings of Liberman et al. (1957, 1961), Sharma and Dorman (1999) found a difference between the brain responses to across-category versus within-category contrasts. For the across-category contrast the standard was voiced /d/ (30 ms VOT) and the deviant was voiceless /t/ (50 ms VOT). For the within-category contrast, the standard was voiceless /t/ (80 ms VOT) and the deviant was voiceless /t/ (60 ms VOT). The tokens were different between the two conditions, but the relative distance between standard and deviant was identical (20 ms). Sharma and Dorman measured a higher amplitude MMN effect to the across-category contrast than the within-category contrast, despite the identical phonetic distance.

Winkler et al. (1999) recorded EEG responses to across-category and within-category contrasts in Hungarian and Finnish vowels. In this case, the standard-deviant pairs were identical, but the two types of grammar of the speakers treat the contrast differently. In Finnish, /e/ and /æ/ belong to different categories, whereas Hungarian does not make this distinction. The /e/-/æ/ contrast is across-category for Finnish speakers, but within-category for Hungarian speakers. Winkler et al. tested native Finnish speakers, L2 Finnish speakers, and “naïve” Hungarians with no knowledge of Finnish. They found that MMN amplitudes were greater for subjects who treated the contrast as across-category (L1 and L2 Finnish speakers) than for subjects who treated the contrast as within-category (naïve Hungarian speakers). Dehaene-Lambertz (1997) similarly recorded EEG responses to across-category and within-category contrasts for speakers of different languages. Dehaene-Lambertz presented a /d/-/ / contrast to French and Hindi speakers. In French, the contrast is within category, but in Hindi it is across category. Dehaene-Lambertz, like Winkler et al., found a larger amplitude mismatch effect to the across-category than the within-category contrast.

The difference in amplitude found in these studies suggests that the amplitude of the MMN effect has multiple sources. At the very least, the phonetic distance as well as the category difference both contribute to the overall amplitude of the effect. The category effect appears to contribute much more to the overall amplitude, though. Näätänen et al. (1997) recorded EEG responses of Finnish speakers to two different across-

category contrasts. The standard was /e/ and the deviants were /ö/ (a phoneme in Finnish) and /õ/ (a sound not corresponding to a phoneme category in Finnish). Näätänen et al. found that the mismatch response to the “prototype” deviant was larger than the response to the non-prototype deviant, even though the latter had a larger acoustic difference from the standard.

These results seem to indicate that listeners are assigning sounds encountered in the oddball paradigm to phonological categories, although they do not rule out a phonetic category interpretation. Although there is an asymmetry present – differences across predefined boundaries contribute more to the mismatch amplitude than arbitrary differences – these boundaries can be defined either by phonological category assignment or by phonetic gradation. From a phonological perspective, the difference between /t/ and /d/ is “greater” than the difference between two tokens of /t/. Belonging to different categories induces a greater mismatch effect than belonging to the same category. From a phonetic perspective, the difference between a “bad” exemplar and a “good” exemplar (when one is very far from the prototype value and one is much closer) is a “greater” difference than the difference between two “good” exemplars. In light of this, we cannot be sure that these results tap into true phonological representations.

Varying standards Several EEG studies have used a “varying-standards” approach to enforce phonological representations. Varying standards provides a simple way to elicit phonological categories by introducing phonetic variation into the standards. If the train of standards varies within a phonological category – say, by varying the VOT or formant frequency by small amounts – this constrains the type of viable memory trace representation that can be generated. A memory trace could simply take the form of a phonological category representation, because all standards belong to the same category. Alternatively, the memory trace could incorporate the varying acoustic information into an “*ad hoc*” representation that computes an average or range of the varying acoustic values.

Phillips et al. (2000b) demonstrated that the auditory cortex accesses phonological representations by measuring the brain response to a phonological condition and an acoustic condition. See Fig. 1 for a summary of these conditions. Their phonological condition varied the standards within one category ([+voice]) and varied the deviants within another category ([-voice]). Varying the standards enforced a phonological memory trace of the standards and generated a mismatch when the across-category deviant was encountered. Their acoustic condition uniformly increased the VOT values of all the stimuli so that some of the standards were now in the [+voice] range and some were in the [-voice] range. Because the standards (as defined by the “frequent” class of stimuli) no longer fell into a single category, a phonological memory trace was impossible. They observed no mismatch effect in this condition, indicating that an *ad hoc* phonetic grouping of the

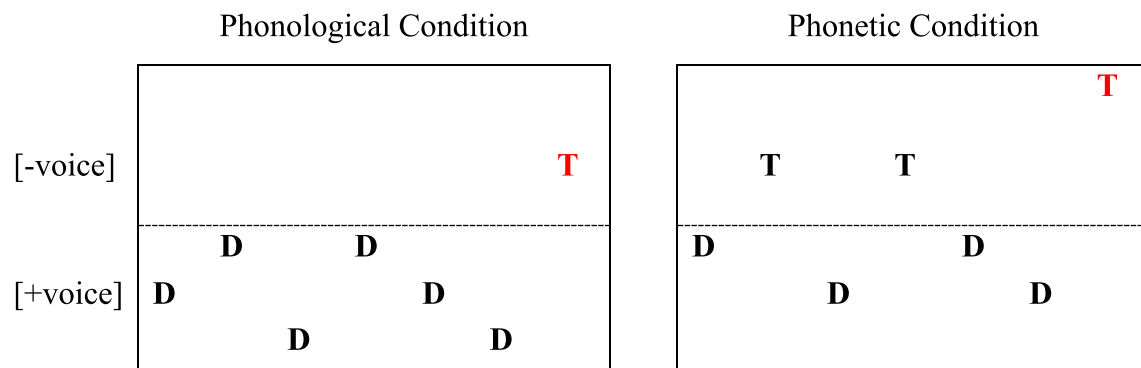


Fig. 1 The two conditions of Phillips et al. (2000b). Standards are marked in black; deviants are marked in red. The phonological condition varied the voice-onset time (VOT) of the standards within the [+voice] category. The deviant fell outside this category, which elicited an MMN. The phonetic condition varied the VOT of the standards by the same amounts, and

the gap between the standards and the deviant was identical to the phonological condition. However, all the VOT values were shifted up, so the standards now fall on both sides of the VOT boundary separating [+voice] from [-voice]. No MMN was found in this condition

standards had not occurred. This serves as strong evidence that the auditory cortex has access to phonological representations, enforced by the varying-standards oddball paradigm.

MMN effects have been elicited in this way by using standards recorded by different speakers (Dehaene-Lambertz & Pena, 2001; Shestakova et al., 2002), standards with variation in their F0 formant frequency (Jacobsen & Schröger, 2004), and variation in VOT (Hestvik & Durvasula, 2016; Phillips, Pellathy, & Marantz, 2000a; Phillips, Pellathy, Marantz, et al., 2000b). Eulitz and Lahiri (2004) and Hestvik and Durvasula (2016) used varying standards to tap into further properties of phonemes, namely phonological underspecification.

However, phonetic categories are still subject to language-specific parameters. The boundary between /t/ and /d/ in English is well established, and a phonetic grouping of sounds on the VOT continuum will cluster around the phonetic prototypes. The inability of the auditory cortex to perform an *ad hoc* grouping of the tokens in the Phillips et al. study may have been due to their distance from category prototypes. If a token is sufficiently distant from a relevant phonetic prototype, it may no longer be possible for the auditory system to assign that token to the necessary category. In other words, phonetic grouping appears to be subject to the same category constraints as phonological grouping. A novel phonetic grouping cannot be created on the fly.

The failure to elicit a mismatch effect in Phillips et al.'s "phonetic" condition demonstrates that *ad hoc* phonetic grouping across categories (such as tokens falling between 30 and 60 ms VOT) is not possible. But the question remains of whether there is detailed phonetic information present in the memory trace, or whether varying the standards necessarily causes the auditory cortex to treat each token as an identical member of a phonological category. Phonological category membership entails no distinction among its members. All tokens are of equivalent status. Phonetic category membership, however, is graded. Members of a phonetic category

may be treated differently, based on the quality of their acoustic properties.

The goal of the current study was to determine whether the auditory cortex is truly accessing *phonological* categories when standards are varied – categories in which every token belonging to that category is treated as identical – or accessing *phonetic* categories – language-specific categories in which differing tokens have unique status. In other words, does the auditory cortex utilize phonological category representations to make predictions about incoming speech sounds, as assumed by Phillips et al. and subsequent studies?

The mismatch negativity (MMN)

This question will be addressed by employing the varying-standards mismatch negativity (MMN) paradigm to see if participants maintain fine-grained phonetic information while making phonological category predictions. The mismatch response is a function of the auditory cortex's ability to generate predictions about incoming sounds.

The MMN is a reflection of the brain's automatic response to any change in auditory stimulation, generated in the auditory cortex (Alho, 1995). Since it is elicited by an unexpected auditory change, the MMN is functionally a neural "surprise" response. Because it is automatic, the mismatch response can be elicited in both attend and non-attend contexts (István Winkler, Czigler, Sussman, Horváth, & Balázs, 2005; Näätänen, 1979, 1985).

In order for the auditory cortex to react to a change in auditory stimulation (and generate an MMN effect), it must make active predictions about incoming sounds. These predictions are informed by a representation (referred to as a memory trace) constructed from the repetitive sequence of standard stimuli. When a deviant is presented, the incoming sensory information "clashes" with the prediction generated by the memory trace representation and elicits an MMN

response (István Winkler, Cowan, Csépe, Czigler, & Näätänen, 1996a; István Winkler, Karmos, & Näätänen, 1996b; Näätänen & Picton, 1987). Studies with backward-masking stimuli suggest that this memory trace is established and held in auditory sensory memory. The memory trace representation can persist for several seconds (Winkler, Schröger, & Cowan, 2001).

The amplitude of the MMN effect increases and the latency shortens as the magnitude of the auditory change increases (Näätänen, Paavilainen, Rinne, & Alho, 2007; Rinne, Särkkä, Degerman, Schröger, & Alho, 2006). When the deviant differs from the standard in more than one attribute, the MMN amplitude shows an additive effect (Schröger, 1995; Wolff & Schröger, 2001). This is evidence that the auditory processing system is sensitive to fine-grained acoustic variation, and that these details can be encoded in memory-trace representations. This is also evidence of a gradient prediction evaluation mechanism. When an incoming sound is evaluated against a prediction, the amplitude of the surprise response (the MMN) is proportional to the difference between the new incoming sound and the representation of the past sounds (i.e., prediction error; cf. Friston, 2005).

Our study compared mismatch responses to two conditions with different distances, operationalized as the distance between oddball VOT and the mean VOT of the standards. In both conditions, the standards will be tokens of /t/ with varying VOT values and the deviant will be /d/ with a VOT of 15 ms. One condition will be a “low-T” condition, in which the standards have an average VOT of 65 ms. The other condition will be a “high-T” condition, in which the standards have an average VOT of 80 ms. These conditions are illustrated in Fig. 2.

Because the amplitude of the mismatch effect is sensitive to the magnitude of the change between stimuli, we predict a “distance effect” predicated on the relative distance between

the deviant VOT and the mean VOT of the standards. Such a distance effect would indicate that the auditory cortex is only able to access phonetically graded representations (rather than phonological category representations). If the auditory cortex keeps track of this information and encodes it into the memory trace, we predict that the condition with higher VOT standards will elicit a higher amplitude mismatch effect than the condition with lower VOT standards, due to the greater distance between the deviant and standards in the “high-T” condition (65 ms) than in the “low-T” condition (50 ms). However, if the auditory cortex utilizes phonological category representations in which all category members are treated with an “all-or-none” principle (i.e., all category members are treated as identical), there should be no difference in amplitude between the two conditions.

Most MMN studies of this type compare the brain response to /d/ as a deviant to the response to /d/ as a non-deviant. This avoids a confound due to the intrinsic difference in the neural responses to /d/ and /t/ sounds. Because the standards and deviants differ in VOT, they may induce latency differences as the brain responds to the consonant burst and the onset of voicing at different times. In the classic MMN paradigm, this confound is avoided by comparing target sounds as deviants with the same sounds as standards. However, this does not rule out the possibility that the MMN effect results from neuronal refractoriness. Repeated stimulation of neuronal populations (e.g., the neuronal population that responds to /t/) induces habituation, with the total activation of that population decreasing over time (Butler, Spreng, & Keidel, 1969). Because the standards appear more frequently than the deviants, neural responses are more habituated to the standards (and are attenuated), while neural responses are less habituated to the deviants.

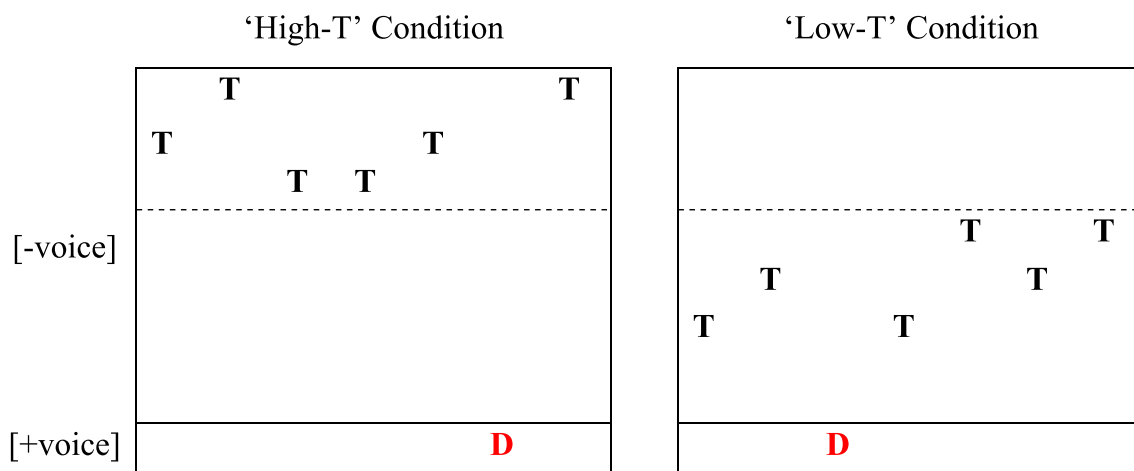


Fig. 2 The two experimental conditions of Experiment 1. The solid black line represents the boundary between [+voice] and [-voice] (at about 35 ms). The dotted line separates the “High-T” standards from the “Low-T” standards. Each condition compares the [-voice] standards of one group

(high or low) against the [+voice] deviant. If the representation contains detailed phonetic information (voice-onset time; VOT), there will be a difference in amplitude between the two conditions

To control for both of these effects, we here use the “random control” paradigm of Jacobsen and Schröger (2001, 2003) to elicit and measure the MMN. Rather than comparing deviants directly to themselves as standards, this method uses a “random-control” block where the occurrence of the target /d/ sound has a frequency equal to its frequency in the standard-deviant blocks, but it is embedded in a continuum of randomly-occurring equi-probable stimuli. By comparing the deviant stimulus to itself in the random control condition, we can control for inherent differences in brain response between the standard and deviant stimuli and ensure that the expected MMNs come from the memory comparison. The deviant in the control condition will serve as a control to compute the mismatch effect in lieu of the standard stimuli.

An MMN difference wave computed from standards and deviants has several underlying sources. One source is simply acoustic difference – acoustically different sounds activate different neuronal populations. Another source is neuronal refractoriness – repeated activation of a neuronal population leads to habituation, decreasing the amplitude of the response. This results in an attenuation of the brain response to the standards relative to the deviants. Another source is the neural “surprise” response, which is a function of a failed prediction about incoming sounds. To be sure that we are measuring the surprise response, rather than differences arising from acoustics and neuronal refractoriness, we use a deviant-deviant comparison – comparing the brain response to the deviant from a standard-deviant block to the deviant in a random-standards control block.

Methods

Participants

Two groups of subjects were recruited. Group 1 consisted of 23 subjects (nine male), and Group 2 consisted of 26 subjects (nine male), for a total of 49 subjects. All subjects were undergraduates at the University of Delaware, native speakers of English, and reported no history of hearing impairments. The average age of the participants was 22.5 years ($SD = 4.6$). Subjects were either paid US\$20 or given extra credit for their participation. All study procedures were approved by the University of Delaware Internal Review Board (IRB) and are compliant with the principles for ethical research established by the Declaration of Helsinki.

Stimuli and design

The stimuli were a sequence of synthesized CV syllables composed of an alveolar stop and the vowel [æ]. Each syllable had a duration of 290 ms. The stimuli were adapted from Hestvik and Durvasula (2016), and were generated via Klatt

Synthesizer to exactly reconstruct the stimuli used in Phillips et al. (2000b). The onset consonant of the deviant stimulus had a VOT value of 15 ms. The consonant onset of the standard stimuli had VOT values of 60, 65, 70, 75, 80, and 85 ms.

VOT values for deviant and standards were chosen on the basis of the robust and reliable VOT boundary separating voiced from voiceless among English speakers, which lies around 30–40 ms (Lisker & Abramson, 1964, 1967; Stevens & Klatt, 1974). The categorical discrimination task in Hestvik and Durvasula (2016) found that the mode of the boundary between /d/ versus /t/ was 40 ms with a standard deviation of 3.6 ms (p. 31). Given that the VOT value in the /d/ category (15 ms) and the smallest VOT value in the /t/ category (60 ms) are both quite far away from the 40-ms threshold, it is reasonable to assume that all subjects accurately perceived the deviant as /d/ and standards as /t/.

The experiment consisted of three blocks, corresponding to three conditions: Low-T, High-T, and Control. Deviants from each experimental block (Low-T, High-T) would be compared with the deviant from the Control block to establish main MMN effect. The two experimental-control differences would then be compared to observe a potential “distance effect.” Each block contained 1000 trials. For both Low-T and High-T blocks, the 1,000 trials consisted of 900 standards (90%) and 100 deviants (10%). The /tæ/ stimuli in the Low-T condition had an onset [t] with a VOT value of 60, 65, or 70 ms. The /tæ/ stimuli in the High-T condition had an onset [t] with a VOT value of 75, 80, or 85 ms. The oddball /dæ/ in both Low-T and High-T conditions had an onset [d] with a VOT value of 15 ms. The presentation of trials in the High-T and Low-T conditions is pseudorandomized such that there are at least three standards between every two deviants. The inter-stimuli interval (ISI) in each condition randomly varied from 410 to 600 ms to avoid phase-locked brain responses.

Following Jacobsen and Schröger (2001, 2003), the Control condition was constructed such that the occurrence of the target /dæ/ appeared in a randomized sequence of equi-probable varying sounds. Because the target /dæ/ is identical to the token used as a deviant in the High-T and Low-T conditions, we refer to it as a “deviant” appearing among “random standards.”

Two groups of subjects were run, with different versions of the random-control condition. For Group 1, the Control condition consisted of ten stimuli with VOT values of 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 ms, with each type constituting 100 trials (10%). For Group 2, the Control condition consisted of five stimuli with VOT values of 5, 10, 15, 20, and 25 ms, with each type constituting 100 trials (20%). Note that the random control standards of Group 1 incidentally crossed a phonetic category boundary (some tokens will be perceived as /t/), which could cause a (small) potential mismatch effect if a relatively high VOT standard appeared immediately before the 15-ms deviant. To test whether this small stimulus mistake

had an effect, Group 2 was recruited and run with random standards all within the same category: random tokens vary at 5-ms intervals from 5 ms to 25 ms VOT. This introduced a different frequency for the tokens (20%), but kept all the tokens in the same category (/d/), so no incidental category contrast was possible. As our results show, the change of stimuli in the random control condition was innocuous (no main effect or interaction was observed; see *Results*).

Procedure

Subjects were seated in a sound-attenuating booth and were exposed to the stimuli presented by two speakers. Subjects were instructed to watch a silent movie (*The Wizard of Oz*) and were told to ignore the auditory stimuli. We did not include an active task so that the results would not be contaminated by an N2b component, which is also sensitive to mismatch, but only occurs when the stimuli are task-relevant (Luck, 2014). The order of the High-T, Low-T, and Control blocks was randomized for each subject. The entire recording session took approximately 1 h.

Apparatus, data acquisition, and data processing

The experiment was programmed using E-Prime v. 2.0, and the E-Prime Extension package for Net Station was used for the EEG data acquisition. Continuous EEG data were recorded from 128 carbon fiber core/silver-coated electrodes in an elastic electrode net (Geodesic Hydrocel 128) and was digitized with the EGI Net Station software v. 4.5 with a sampling rate of 250 Hz. Before data acquisition, electrode impedances were lowered to below 50 k Ω . Subjects' electro-ocular activity were recorded from two bipolar channels. The vertical electro-oculogram (EOG) was recorded with the supraorbital and infraorbital electrodes of both eyes; the horizontal EOG was recorded with the electrodes located at the outer canthi of both eyes. The Cz electrode placed on the center point of the head was used as the reference site.

After the acquisition, the data were passed through a 0.3-Hz FIR high-pass filter. Then the continuous EEG was segmented into epochs of 1,000 ms. Each epoch included a 200-ms pre-stimulus period and an 800-ms period after the stimulus onset. The segmented data was baseline-corrected based on the mean voltage of the 200-ms pre-stimulus period. The data were then submitted to an automated process of eyeblink subtraction using ICA with the ERP PCA toolkit v. 2.64 run on MATLAB R2016b. An eyeblink template was generated from each subject's data via visual inspection. An ICA component was marked as an eyeblink component and was subtracted from the data if it was correlated at $r = .9$ or greater with the manually created eyeblink template. After eyeblink subtraction, the data were submitted to the artifact correction procedure for bad channels and movement artifacts. A channel was

marked bad if its best absolute correlation with its neighboring channels fell below .4 across all time points. Bad channels were replaced via interpolation from surrounding good channels. If a channel was marked bad in over 20% of trials, it was considered bad in all trials. A trial was zeroed out if it contained more than 10% bad channels. The remaining trials were averaged into six cells: High-T-deviants, High-T-standards, Low-T-deviants, Low-T-standards, Control-deviants, and Control-standards. The data were then re-referenced to the average voltage across all electrodes.

Evoked-response potential (ERP) analysis

We utilized a principal-components analysis (PCA) to determine the electrode regions and time windows for evoked-response potential (ERP) analysis. PCA is suitable for a high-density recording montage, and it provides a more objective way of selecting time windows and electrode regions for analysis than visual inspection (Dien & Frishkoff, 2005; Dien, Khoe, & Mangun, 2007; Dien, 2010, 2012; Spencer, Dien, & Donchin, 1999, 2001), following one of the recommendations of Luck and Gaspelin (2017) to remove experimenter bias. The PCA decomposes the temporal and spatial dimensions into a linear combination of a smaller set of abstract ERP factors based on covariance patterns among time points and electrode sets. In this way, the PCA can tease apart the underlying contributions of the factors to the summed scalp activity.

For the input to the PCA, we used two difference waveforms (Low-T minus Control and High-T minus Control) to identify the main mismatch effect of the Low-T condition and the High-T condition. We used the difference waves instead of the absolute waveforms as we wanted the PCA to focus on the temporal and spatial fluctuations of the mismatch effect itself. The first difference waveform was computed by subtracting the response to the deviant in the Low-T condition minus the same deviant in the Control condition. The second difference waveform was computed by subtracting the response to the deviant in the High-T condition minus the response to the same deviant in the Control condition. These difference waveforms represent the attenuation of the Auditory Evoked Potential (an automatic response to sound) as a result of mismatch, and were used as the input for the PCA to extract the temporal and spatial factors. We then used these factors to constrain the selection of time windows and electrode regions used in the statistical analysis on the undecomposed voltage data.

For the statistical analysis, we first conducted a conventional comparison between standards and deviants in the High-T and Low-T conditions, then we utilized the random Control condition to confirm that the observed effect was not due to the intrinsic difference of the stimuli.

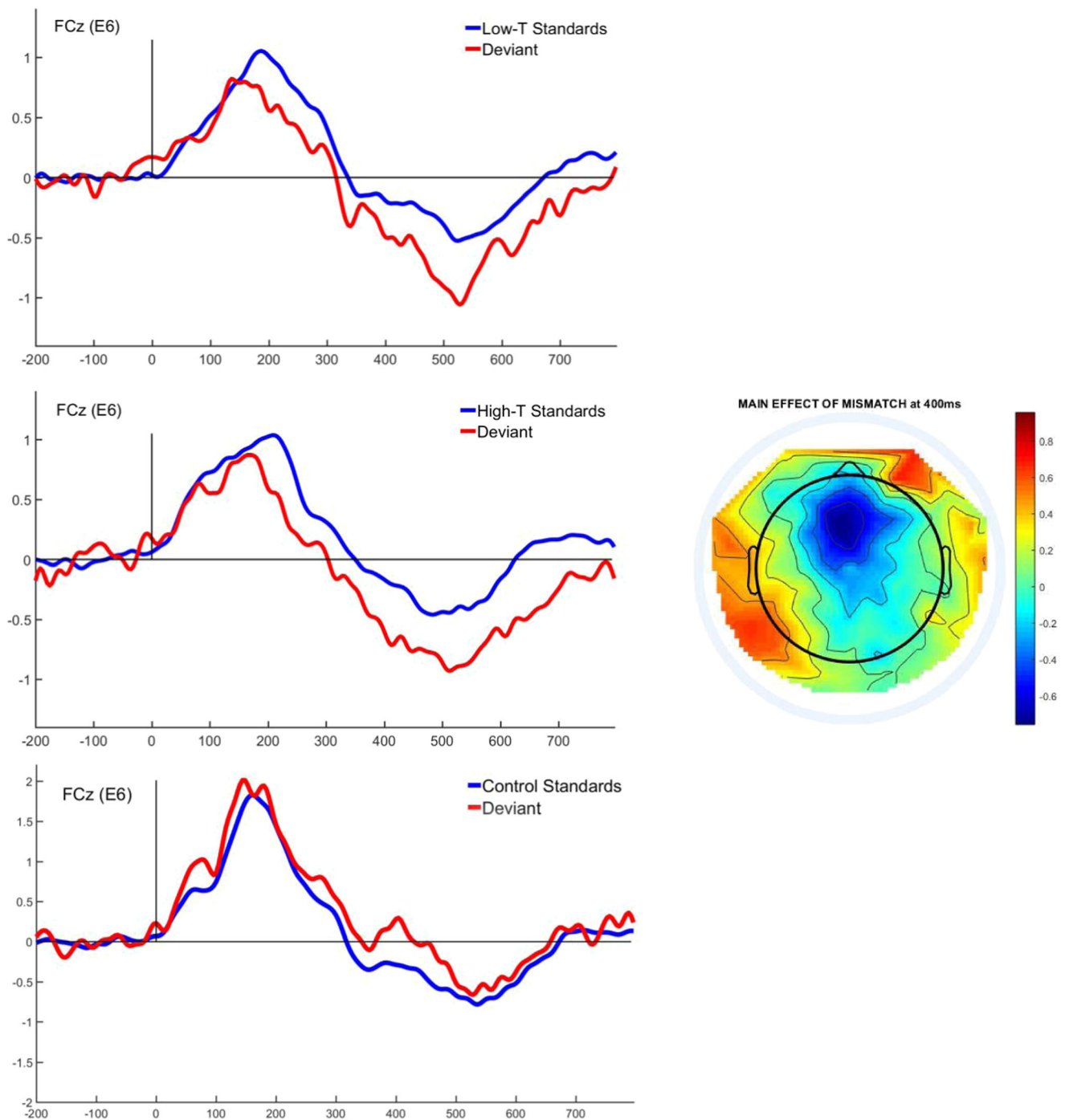


Fig. 3 Grand average waveforms of deviants vs. standards in both conditions. From top: Low-T condition, High-T condition, and Control condition. There is a persistent negativity in the experimental conditions,

but none in the control condition, as expected. The topoplot (far right) shows the difference between the control deviant and the mean of the two deviant waveforms

Results

After the EEG recording and pre-processing, nine subjects' data were excluded due to having either more than 10% of bad channels or more than 25% of bad trials in the whole session. The data of the remaining 37 subjects (24 female) entered the analysis process. Of

the 37 subjects, 20 were from Group 1, and 17 were from Group 2. The mean proportions of good trials were 94% (SD = 10%) in the High-T condition, 93% (SD = 13%) in the Low-T condition, and 95% (SD = 7%) in the Control condition

A visual inspection of the grand average waveforms of the difference between standards and deviants in the

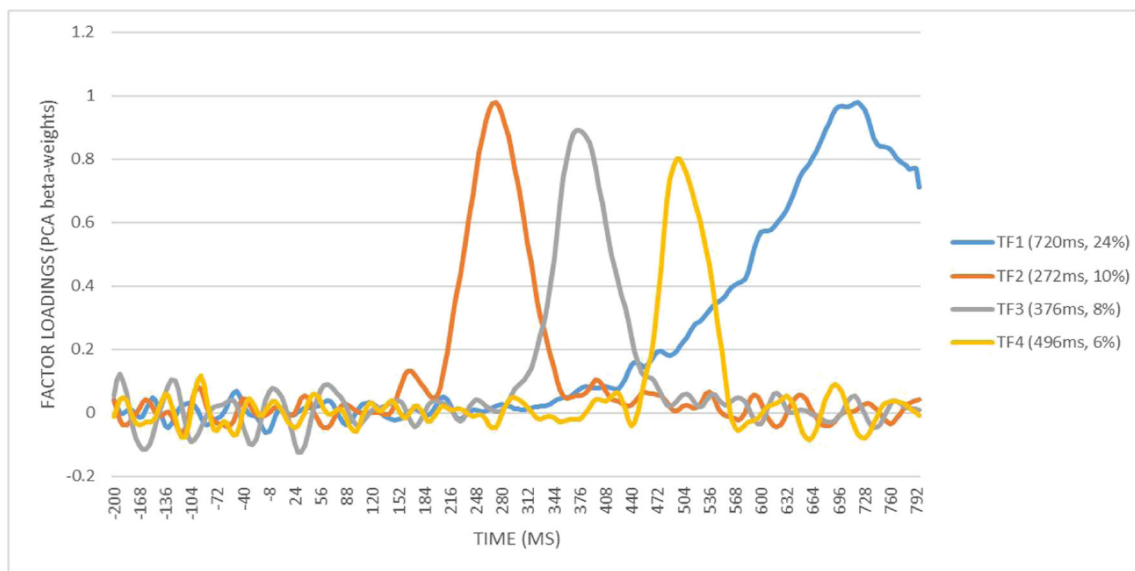


Fig. 4 Difference waves temporal factor loadings for factors accounting for at least 5% of the variance, labeled by their peak latency and amount of variance accounted for

high versus low conditions showed that the mismatch was greatest around 500 ms (see Fig. 3). A topoplot of the main effect difference between the two deviants and the control deviant showed the effect to have a typical fronto-central distribution.

However, because informal visual inspection of the data is subject to experimenter bias (Luck & Gaspelin, 2017), we used a PCA to objectively define the time window and spatial distribution of the mismatch effect for further analysis. We first applied a temporal PCA to the two difference waveforms (i.e., Low-T deviant minus Control deviant, and High-T deviant minus Control deviant), to capture the temporal and spatial dynamics of the mismatch effect. The scree test in combination with the Parallel Test of the temporal PCA generated 22 temporal factors, accounting for 90% of the total variance. We then screened out the factors that accounted for less than 5% of the total variance. This left four temporal factors with the following temporal factor loadings: TF1 accounted for 24% of the total variance and peaked at 720 ms; TF2 accounted for 10% of the total variance and peaked at 272 ms; TF3 accounted for 8% and peaked at 376 ms; TF4 accounted for 6% and peaked at 496 ms. See Fig. 4 for a summary of the factor loadings and their temporal distribution.

The temporal and spatial distributions of the factor score waves (back-projected into voltage space) of all four temporal factors are shown in Fig. 5. Note that only TF3 and TF4 had a spatial distribution consistent with an MMN effect, with TF3 showing a central distribution and TF4 a more anterior distribution, and are thus retained for the further analysis.

In the next step, the 22 temporal factors were used as input for an Independent Component Analysis (ICA) using Infomax

rotation to further decompose each temporal factor into spatial sources of variance. A scree test found seven spatial factors for each temporal factor. Visual inspection of the resulting spatial subfactors revealed that only two temporal-spatial factors, TF3SF2 (peak latency 376 ms) and TF4SF1 (peak latency 496 ms), had a topography consistent with an MMN effect (centro-frontal negativity with inversion at the mastoids (Näätänen et al., 2007; Näätänen & Picton, 1987)). However, only TF4SF1 was retained as the temporal factor 4 most closely corresponded to the peak of the difference wave (around 500 ms) in the undecomposed voltage data.

We therefore took TF4SF1 to represent the most prominent mismatch ERP component. We defined the time window as each time sample in TF4 with a factor loading score of 0.6 or greater, which gave a time window of 484–516 ms. We defined the spatial region of interest (ROI) as every electrode in TF4SF1 with a factor loading score of 0.6 or greater. The ROI is shown in Fig. 6 (shown in red).

The mismatch effect was tested first by conducting a two-way ANOVA for the conventional standard-deviant comparison. The dependent measure was the voltage averaged over the defined time window and the ROI. The two within-subject factors were Condition Type (High-T vs. Low-T) and Stimulus Type (Standard vs. Deviant). As the MMN effect is expected in both High-T and Low-T conditions, we should observe a main effect of Stimulus Type. If the deviant after high standards exhibited a greater attenuation than the deviant after low standards did (supporting the brain tracking the fine-grained acoustic details), we should also observe an interaction between Condition Type and Stimulus Type. The ANOVA result

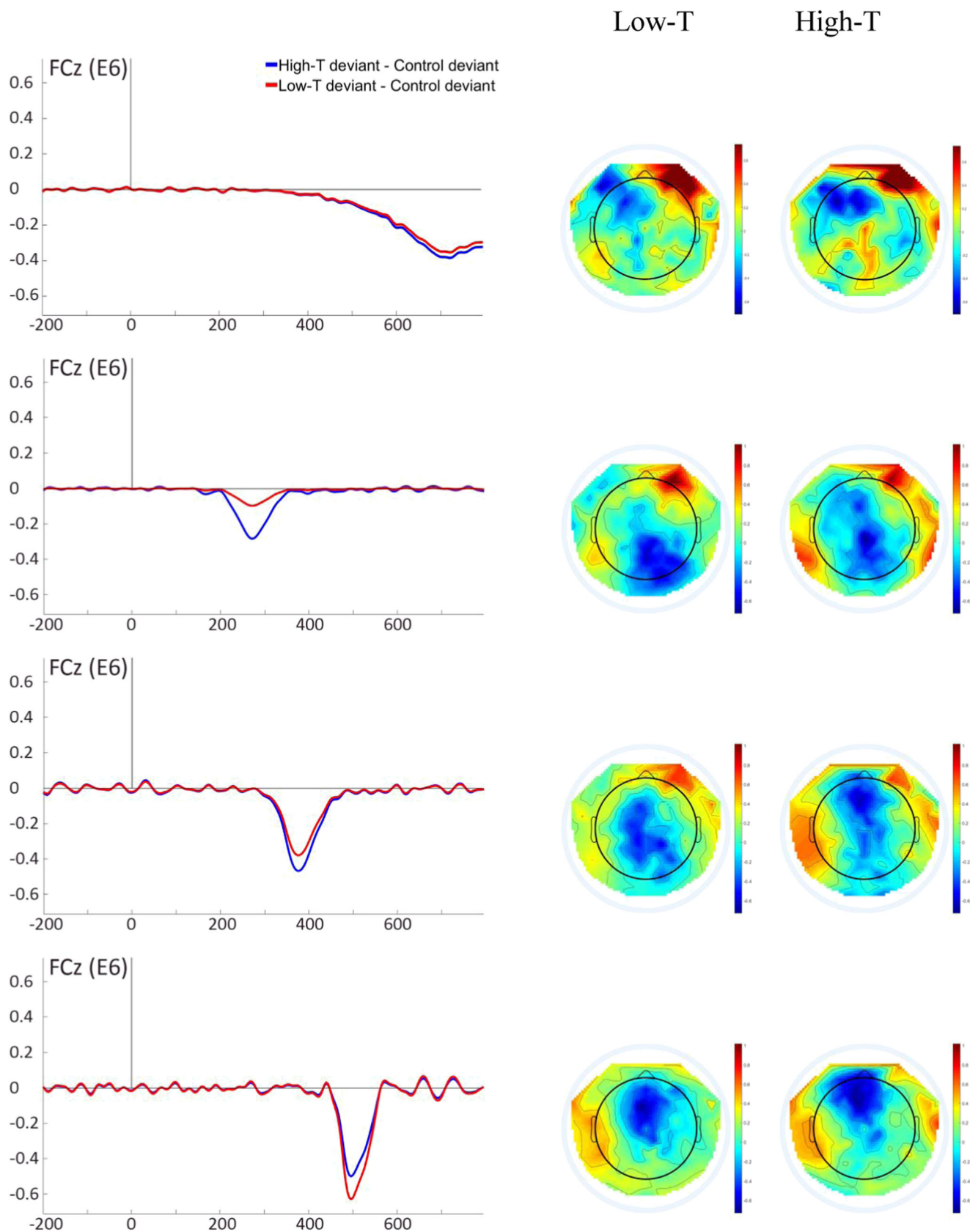


Fig. 5 Factor scores by condition back-projected into voltage space (shown at FCz) and their spatial distribution (Top-down: TF1, TF2, TF3, TF4). The Low-T condition is shown with the topoplots on the left;

the High-T condition is shown with the topoplots on the right. All data shown are the grand average of all subjects

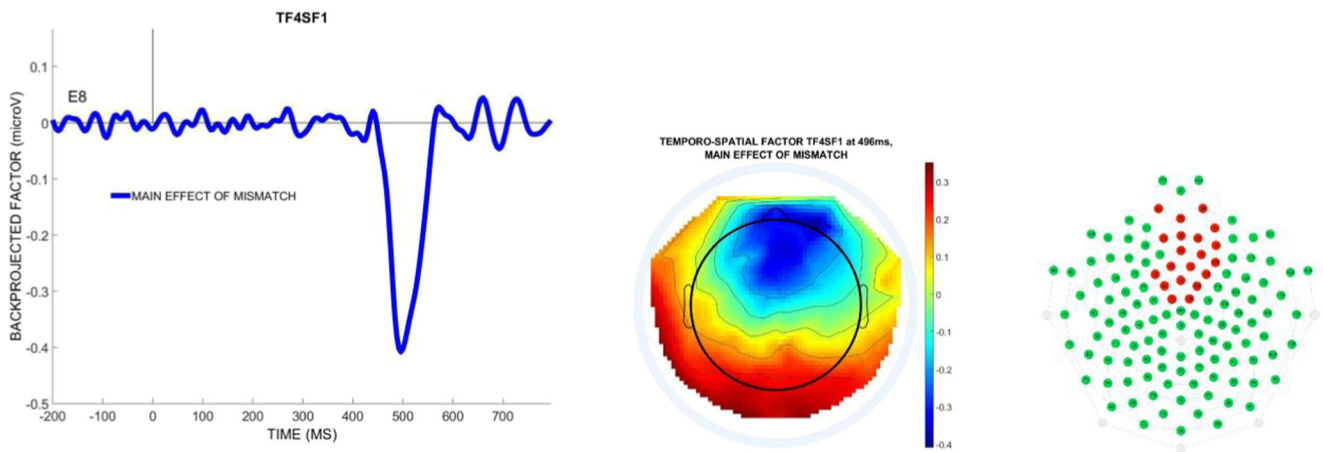


Fig. 6 Time course, spatial distribution, and region of interest of TF4SF1, defined by the principal-components analysis (PCA)

did not yield a significant interaction ($p = .962$, with the observed power of .5), indicating a lack of support for the fine-grained phonetic memory trace. There was a main effect of Stimuli Type [$F(1, 36) = 8.366$, $p = .006$, partial $\eta^2 = .189$], reflecting an overall mismatch effect. There was no effect of Condition Type ($p = .276$).

To make sure that the mismatch effect obtained above is not an artifact of physical stimulus differences between standard stimuli and deviant stimuli, we compared the deviant waveform in the High-T and Low-T conditions to the same 15-ms VOT stimulus in the random standards control condition. A pairwise t-test comparing Standard and Deviant in the Control condition showed no significant difference, as expected [$t(36) = 0.376$, $p = 0.7$]. This indicates that there was no mismatch response to the 15-ms “deviant” in the Control condition. This response to the deviant in the absence of a memory trace prediction will allow it to serve as a neutral point of comparison, in lieu of a conventional standard-deviant comparison. Figure 7 shows the computed mean deviant

waveforms for the PCA-defined electrode region for TF4SF1 for the three conditions.

We then conducted a one-way repeated measures ANOVA with the within-subject factor DEVIANT TYPE (High-T vs. Low-T vs. Control). A genuine effect would be reflected in a significant difference between the High-T versus Control, and Low-T versus Control. Since we have two versions of Control, we also included the between-subject factor GROUP (Control Condition Group 1 vs. Control Condition Group 2). There was no interaction ($p = .444$) or main effect of Control Condition Group ($p = .369$), but there was a main effect of DEVIANT TYPE [$F(2, 70) = 3.303$, $p = .043$, $\eta^2 = .086$], with the observed power of .608. This means that the three deviant waveforms were not of the same amplitude. Planned contrasts revealed a significant difference between High-T and Control [$t(36) = 2.253$, $p = .030$], as well as a difference between Low-T and Control [$t(36) = 2.11$, $p = .042$], which indicated that the mismatch effect observed in the conventional standard-deviant comparison was not subject

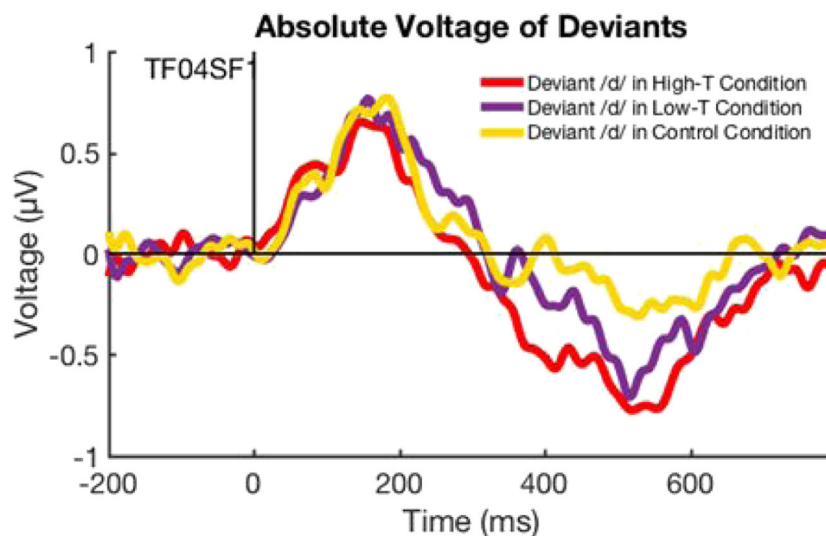


Fig. 7 The deviant waveforms in High-T, Low-T, and Control conditions. The deviant had a voice-onset time (VOT) value of 15 ms in each condition

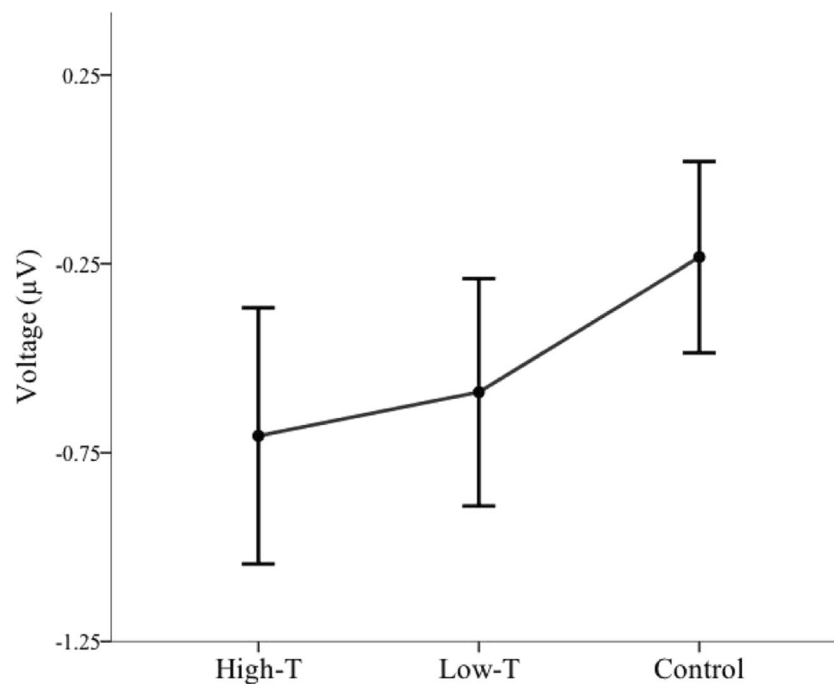


Fig. 8 Mean voltage of the deviant in the three conditions (vertical bars denoting 95% confidence interval)

to the intrinsic difference of the stimulus in both conditions. Consistent with the result of the lack of interaction in the conventional standard-deviant comparison, there was no difference between the High-T and Low-T ($p = .497$). Figure 8 shows the mean voltage of the responses to deviants as a function of DEVIANT TYPE.

The probability of a true null effect was examined by estimating a Bayes factor using Bayesian Information Criteria (BIC; Wagenmakers, 2007). This compares the fit of the data under the null hypothesis compared to the alternative hypothesis by comparing models that include and models that do not include each factor and interaction. This analysis produces an inverted Bayes Factor (BF01) ratio, indicating the likelihood of the null hypothesis being true. Higher BF01 values indicate greater evidence in favor of the null hypothesis.

We performed two Bayesian repeated measures ANOVAs with default priors (Rouder, Morey, Speckman, & Province, 2012). A repeated measure 2×2 ANOVA comparing Condition Type (High-T, Low-T) and Stimulus Type (Standard, Deviant) found moderate evidence for the null hypothesis (Condition*Stimulus interaction BF01 = 4.79). This suggests that the null is 4.79 times more likely to be true than the alternative. A repeated measure one-way ANOVA was also performed comparing the values of deviants in each condition (Control, High-T, Low-T). A *post hoc* comparison between High-T and Low-T found moderate favor for the null (BF01 = 4.54), suggesting the null is 4.54 times more likely than the alternative.

These results suggest that the amplitude of the MMN response is not significantly affected by distance between standards and deviant (measured in VOT).

Discussion

We found a significant mismatch effect in both High-T and Low-T conditions, as expected. The largest underlying component of this mismatch effect (as determined by the PCA) occurred in the relatively late time window of 484–516 ms. Negativity in this time window, alternately labeled either late MMN or LDN (late discriminatory negativity), has been reported by several previous MMN studies (Cheour et al., 1998; Datta, Shafer, Morr, Kurtzberg, & Schwartz, 2010; Shafer, Morr, Datta, Kurtzberg, & Schwartz, 2005). The late MMN/LDN has a spatial topography consistent with a typical early MMN, consisting of a fronto-central negative component and a positivity at the mastoids, which we also observed (see, e.g., Hestvik & Durvasula, 2016; Neuhoﬀ et al., 2012).

We also observed a small negativity in a more typical earlier time window (~150 ms), but this was determined by the PCA to account for only a small amount of the total variance. For this reason, we focused our analysis on the late time window. The significant main effect indicates that the /t/ standards were perceived as different than the /d/ deviant, as expected.

If the memory trace was phonetic (rather than phonological), we should have found a difference in amplitude between the two conditions, due to the difference in distance between standards

and deviant (50 ms for the Low-T condition and 65 ms for the High-T condition). A difference in the amplitude of the mismatch effect between the conditions would have indicated the presence of phonetic information (VOT) in the memory trace. However, we find no significant difference in the amplitude of the mismatch effect between the two conditions, despite the difference in distance between standards and deviants.

The absence of a significant difference between the phonetically different conditions supports the interpretation of Phillips et al. (2000a, 2000b) that the auditory cortex accesses phonological representations when the standards are varying within category. Because phonological categories treat all members equivalently, there is no distinction to be made between the High-T standards and Low-T standards – from a phonological perspective they are all simply /t/. The absence of a distance effect suggests that the representation that formed the memory trace did not make use of any fine-grained phonetic information, but treated the distance between deviant and standard as equivalent in both conditions. While Phillips et al. showed that listeners cannot perform *ad hoc* grouping of sounds belonging to phonological categories, we expanded on their results by demonstrating that phonetic information is absent from the memory trace even when all the standards are varying within a single phonological category. This serves as further evidence that phonological category representations are available to the auditory cortex.

Mismatch

In order to make a phonological categorization, it is necessary to map the acoustic properties of the sound to phonological features. This may be done by creating a phonetic category representation or judging the proximity of the sound to a phonetic prototype. Our results suggest that by the time several variant sounds have been encountered, this type of phonetic information is absent. Because the MMN response is a function of prediction – the auditory cortex creates a memory trace of encountered sounds to predict incoming sounds – these results imply that once a phonological category has been assigned, there is little value for phonetic information. The ultimate goal of speech perception must be to assign categories to sounds to build representations that allow lexical retrieval. Assigning a phonological category to a sound seems to satisfy this demand.

The problem of invariance, as well as phonological processes that result in a many-to-one mapping of phonetic exemplars to phonological category representations, necessitates that phonetics play a critical role in this process. Here we have shown that even though this phonetic information is available, it may not necessarily be of predictive value. In a stream of phonetically varying sounds, the brain may opt for a simple predictive solution: pre-defined phonological categories.

Conclusion

In this study, we used the “varying-standards” oddball paradigm to enforce genuine phonological memory traces – i.e., representations of speech sounds which do not maintain any relevant phonetic information. Our results support the interpretation of previous studies that the auditory cortex can access phonological category representations and use those representations to make predictions about incoming sounds. Because phonological categories have no internal structure, all category members are treated identically. By contrast, phonetic categories have internal structure predicated on acoustic properties such as VOT, with the category labels clustering around prototypes (good exemplars of the phonetic category). Because we failed to observe a significant difference between phonetically different (but phonologically identical) conditions, we conclude that the auditory processing system in fact has access to and actively utilizes phonological representations for making predictions about incoming sounds.

Data availability None of the data or materials for the experiments reported here are available, and none of the experiments were preregistered.

References

- Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear & Hearing, 16*, 38–51.
- Bertoncini, J., & Bijeljac-Babic, R. (1987). Discrimination in neonates of very short CVs. *The Journal of the Acoustical Society of America, 82*(1), 31. <https://doi.org/10.1121/1.395570>
- Butler, R. A., Spreng, M., & Keidel, W. D. (1969). Stimulus repetition rate factors which influence the auditory evoked potential in man. *Psychophysiology, 5*(6), 665–673.
- Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America, 62*(4), 961–970. <https://doi.org/10.1121/1.381590>
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351–353. <https://doi.org/10.1038/1561>
- Datta, H., Shafer, V. L., Morr, M. L., Kurtzberg, D., & Schwartz, R. G. (2010). Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development. *Journal of Speech, Language, and Hearing Research, 53*, 757–778.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*(4), 919–924.
- Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Cognitive Neuroscience and Neuropsychology, 12*(14), 3155–3158.
- Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138–145. <https://doi.org/10.1016/j.jneumeth.2009.12.009>

- Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Developmental Neuropsychology*, 37(6), 497–517. <https://doi.org/10.1080/87565641.2012.697503>
- Dien, J., & Frishkoff, G. A. (2005). Introduction to principal components analysis of event-related potentials. In T. Handy (Ed.), *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press.
- Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax vs. infomax rotations. *Human Brain Mapping*, 28(8), 742–763. <https://doi.org/10.1002/hbm.20304>
- Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 16(3), 513–521. <https://doi.org/10.3758/BF03198580>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189.
- Hestvik, A., & Durvasula, K. (2016). Brain & Language Neurobiological evidence for voicing underspecification in English. *152*, 28–43.
- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97(1), 553–562.
- Jacobsen, T., & Schröger, E. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, 41, 654–659. <https://doi.org/10.1111/j.1469-8986.2004.00175.x>
- Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107. Retrieved from <https://link-springer-com.proxy3.library.mcgill.ca/content/pdf/10.3758/BF03212211.pdf>
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190(4209), 69–72. <https://doi.org/10.1126/science.1166301>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368. <https://doi.org/10.1037/h0044417>
- Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61(5), 379–388. <https://doi.org/10.1037/h0049038>
- Lisker, L., & Abramson, A. S. (1964). Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*, 20(3), 384–422. <https://doi.org/10.1121/1.2142685>
- Lisker, L., & Abramson, A. S. (1967). Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech*, 10(February), 1–28. <https://doi.org/10.1177/002383096701000101>
- Luck, S. J. (2014). *An introduction to the event-related potential technique, Second Edition*. Cambridge, MA: MIT Press.
- Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146–157. <https://doi.org/10.1111/psyp.12639>
- Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *The Journal of the Acoustical Society of America*, 60(2), 410–417. <https://doi.org/10.1121/1.381097>
- Miller, J. L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, 50(1–3), 271–285. [https://doi.org/10.1016/0010-0277\(94\)90031-0](https://doi.org/10.1016/0010-0277(94)90031-0)
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46(6), 505–512. <https://doi.org/10.3758/BF03208147>
- Näätänen, R. (1979). Orienting and evoked potentials. In H. Kimmel, E. van Olst, & J. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 61–75). Hillsdale, NJ: Erlbaum.
- Näätänen, R. (1985). Selective attention and stimulus processing: reflections in event-related potentials, magnetoencephalogram, and regional cerebral blood flow. In M. Posner & O. Marin (Eds.), *Attention and performance* (vol. XI, pp. 355–373). Hillsdale, NJ: Erlbaum.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434.
- 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>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of component structure. *Psychophysiology*, 24, 375–425.
- Neuhoff, N., Bruder, J., Bartling, J., Warnke, A., Remschmidt, H., Schulte-Körne, G., & Müller-Meyhok, B. (2012). Evidence for the Late MMN as a Neurophysiological Endophenotype for Dyslexia. *7(5)*, 1–8. <https://doi.org/10.1371/journal.pone.0034909>
- Pastore, R. E., Ahroon, W. A., Buffuto, K. A., Friedman, C. J., Puleo, J. S., & Fink, E. A. (1977). Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 686–696.
- Phillips, C., Pellathy, T., & Marantz, A. (2000a). Phonological Feature Representations in Auditory Cortex. *Unpublished Manuscript*.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., ... Roberts, T. (2000b). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6), 1038–1055. <https://doi.org/10.1162/08989290051137567>
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260. <https://doi.org/10.3758/BF03214136>
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285–290. <https://doi.org/10.3758/BF03213946>
- Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention, 7, 0–8. <https://doi.org/10.1016/j.brainres.2006.01.043>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31(4), 307–314. <https://doi.org/10.3758/BF03202653>
- Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, 32, 55–65.
- Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience*, 17(7), 1168–1180. <https://doi.org/10.1162/0898929054475217>
- Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, 106(2), 1078–83. <https://doi.org/10.1121/1.428048>
- Shestakova, A., Brattico, C. A. E., Huottilainen, M., Galunov, V., Soloviev, A., Sams, M., ... Niitinen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Cognitive Neuroscience and Neuropsychology*, 13(0), 1–5.

- Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research*, 7(2), 215–219. [https://doi.org/10.1016/S0926-6410\(98\)00037-8](https://doi.org/10.1016/S0926-6410(98)00037-8)
- Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, 36(3), 409–414. <https://doi.org/10.1017/S0048577299981180>
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP to deviant stimuli.pdf. *Psychophysiology*, 38, 343–358. <https://doi.org/10.1111/1469-8986.3820343>
- Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. a. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, 82(5), 2346–2357. <https://doi.org/10.1016/j.heares.2009.04.010>
- Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, 55(3), 653–659. <https://doi.org/10.1121/1.1914578>
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, 92(2), 723–735. <https://doi.org/10.1121/1.403997>
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804.
- Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen, R. (1996a). Interactions between Transient and Long-Term Auditory Memory as Reflected by the Mismatch Negativity. *Journal of Cognitive Neuroscience*, 8, 403–415.
- Winkler, I., Czigler, I., Sussman, E., Horváth, J., & Balázs, L. (2005). Preattentive Binding of Auditory and Visual Stimulus Features. *Journal of Cognitive Neuroscience*, 17(March), 320–339. <https://doi.org/10.1162/0898929053124866>
- Winkler, I., Karmos, G., & Näätänen, R. (1996b). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential, 742, 239–252.
- Winkler, I., Kujala, T., Tiitinen, H., Alku, P., Lehtokoski, A., Ilmoniemi, R. J., ... Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638–642.
- Winkler, I., Schröger, E., & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience*, 13(1), 59–71. <https://doi.org/10.1162/089892901564171>
- Wolff, C., & Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters*, 311, 37–40.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.