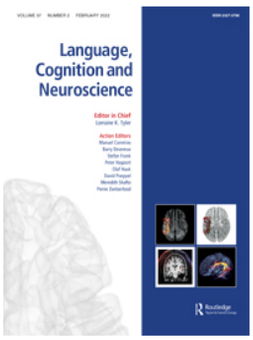

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Auditory predictions are phonological when phonetic information is variable

Ryan Rhodes, Enes Avcu, Chao Han & Arild Hestvik

To cite this article: Ryan Rhodes, Enes Avcu, Chao Han & Arild Hestvik (2022): Auditory predictions are phonological when phonetic information is variable, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2022.2043395

To link to this article: https://doi.org/10.1080/23273798.2022.2043395

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2022.2043395

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Auditory predictions are phonological when phonetic information is variable

Ryan Rhodes [a,c], Enes Avcu[b,c], Chao Han[c] and Arild Hestvik [c]

[a]Rutgers University Center for Cognitive Science (RuCCS), Rutgers University, New Brunswick, NJ, USA; [b]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; [c]Department of Linguistics & Cognitive Science, University of Delaware, Newark, DE, USA

ABSTRACT

An emerging body of studies has claimed that exposure to phonetically varying speech sounds results in strictly phonological representations of speech sounds by the brain's auditory prediction system (Cornell et al., 2011; Eulitz & Lahiri, 2004; Hestvik et al., 2020; Hestvik & Durvasula, 2016; Phillips et al., 2000). We test this claim by measuring mismatch negativity (MMN) to a subcategorical contrast with two sets of phonetically varying standards. When controlling for non-prediction related contributors to the MMN, we find a mismatch in a late time window. However, the mismatch effect is only significant for participants who perceived the contrast as an across-category contrast (/t/ vs /d/). Additionally, we find no modulation of mismatch amplitude predicated on phonetic distance between standards and deviant. Taken together, this indicates that the prediction generated by the auditory system in response to phonetically varying speech sounds is indeed phonological and lacks any fine-grained phonetic content.

ARTICLE HISTORY
Received 20 October 2020
Accepted 9 February 2022

KEYWORDS
Mismatch; prediction; phonetic; phonological

Introduction

To comprehend human speech, listeners have to bridge the gap between continuous spectro-temporal properties of sound and discrete linguistic categories. Under many speech perception models, the brain represents speech sounds as acoustic tokens before speaker normalisation or phoneme identification takes place (Kingston & Diehl, 1994; Liberman et al., 1967; Theodore et al., 2009), and there is evidence that speech sounds are neurally encoded with great acoustic detail (Andruski et al., 1994; Bidelman et al., 2013; Toscano et al., 2010). However, sounds are also represented via language-specific distinctive features (Monahan, 2018), and such features would be the basic targets for phonological processes (Chomsky & Halle, 1968; Kenstowicz, 1994). Mediating between these levels is a phonetic level which groups sounds into language-specific categories. A key feature of this level of representation is that phonetic categories – unlike phonological categories – contain gradient, within-category information (Pierrehumbert, 1990; Werker & Logan, 1985).

An emerging body of studies has claimed that phonetic variance precludes phonetic prediction – that exposure to phonetically varying speech sounds results in strictly phonological neural representations of speech sounds (Cornell et al., 2011; Eulitz & Lahiri, 2004; Hestvik et al., 2020; Hestvik & Durvasula, 2016; Phillips et al., 2000). We argue that it has not been sufficiently established that the observed brain responses in these cases is the result of wholly phonemic categorical predictions which lack phonetic content. In this study we measure the mismatch negativity (MMN, an ERP component indicating a prediction error-related brain response) to a nominally sub-phonemic contrast with phonetically varying standards. We observe an MMN to this contrast, but analysis reveals the mismatch may be primarily driven by individual differences in voicing category threshold. In other words, only participants who perceived the contrast as across-category showed a significant MMN effect. We take this as further evidence of a phonologically driven prediction error signal in response to phonetically variable input.

Background

There is a large body of literature using electrophysiological measures to record brain responses to speech sounds. Many studies have used an oddball paradigm to elicit mismatch negativity (MMN), an event-related potential first described by Näätänen et al. (1978) representing an automatic response to any detected change in auditory stimulation generated in the auditory cortex (Alho, 1995; Näätänen et al., 2007).

The MMN has been argued to index prediction error resulting from the mismatch between predicted stimuli and deviant stimuli (Carbajal & Malmierca, 2018; Font-

CONTACT Ryan Rhodes ryan.rhodes@rutgers.edu
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2022.2043395.

Alaminos et al., 2021; Garrido et al., 2008, 2009; Lieder et al., 2013; Todorovic et al., 2011; Winkler & Czigler, 2012). Under the predictive coding approach (Friston, 2005, 2010), the auditory system is a Bayesian prediction generator built atop a representational system in which speech sounds are represented at several hierarchically organised levels (Dürschmid et al., 2016; Escera & Malmierca, 2014; Rubin et al., 2016; for a review, see Heilbron & Chait, 2018). Each of these levels maintains different types of information simultaneously. Lower levels of the hierarchy contain acoustic sensory information, while higher levels contain more abstract information such as phonological features or phoneme category labels. Speech sound representations are generated by a hierarchical, generative model that is used to make predictions about the auditory environment and is continually updated from new sensory input.

Several studies using MMN have shown evidence of predictive coding (Bendixen et al., 2012; Chennu et al., 2013; Grimm & Escera, 2012; Wacongne et al., 2012), indicating that the MMN is at least partially driven by a hierarchically organised prediction system. Predictions occur at several structural levels: non-linguistic acoustic contrasts, language-specific phonetic contrasts, phoneme category contrasts, and even higher units of structure like lexical categories (Pulvermüller & Shtyrov, 2006). This makes the MMN a useful tool for probing the content of auditory predictions and the forward-propagating prediction error signal.

When faced with speech sounds, the auditory system will generate predictions at multiple levels. Acoustic-phonetic predictions will include details about properties such as formant frequency, duration, and voice onset time (VOT). Phonological predictions will include only phoneme category labels or phonological features. Ordinarily, the predictive system interfaces with statistical learning mechanisms to track the properties of sensory input and generate detailed predictions, resulting in predictions which synthesise the properties of the input in a dynamic way, weighting different stimuli and features according to the frequency of their appearance (McMurray et al., 2009; Paavilainen et al., 2001; Saffran et al., 1999).

However, prior knowledge of linguistic categories may interfere with this general statistical learning process, evidenced by sensitivity to speech sound categories (Barrios et al., 2016; Dehaene-Lambertz, 1997; Kazanina et al., 2006; Näätänen et al., 1997; Rodrigues Silva et al., 2017; Sharma & Dorman, 2000; Yu et al., 2017). Rather than synthesising all the properties of the input to generate the most detailed possible prediction, the auditory system may instead rely exclusively on linguistic categories from long term memory to make predictions about incoming sounds, particularly under conditions of phonetic variance (Phillips et al., 2000). A growing body of studies has used a "varying standards" paradigm in which acoustically and phonetically varying stimuli are used to elicit MMN. These studies have used standards recorded by different speakers (Dehaene-Lambertz & Pena, 2001; Shestakova et al., 2002), by a single speaker recording different tokens (Schluter et al., 2016), by manipulations in formant frequency (Hill et al., 2004; Jacobsen et al., 2004), and by manipulations in VOT (Hestvik et al., 2020; Hestvik & Durvasula, 2016; Phillips et al., 2000).

Phillips et al. (2000) demonstrated that the auditory cortex accesses phoneme category knowledge by comparing a phonological condition, and an acoustic condition. Their phonological condition consisted of phonetically varying standard stimuli belonging to the /t/ category and a deviant stimulus belonging to the /d/ category, making the standard-deviant contrast a categorical /t/-/d/ contrast. Their acoustic condition shifted the voice onset time of the standard stimuli so that the median VOT of the standards would be equal to the boundary value separating voiced from voiceless. This resulted in half the standards belonging to the /t/ category and half belonging to the /d/ category. The standard-deviant contrast was no longer categorical – the difference could only be detected if the phonetic properties of the stimuli were accurately represented in the prediction. Phillips et al. found a significant MMN in their phonological condition but no corresponding MMN in their acoustic condition, which they interpreted as evidence that the auditory cortex has access to abstract phoneme category representations. The crucial finding, which has informed many subsequent MMN studies, is that the auditory system seems to have been incapable of generating an *ad hoc* phonetic category to characterise the difference between standards and deviants in the acoustic condition.

The standards in Phillips et al.'s acoustic condition varied *across* the /d/-/t/ category boundary, with stimuli falling both on the /d/ side and on the /t/ side of the boundary (adjusted for each participant). Phoneme category knowledge seems to provide a strong enough bias to the perceptual system to prevent the formation of an *ad hoc* phonetic prediction when the phonetic values run afoul of the category identity. In other words, phonetic predictions may still be possible when the variance is within-category. Mah et al. (2016), investigating native and non-native contrasts with a varying standards paradigm, found a phonetic prediction when stimuli were simple fricative noise bursts – an effect which disappeared when the

stimuli were full syllables. This suggests that speakers' linguistic knowledge plays a large role in auditory prediction mechanisms, even affecting predictions maintained at lower acoustic-phonetic levels of the predictive hierarchy. And this effect of linguistic knowledge seems to eclipse any effects induced by phonetic variance.

An emerging body of work using the varying standards paradigm has relied on the assumption that the introduction of acoustic-phonetic variance precludes predictions containing phonetic information. Several studies have used the varying standards paradigm to provide evidence for phonological theories such as the Featurally Underspecified Lexicon (Cummings et al., 2017; Eulitz & Lahiri, 2004; Hestvik et al., 2020; Hestvik & Durvasula, 2016; Scharinger et al., 2010; Schluter et al., 2016). The logic of these experiments relies on the absence of phonetic information in the predictions generated in response to phonetically varying standards. But evidence contrary to this operational assumption have emerged, with some studies finding underspecification effects using single standard paradigms (Scharinger, Bendixen, et al., 2012; Walter & Hacquard, 2004). In addition, at least one study using the varying standards paradigm has found a sub-phonemic mismatch effect – which should only be possible if the varying standards and subsequent prediction are represented by the auditory system with phonetic detail (Miglietta et al., 2013). These results imply that phonetic variance is not either necessary or sufficient to "erase" phonetic content from the prediction. Given these contrary findings, we question whether phonetically detailed predictions in response to phonetically varying speech sound input are in fact possible – and more importantly, what particular acoustic-phonetic information is being fed forward as prediction error. In other words, how exactly does the prediction error signal correspond to phonetically variable linguistic input?

It has been observed that the amplitude of the MMN increases as the magnitude of the auditory change increases (Alain et al., 1999; Amenedo & Escera, 2000; Berti et al., 2004; Joanisse et al., 2007; Näätänen et al., 2007; Pakarinen et al., 2007; Rinne et al., 2006; Schröger, 1995; Tiitinen et al., 1994; Wolff & Schröger, 2001). Rhodes et al. (2019) elicited MMN in response to a deviant /d/ (15 ms) which appeared infrequently in two conditions of varying standards, comparing standards varying in a "high" (75-85 ms) versus a "low" (60-70 ms) range of VOT to observe an effect of deviance magnitude in the MMN. They predicted a "phonetic distance effect" in the MMN – a positive correlation between MMN amplitude and phonetic distance between standards and deviant – which would indicate a prediction containing phonetic information relating to the standards. While they observed a significant mismatch response, there was no significant difference between conditions, suggesting that this fine-grained phonetic information was absent from the prediction generated in response to the phonetically varying standards. This accords with previous studies using the varying standards paradigm to elicit phonological representations.

However, since Rhodes et al. used an across-category contrast (/d/ vs /t/), there is the possibility that a relatively small phonetic effect may have been obscured by a larger category effect. Both phonetic and phoneme-category differences contribute to MMN amplitude, but across-category differences generate significantly greater MMN amplitudes than within-category differences, even when phonetic distance is held constant (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Rodrigues Silva et al., 2017; Sharma & Dorman, 1999; Winkler, Kujala, et al., 1999a; Winkler, Lehtokoski, et al. 1999b). For this reason, the current study replicates the design of Rhodes et al. (2019) with the modification that all contrasts are within-category (/t/ vs /t/).

Methods

Participants

32 participants were recruited (2 male) from the University of Delaware. Participant ages ranged from 18–27 with a mean age of 20 (SD = 2). 1 participant was left-handed. All participants were students at the University of Delaware, native speakers of English, and reported no history of language or hearing impairments. Participants were either paid $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.

Four participants' data were excluded from analysis due to the raw recorded EEG data having greater than 10% bad channels; two participants' data were excluded due to having fewer good trials than two standard deviations from average; and three[1] participants' data were excluded because their pre- or post-test categorisation thresholds were above 100 ms (indicating they perceived all stimuli as /d/). After applying these criteria, 21 participants' data were retained for analysis.

After exclusion, the average number of total good trials per participant was 2,797.9 (of a possible 3,000). There were six cells in the experimental design: High Standard (845.2 average good trials per participant), High Deviant (93.4), Low Standard (855.8), Low Deviant (94.8), Control Standard (818), and Control Deviant[2]

(90.6). Because standard and deviant stimuli appeared in a 9:1 ratio, there were 900 total possible trials in the standard cells and 100 total possible trials in the deviant cells.

Study design

The current study consists of two experimental conditions and a control condition. Each experimental condition contains three standards with VOT values in different range bands. In both experimental conditions, the standards will be tokens of /t/ with varying VOT values and the deviant will be /t/ with a VOT of 50 ms. In the Low Standards condition the standards range from 95–105 ms in 5 ms increments ({95, 100, 105 ms}, mean: 100 ms). In the High Standards condition the standards range from 110–120 ms ({110, 115, 120 ms}, mean: 115 ms). The mean difference between standard and deviant is 50 ms in the Low condition and 65 ms in the High condition, with a 15 ms difference between the two conditions. Single-subject data from Carney et al. (1977) suggests listeners can behaviourally discriminate VOT differences at least as small as 10 ms in high ranges (in their study, centred around 80 ms VOT). The Random Standards Control condition consists of syllables with VOT values ranging from 50–95 ms in 5 ms increments {50, 55, 60, 65, 70, 75, 80, 85, 90, 95 ms}. All stimuli appear with equal frequency in the Control condition (10% for each stimulus). See Figure 1 for an illustration of the stimuli used for each condition.

Because the standard-deviant contrast is within-category for both experimental conditions, an MMN is only obtainable if the prediction contains phonetic information. Standards and deviant both belong to the /t/ category, meaning the phoneme category label alone cannot differentiate standard from deviant. Successful discrimination requires a comparison of the VOT values of the standards and deviant. If the standard sounds are represented at the acoustic-phonetic level, we will see a significant MMN in both conditions (high and low). Additionally, we should see a difference in MMN amplitude between the conditions due to the phonetic distance between standards and deviant. If the standard sounds are only represented at the phoneme level, there should be no MMN – because all stimuli (standards and deviants) belong to the same category /t/.

In addition to prediction mechanisms, the MMN may also be partially driven by neuronal refractoriness (Auksztulewicz & Friston, 2016; Escera & Malmierca, 2014; May & Tiitinen, 2010; Ulanovsky et al., 2003) (cf. Ruusuvirta, 2021) and gradient perceptual encoding (Frye et al., 2007; Näätänen, 1992; Toscano et al., 2010). When MMN is used as an index of prediction error, these effects can introduce confounds. For this reason, we use a "random standards" control paradigm (Jacobsen & Schröger, 2001, 2003; Schröger & Wolff, 1996; see also Carbajal & Malmierca, 2018; Font-Alaminos et al., 2021) to control for both perceptual encoding and neuronal refractoriness. The random standards control paradigm introduces a control block in which

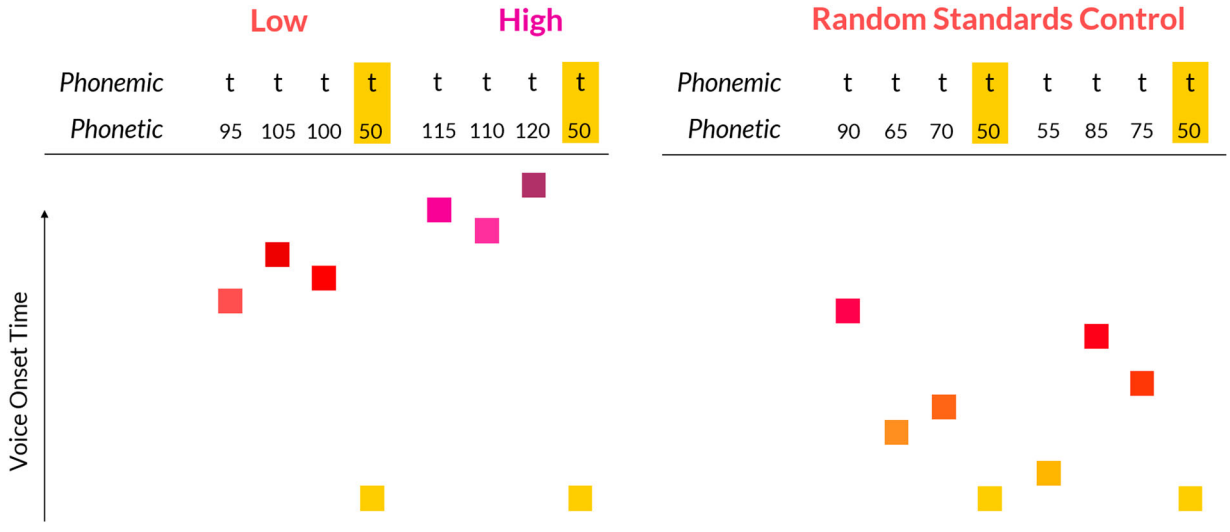

Figure 1. Illustration of the experimental and control conditions. The standards for the "Low" condition vary between 95 and 105 ms VOT (shown in shades of red). The standards for the "High" condition vary between 110 and 120 ms VOT (shown in shades of purple). The highlighted stimuli represent the infrequent deviants (shown in yellow). The standards and deviants appear in a 9:1 ratio. The highlighted (yellow) stimuli here are phonetically identical to and appear with the same frequency as the deviant stimuli in the experimental conditions. The other stimuli range from 55–95 ms VOT in 5 ms increments. All stimuli in the control condition appear with equal frequency (10%).

the deviant stimulus from the experimental block appears among a set of randomly varying sounds that all appear with equal frequency. Since all stimuli in this condition appear with equal frequency, it is assumed that no prediction can be made on the basis of frequency-based groupings. And because the deviant stimulus appears with the same frequency in both control and experimental blocks, the effect of neuronal refractoriness is controlled. Finally, because we are comparing the deviant stimulus to an identical stimulus in the control condition, the effect of gradient encoding is controlled.

Materials

All stimuli were synthesised using Klatt synthesiser. Stimuli were adapted from a previous study (Hestvik & Durvasula, 2016) and were originally created to match the stimuli used by Philips et al. (2000). Hestvik & Durvasula used stimuli with VOT ranging from 0–100 ms in 5 ms steps. In addition to these stimuli, we generated four new stimuli using the same Klatt synthesiser script for VOT values 105, 110, 115, and 120 ms. See Figure 2 for a spectrogram and waveform of sample stimuli.

Each synthesised syllable had a duration of 300 ms. Because the voice onset time varied across the stimuli (95-105 ms in the "low" condition, 110–120 ms in the "high" condition, 50–95 ms in the control condition), the vowel of each stimulus varied in duration (the duration of the vowel for each stimulus was simply the difference between the total duration of 300 ms and the duration of the voice onset time). The script used to generate the stimuli can be found in Appendix A.

Procedure

Participants performed an offline categorisation task both before and after the EEG session. This was to ensure both that participants perceived all stimuli as belonging to the same category (/t/), and exposure to the very high VOT values would not shift their perceptual boundary between voiced and voiceless.

Studies have shown that repeated exposure to a stimulus can cause adaptation, increasing the probability that a subsequent stimulus will be perceived as belonging to the opposite category (Eimas & Corbit, 1973). This is particularly true when the repeated stimulus is near the category boundary. Adaptation is also particularly effective if the repeated stimulus is close to the listener's prototype value (Samuel, 1982). Given the very high range of VOT of the standards in this experiment, it is unlikely that any of the standards are near any particular listener's prototype. The mean VOT for a typical /t/ in English is around 70 ms, with a category boundary around 25 ms (estimated from Lisker & Abramson, 1964). However, there is broad variability in production of VOT among English speakers (Allen et al., 2003). Given all these facts, it is impossible to know a priori what kind of impact repeated exposure to these stimuli would have on perception for each subject. Measuring their categorisation function both before

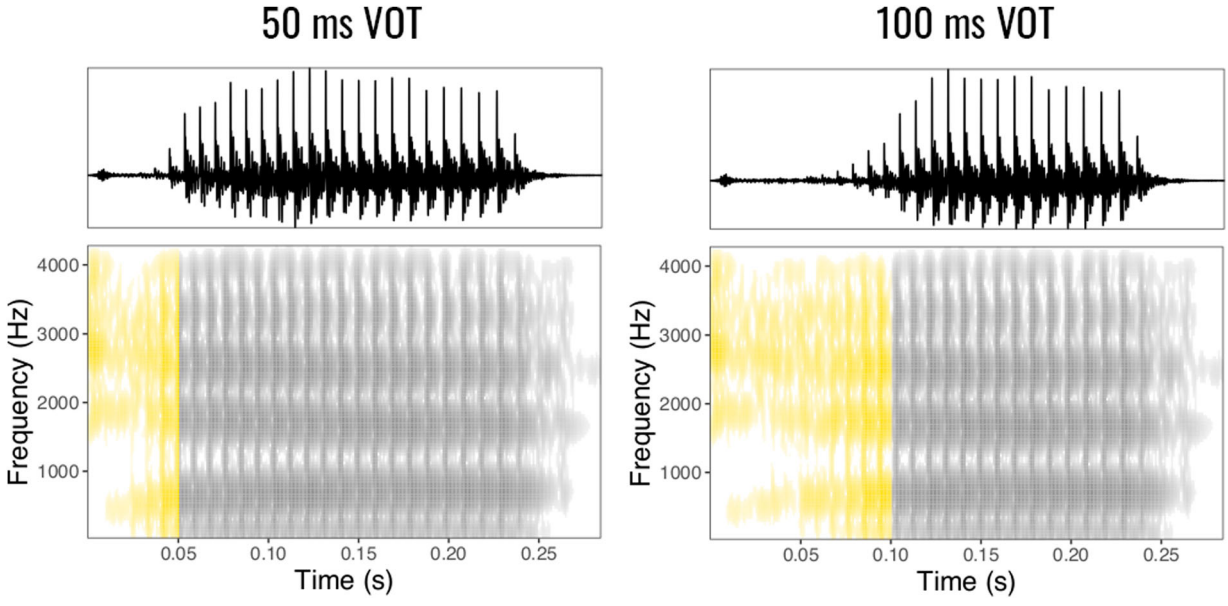

Figure 2. Waveforms and spectrograms of two sample stimuli: 50 ms VOT deviant (left), and 100 ms VOT standard (right). Lag time (VOT) is highlighted in yellow.

and after the EEG session should determine whether their perceptual boundary has been significantly shifted.

For the categorisation task, participants were presented with CV stimuli with VOT values ranging from 0 to 100 ms and instructed to press a button to categorise each sound as either a /t/ or /d/. Stimulus presentation was randomised. The categorisation task consisted of 6 blocks of 21 trials each, for a total of 126 trials. After each block, the participant was given a brief break. Every other block had reversed key assignments (e.g. if button 1 corresponded to /t/ and button 4 to /d/, in the next block button 1 would correspond to /d/ and button 4 to /d/). Key assignments were present on screen during every trial. The categorisation task took about 10 min to complete.

After the conclusion of the categorisation task, the participant was outfitted with the electrode net and sat in the booth for the EEG session. After the conclusion of the EEG session, the participant completed the same categorisation task a second time. During the EEG recording session, participants were seated in a sound-attenuating booth in a chair facing a computer monitor. Stimuli were presented via two free-field speakers. The order of the two experimental blocks and control block was randomised to control for block order effects, including distributional learning or adaptation (see e.g. Frost, Winkler, Provost, & Todd, 2016; Todd et al., 2014). The interstimulus interval (ISI) in each condition randomly varied from 410 to 600 ms to avoid phase-locked brain responses. The total recording time was about one hour (approximately 20 min per block).

Continuous EEG was recorded passively with no task or directed attention. Participants were instructed to watch a movie with no sound and no subtitles (*Finding Nemo*). Participants were not instructed to pay attention to the sounds. They were told the experiment required nothing from them – there was no task, and they did not need to pay attention to or remember any details of what they would see or hear. Participants were instructed only to sit still and stay awake. The movie was offered to help participants stay awake and prevent boredom. Participants were given regular breaks between blocks (each block was about 18 min long) during which they could stand, stretch, eat a chocolate, or take a drink of water. This was done to ensure their comfort and keep them awake.

Data acquisition and pre-processing

Continuous EEG was recorded from 128 carbon fibre core/silver-coated electrodes in an elastic net (Hydrocel Geodesic Sensor Net 128, a high-impedance saline-based EEG system). Before data acquisition, electrode impedances were lowered to below 50 kΩ. Participants' electro-ocular activity was measured from 2 bipolar channels at the supraorbital and infraorbital electrodes and the outer canthi electrodes of each eye. The continuous EEG was recorded with an analog on-line 100 Hz low-pass filter and digitised with a sampling rate of 250 Hz.

After acquisition, the raw EEG data were run through a 0.1 Hz high-pass filter before segmentation to remove slow drifts (within the limit recommended by Tanner et al., 2015). The continuous EEG was segmented into 800 ms epochs with a 200 ms pre-stimulus period. Segmented data underwent artifact correction, bad channel replacement, and baseline correction using the ERP PCA toolbox (Dien, 2010). Segmented data were baseline corrected to the mean voltage of the 200 ms pre-stimulus period. Subsequent processing removed eyeblink and eye movement artifacts using EEGLAB's Independent Component Analysis. 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 a manually created eyeblink template. After eyeblink subtraction, the data were submitted to the artifact correction procedure for bad channels. A channel was marked bad if its best absolute correlation with its neighbouring channels fell below .4 across all time points. Bad channels were replaced via interpolation from neighbouring good channels. If a channel was marked bad in over 20% of trials, it was considered bad in all trials. A trial was marked bad if it contained more than 10% bad channels. Bad trials were dropped from the analysis.

The remaining trials were averaged into six cells: Control Standard, Control Deviant, High Standard, High Deviant, Low Standard, and Low Deviant. The averaged data were then re-referenced to the linked mastoids. A 40 Hz low-pass filter was applied as a final step.

Results

Behavioural results

A threshold analysis was applied to the categorisation data to determine the perceptual boundary between the voiced (/d/) and voiceless (/t/) categories for each participant. For each participant, we built a logit regression model to estimate the log odds of a /t/ response (vs. a /d/ response) for each VOT value. Trials with no responses did not enter the regression. We then calculated the VOT value at which the participant would respond with exactly 50% probability (where the log odds = 0).

A paired samples t-test comparing the threshold values before and after the EEG recording session

show no significant difference ($p > 0.05$; 1-β = 0.075359). The mean threshold value for the pre-EEG task was 50.7 ms (SD = 13.1), and the mean threshold value after the EEG task was 52.5 ms (SD = 16.6). These values are higher than the commonly-cited perceptual threshold for stop voicing in English – with typical VOT values cited between 25 ms (Lisker & Abramson, 1964) and 40 ms (Hestvik & Durvasula, 2016). Perceptual category boundaries can be affected by the range of VOT values used in the task (Brady & Darwin, 1978; Rosen, 1979), which could explain our higher than typical threshold scores. Our stimuli and design for the behavioural categorisation task were identical to those used by Hestvik & Durvasula, who reported a mean threshold score of 40 ms (SD = 5 ms).

To test for the possibility of distributional learning or adaptation effects, a one-way ANOVA was conducted to determine whether there was an effect of Block Order on the threshold scores. This one-way ANOVA found no significant effect of Block Order on threshold scores ($p > .05$). This indicates that exposure to the high VOT stimuli during the EEG recording session either did not induce any learning or adaptation effects, or that these effects did not persist long enough after the session to be measured. Figure 3 shows the categorisation results and histograms of pre- and post-test threshold scores across participants.

There are other possible explanations for these higher-than-expected perceptual thresholds. Listeners can use a variety of phonetic cues to identify stop voicing contrasts, including *F*0 (Haggard et al., 1970; Ohde, 1984; Soli, 1983; Winn et al., 2013) and centre of gravity (Forrest et al., 1988). It is possible a subset of our participants relied on cues other than VOT to identify the voicing of the synthesised stop consonants, which could have introduced variance into the behavioural responses. 6 participants showed perceptual threshold scores greater than 100 ms (meaning they selected /d/

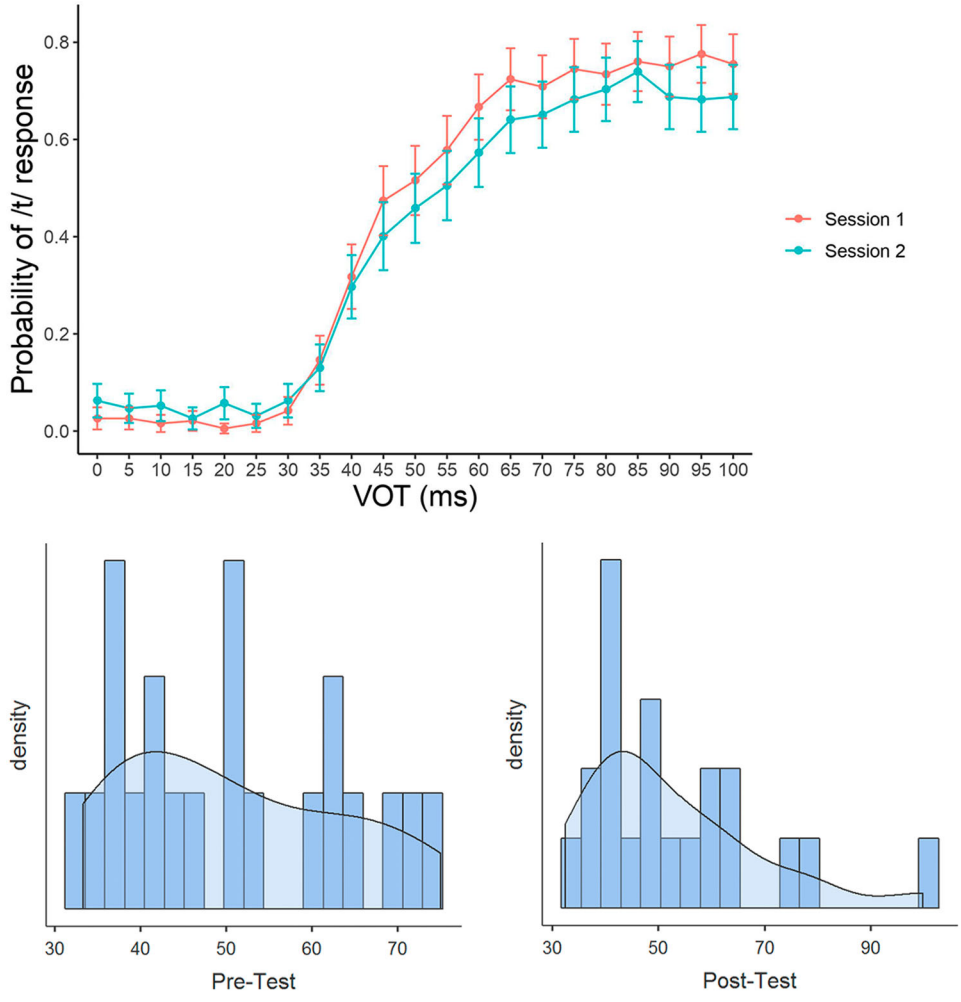

Figure 3. Top: categorisation data showing the proportion of /t/ responses plotted against VOT. Session 1 (pre-test) was recorded before the EEG session. Session 2 (post-test) was recorded after the EEG session. Bottom: Histograms showing distribution of threshold scores for the categorisation pre-test (left) and post-test (right). Voice onset times (VOTs) are shown on the x-axis.

on almost every trial, regardless of VOT). These participants were excluded from further analysis.

Eeg results

We present an analysis comparing the deviants from the experimental conditions to the same stimulus in the random standards control condition (as suggested by Jacobsen & Schröger, 2001, 2003; and Schröger & Wolff, 1996). This analysis is intended to control for gradient perceptual encoding effects and neuronal refractoriness and isolate the prediction error related component of the MMN. The electrode Fz was selected for analysis *a priori* from previous studies. Time windows for the analysis were selected via visual inspection of the averaged difference wave (computed as the average of the high-control and low-control difference waves) to capture the clearest mismatch effects. For conventional deviant-standards waveforms and topoplots[3], see Figures 4 and 5.

For this analysis, we compared the 50 ms VOT deviants in the experimental conditions directly with the same stimulus in the control condition (see Figure 9). These differences were analysed with a repeated measures ANOVA with the factor Condition (High, Low, Control) and the between subject factor Threshold (perceptual threshold score, binned into >50 ms and <50 ms levels). We conducted additional planned comparison paired samples t-tests to determine if the brain response to the 50 ms VOT deviants in each experimental condition significantly differed from the brain response to the same stimulus in the control condition, as well as whether the deviants in the High and Low conditions differed from each other.

When comparing the deviants in the experimental conditions to the same stimulus in the control condition, we find a clear mismatch effect in a very late time window (peaking ~500 ms). A repeated measures ANOVA with the factor Condition (High, Low, Control) and between subject factor Threshold (>50 ms, <50 ms) found a significant main effect of Condition ($F(2,38) = 3.34$, $p < .05$) but no significant Condition*Threshold interaction ($F(2,38) = 2.18$, $p > .05$).

Planned comparison paired samples t-tests found a significant difference between High and Control ($t(19)$ 3.034; $p < .05$; $1\text{-}\beta = 0.738$) and a significant difference between Low and Control ($t(19)$ 2.145; $p = 0.045$; $1\text{-}\beta = 0.765$). There was no significant difference between the deviants in the High and Low conditions ($t(19)$ 0.205; $p > 0.05$).

Paired samples t-tests were also conducted separately for the <50 ms threshold group and >50 ms threshold group to determine whether these significant effects were due to differences in perceptual category boundaries. T-tests for the >50 ms threshold group found significant differences between High and Control ($t(8)$ 3.918; $p = 0.002$; $1\text{-}\beta = 0.999$; BF10 = 24.30) and a significant difference between Low and Control ($t(8)$ 2.375; $p = 0.022$; $1\text{-}\beta = 0.920$; BF10 = 3.83), but not with High and Low ($t(8) = 0.596$; $p > 0.05$; $1\text{-}\beta = 0.181$; BF10 = 0.223). *t*-tests for the <50 ms threshold group found no significant differences between deviants in the experimental conditions and the same stimulus in the control condition (High and Control ($t(11)$ 0.282; $p = 0.392$;; BF10 = 0.36) and Low and Control ($t(11)$ 0.584; $p = 0.285$;; BF10 = 0.47)), and no difference between experimental conditions (High and Low ($t(11)$ 0.278; $p = 0.393$;; BF10 = 0.36)) Figures 6 and 7.

A central aim of this study was to determine whether phonetic distance between standard and deviant would affect MMN amplitude (i.e. prediction error). To determine the probability of a true null effect (no difference between High and Low), a Bayes factor was computed for each paired samples t-test, using a default prior of 0.707 (Rouder et al., 2012). This analysis found a BF01 = 4.37 for all participants when comparing the deviants in the High and Low conditions, meaning the null hypothesis is four times more likely to be true than the alternative. This is in line with previous findings of BF01 = 4.54 in a previous experiment with the same phonetic distance between conditions (Rhodes et al., 2019). These results suggest that the amplitude of the MMN response is not significantly affected by the phonetic distance between standards and deviant.

Discussion

This study was designed to test whether auditory predictions generated in response to phonetically varying stimuli would contain phonetic information – or whether the auditory prediction system would instead rely solely on linguistic category knowledge from long term memory. We used a random standards control paradigm to control for effects related to neuronal refractoriness and gradient perceptual encoding, and to isolate a true prediction error brain response. We computed MMN difference waves by subtracting the brain response to the deviant in the experimental conditions minus the brain response to the same stimulus in the control condition.

We find a significant mismatch effect in a late time window (peaking ~500 ms). Our hypothesis (contra the prevailing assumption about the varying standards paradigm) was that varying standards would not "erase" phonetic information during auditory prediction – that is, we expected a mismatch to the sub-phonemic contrast presented in the study. The interpretation of this mismatch

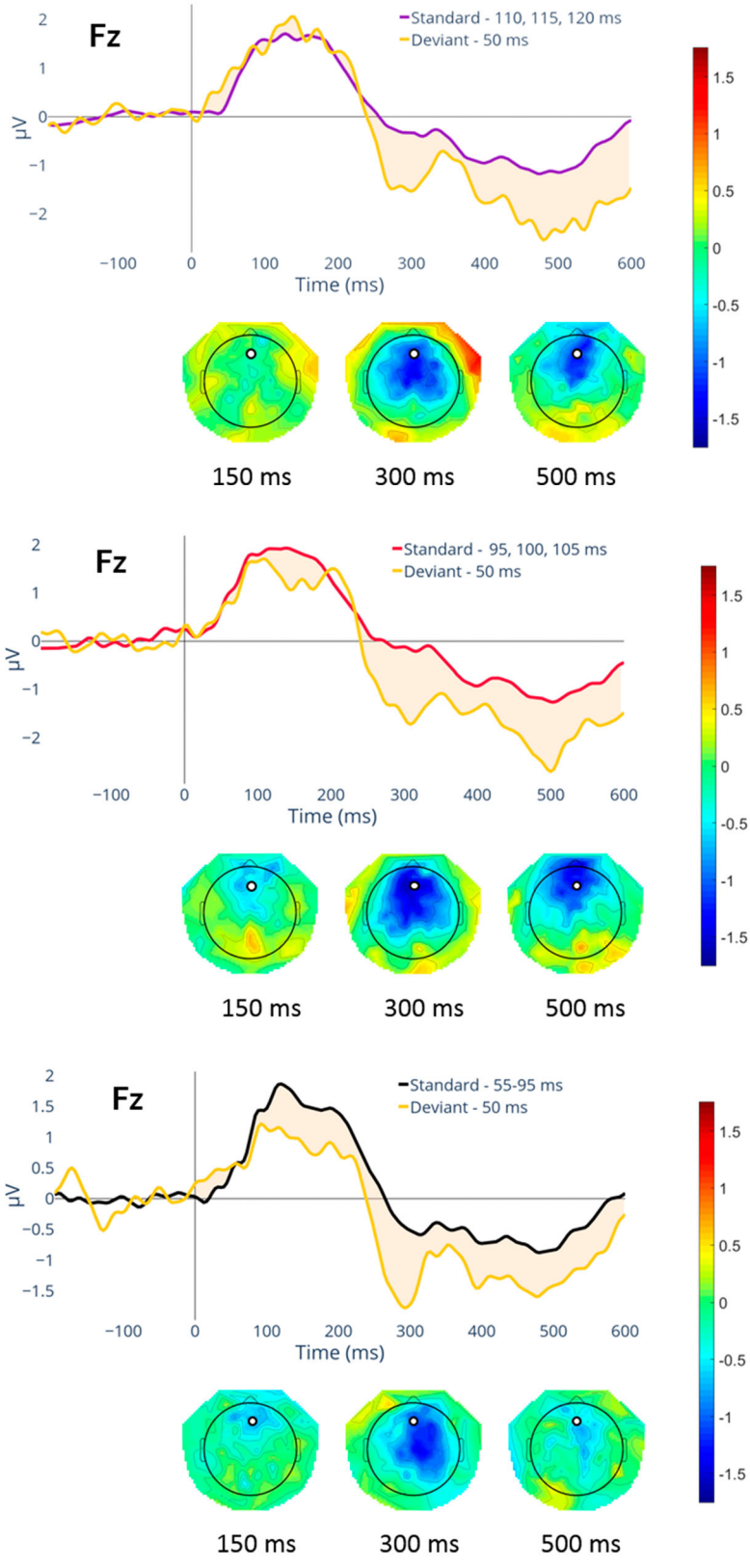

Figure 4. Standard vs deviant waveforms (at Fz) and topoplots. The High condition is shown on the top, the Low condition in the middle, and the Random Standards control condition on the bottom. The standard wave is an average of the brain response to the set of varying standards (for the control condition, the standard wave is the average brain response to all stimuli except the 50 ms VOT stimulus). The topoplots are computed from the deviant-minus-standard difference wave at 150, 300, and 500 ms.

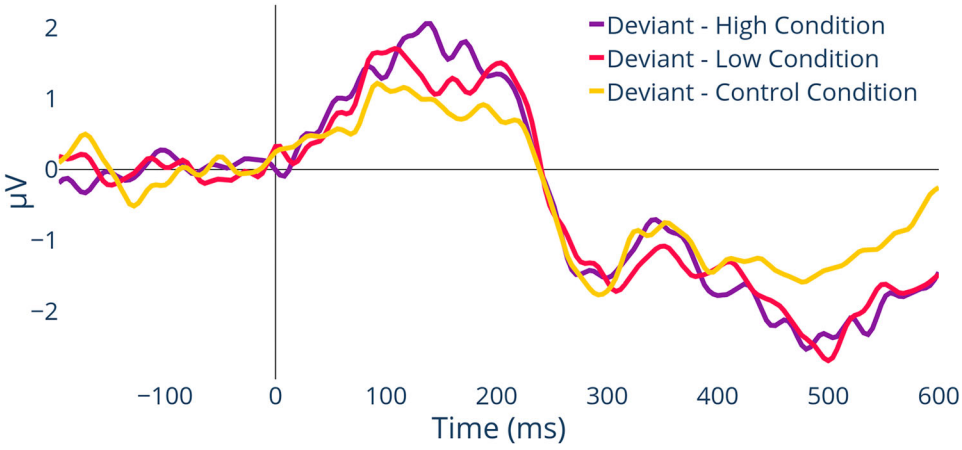

Figure 5. Comparison of the 50 ms VOT deviant waves in each condition. Each wave represents the brain response to a phonetically identical token, appearing with identical frequency (10%) in each condition (High, Low, Control). The two experimental conditions (High, Low) do not differ from the Control in the N2 time window (∼300 ms). However, they do differ from control in the later time window (∼500 ms).

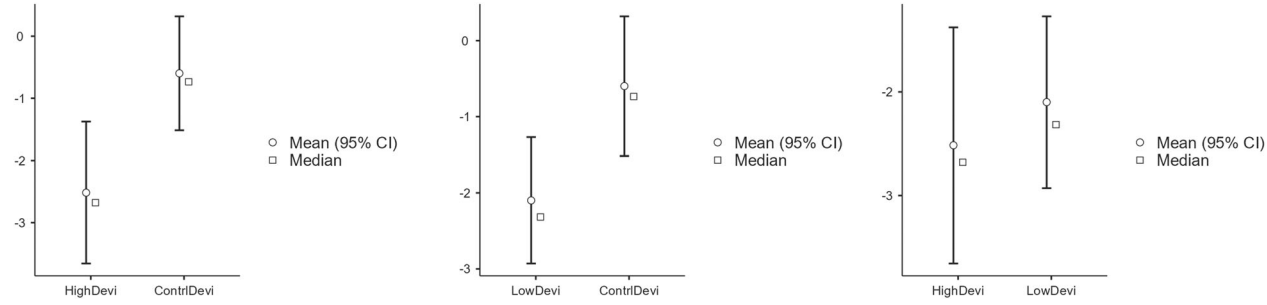

Figure 6. Plots of t-tests for the >50 ms group. Left panel: high deviant vs control; middle panel: low deviant vs control; right panel: high deviant vs low deviant.

depends on the participants' having perceived the presented contrasts as within-category (/t/ vs /t/). However, while we do not find a significant interaction between MMN effect and participant threshold scores (measured by an offline phoneme categorisation task), we do see a difference between two (post-hoc) participant groups. Participants with threshold scores *below* 50 ms were likely to have categorised the 50 ms deviant stimulus as /t/. For these participants, the experimental contrast would have been perceived as a within-category contrast (as intended by the experimental design). Participants with threshold scores *above* 50 ms were likely to categorise the 50 ms deviant stimulus as /d/ – meaning that for these participants the standard-deviant contrast in the experimental conditions was likely to have been perceived as an across-category contrast. The fact that we see a highly significant mismatch for the >50 ms group and no significant mismatch

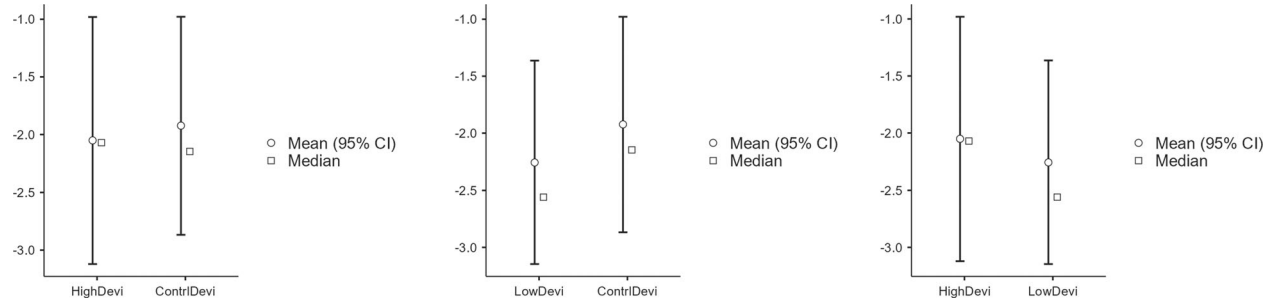

Figure 7. Plots of t-tests for the <50 ms group. Left panel: high deviant vs control; middle panel: low deviant vs control; right panel: high deviant vs low deviant.

for the <50 ms group suggests that the MMN effect observed here is driven at least in part by phonological category identity.

We also predicted a modulation of the MMN by phonetic distance – deviants which are phonetically farther removed from the standards should elicit a higher amplitude mismatch effect. However, we find no significant difference between deviants in the High and Low conditions. The fact that we only see significant mismatch effects for participants who perceived the contrast as across-category (but not for participants who perceived the contrast as within-category), coupled with the fact that mismatch amplitude is not modulated by phonetic distance between standards and deviant for either group of participants, strongly suggests that the observed mismatch is predicated on phonological category identity rather than phonetic content.

This phonological category-only mismatch is exactly what previous studies have assumed as a consequence of the varying standards paradigm (e.g. Cornell et al., 2011; Cummings et al., 2017; Eulitz & Lahiri, 2004; Hestvik et al., 2020; Hestvik & Durvasula, 2016; Scharinger et al., 2010, 2012; Schluter et al., 2016). Several other studies have found sensitivity in the MMN to sub-phonemic contrasts (e.g. Joanisse et al., 2007; Miglietta et al., 2013), but these studies did not include the same paradigms or controls used in our experiment.

The time course and spatial distribution of our observed mismatch effect (452-552 ms) is consistent with the so-called Late Discriminative Negativity (LDN), which is most often reported in infants (Cheour et al., 1998; Martynova et al., 2003) and children (Datta et al., 2010; Halliday et al., 2014; Hong et al., 2018; Korpilahti et al., 2001; Neuhoff et al., 2012; Shafer et al., 2005; Strotseva-Feinschmidt et al., 2015). However, the effect has also been observed in adults (Hestvik & Durvasula, 2016; Hill et al., 2004; Korpilahti et al., 1995; Zachau et al., 2005). Late mismatch responses (>200 ms) have been argued to index top-down processing (Garrido et al., 2007; Schiff et al., 2006). Hill et al. (2004) find late mismatch effects in a phonological but not an acoustic condition and argue that this late effect corresponds to phonological categorisation.

Zachau et al. (2005) observe an LDN to an abstract tone pattern and speculate that this may reflect auditory rule extraction and transfer of rules to long-term memory. Korpilahti et al. (1995, 2001) find a larger LDN in response to words than tones, suggesting the LDN might correspond to processing complex or linguistic stimuli. Hestvik & Durvasula (2016) likewise find an LDN in response to linguistic stimuli in a paradigm very similar to the current study.

However, Rodrigues Silva & Rothe-Neves (2020) report contrary findings. In an MMN study looking for evidence of positional neutralisation, they find an LDN which is sensitive to a sub-phonemic contrast – even when the MMN is not. Miglietta et al. (2013) also report a late mismatch effect to a sub-phonemic (allophonic) contrast. Winkler et al. (1999) suggest that auditory change detection might be faster for categorical rather than acoustic differences, a view that is shared by van Hessen & Shouten (1992), who argue that it takes time to create a detailed representation and to compare the representation to a category prototype to determine its phonetic distance. Most of these studies examined vowel contrasts, which has been shown to be less categorical than consonant contrasts (Fry et al., 1962; Kronrod et al., 2016; Pisoni, 1973). Still, these findings are at odds with the late phonemic discrimination effect we observe here, which raises some questions about the time course of auditory processing and speech sound identification.

Taken together, it seems the LDN is strongly implicated in speech perception and the processing of linguistic stimuli. Although it has been argued to index phoneme categorisation processes – which seems to be the case for our results as well – the findings of Rodrigues Silva & Rothe-Neves among others indicate that it must also be sensitive to sub-phonemic differences. Future work will be needed to investigate the LDN in more depth to resolve this apparent contradiction.

Although we observe a genuine prediction error response, we still do not see any significant difference between the High and Low conditions. There are several potential explanations for this absence of effect. First, it is possible that the difference is simply too small to be visible in the MMN. Ideally, a larger difference between conditions would have been used to elicit MMNs which may then show a phonetic distance effect. However, there is the practical concern of perceptual space – each phoneme category only has finite space for the phonetic contrasts to occupy before the stimuli cross a perceptual boundary or become extremely unnatural.

Additionally, the mean distance between conditions (the difference between the mean VOT of the standards in each condition) is far greater than the difference between the standards within each condition (15 ms vs 5 ms). The logic of the varying standards paradigm requires the auditory system to perceive these small differences in VOT – if a 5 ms VOT difference was below the threshold of a just noticeable difference, the auditory system would not be able to detect the phonetic variance in the first place. Prior studies have used intervals as short as 4 ms between standards to find phonological mismatch effects, so we assume a 15 ms

difference between conditions should be detectable by the auditory system, even in a high range of VOT (see Phillips et al., 2000; Hestvik & Durvasula, 2016).

A second possibility which has been put forward is that the MMN simply does not encode deviance magnitude when all factors (refractoriness, gradient encoding) are controlled for. Horváth et al. (2007) and Font-Alaminos et al. (2021) both conduct MMN experiments using similar control paradigms and find no deviance magnitude effects, arguing that these previously reported effects are artifactual. With proper control paradigms, deviance magnitude simply disappears. This is congruent with our findings here, as well as similar previous findings (Rhodes et al., 2019). This would imply that either the auditory predictive system simply does not propagate detailed information about a failed prediction via the prediction error signal (a claim fundamentally at odds with the central claims of the predictive coding framework), or that the MMN correlates only loosely with this prediction error signal (conveying the presence of prediction error but no additional information). We cannot adjudicate this claim here, but we maintain that the MMN is indexing at least the presence of prediction error, which is sufficient for our conclusions (predicated on presence vs absence of prediction error). Further research will be needed to settle the issue relating MMN and prediction error propagation.

By isolating a prediction error brain response to phonetically varying speech sounds, we add to a growing body of literature probing the nature and contents of the predictions generated by the brain's hierarchical auditory prediction machinery. In this study, we observe a late mismatch (LDN) effect. When controlling for mismatch sources other than prediction error (neuronal refractoriness, gradient encoding effects) we observe an LDN, representing a genuine prediction error signal in response to the presented VOT contrast. However, this mismatch is only significant for participants who perceived the contrast as across-category. This fact, coupled with a lack of phonetic distance effect, indicates that the observed prediction error is a function of phonological category identity and is not predicated on phonetic detail. In other words, when input is phonetically variable, the auditory system fails to incorporate that phonetic information into its predictions and instead relies solely on phonological categories to update its model of the auditory environment.

Notes

1. Six participants had categorization thresholds above 100 ms, but three of these were excluded independently for previously specified EEG data quality issues.
2. "Deviant" here refers simply to the 50 ms VOT stimulus. In the control condition, all stimuli appear with equal frequency, so the 50 ms stimulus is not a true deviant. However, because the 50 ms stimulus will be compared to deviants from the experimental conditions, we refer to it as "control deviant" as a shorthand.
3. For a more conventional deviant-minus-standard analysis, see the Supplemental Material. The effects found in the deviant-standard analysis differs substantially from the deviant-control analysis. In the deviant-standard analysis, we find a significant mismatch effect for both groups of participants in an early (252-352 ms) and late (452-552 ms) time window. In the deviant-control analysis, we find significant results only in the late (452-552 ms time window) and only for the >50 ms group. Although the results of the deviant-control analysis are more limited, we believe they have greater validity because the control paradigm controls for stimulus quality and stimulus frequency.

Acknowledgements

The authors would like to acknowledge the work of undergraduate research assistants Lena Herman and Megan Bahnson for their valuable contribution to the execution of this study.

Disclosure statement

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

ORCID

Ryan Rhodes http://orcid.org/0000-0002-3397-2012
Arild Hestvik http://orcid.org/0000-0003-4561-7584

References

Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, *10*(11), 2429–2434. https://doi.org/10.1097/00001756-199908020-00038

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, *16*(1), 38–51. https://doi.org/10.1097/00003446-199502000-00004

Allen, J. S., Miller, J. L., & Desteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552. https://doi.org/10.1121/1.1528172

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, *12*(7), 2570–2574. https://doi.org/10.1046/j.1460-9568.2000.00114.x

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187. https://doi.org/10.1016/0010-0277(94)90042-6

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding.

Cortex, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Barrios, S., Jiang, N., & Idsardi, W. J. (2016). Similarity in L2 phonology: Evidence from L1 spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Language Research*, *32*(3), 367–395. https://doi.org/10.1177/0267658316630784

Bendixen, A., Schröger, E., Ritter, W., & Winkler, I. (2012). Regularity extraction from non-adjacent sounds. *Frontiers in Psychology*, *3* (143), 1–12. https://doi.org/10.3389/fpsyg.2012.00143

Berti, S., Roeber, U., & Schröger, E. (2004). Bottom-up influences on working memory: Behavioral and electrophysiological distraction varies with distractor strength. *Experimental Psychology*, *51*(4), 249–257. https://doi.org/10.1027/1618-3169.51.4.249

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. https://doi.org/10.1016/j.neuroimage.2013.04.093

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *The Journal of the Acoustical Society of America*, *63*(5), 1556–1558. https://doi.org/10.1121/1.381849

Carbajal, G. V., & Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: From the subcortical roots to cortical deviance detection. *Trends in Hearing*, *22*, 233121651878482–33. https://doi.org/10.1177/2331216518784822

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

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A. M., & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, *33*(27), 11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013

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

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

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

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

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*(3), 757–777. https://doi.org/10.1044/1092-4388(2009/08-0123)

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, *8* (4), 919–924. https://doi.org/10.1097/00001756-199703030-00021

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

Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H. J., Kirsch, H. E., Dalal, S. S., Deouell, L. Y., & Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences*, *113*(24), 6755–6760. https://doi.org/10.1073/pnas.1525030113

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99–109. https://doi.org/10.1016/0010-0285(73)90006-6

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

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

Font-Alaminos, M., Ribas-Prats, T., Gorina-Careta, N., & Escera, C. (2021). Emergence of prediction error along the human auditory hierarchy. *Hearing Research*, *399*. https://doi.org/10.1016/j.heares.2020.107954

Forrest, K., Weismer, G., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115–123. https://doi.org/10.1121/1.396977

Friston, K. J. (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

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Frost, J. D., Winkler, I., Provost, A., & Todd, J. (2016). Surprising sequential effects on MMN. *Biological Psychology*, *116*, 47–56. https://doi.org/10.1016/j.biopsycho.2015.10.005

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. https://doi.org/10.1177/002383096200500401

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy 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., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences*, *104*(52), 20961–20966. https://doi.org/10.1073/pnas.0706274105

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

Grimm, S., & Escera, C. (2012). Auditory deviance detection revisited: Evidence for a hierarchical novelty system. *International Journal of Psychophysiology*, *85*(1), 88–92. https://doi.org/10.1016/j.ijpsycho.2011.05.012

Haggard, M., Ambler, S., Callow, M., & Haggard, M. (1970). Pitch as a voicing Cue. *The Journal of the Acoustical Society of America*, *47*(2), 613–617. https://doi.org/10.1121/1.1911936

Halliday, L. F., Barry, J. G., Hardiman, M. J., & Bishop, D. V. (2014). Late, not early mismatch responses to changes in frequency are reduced or deviant in children with dyslexia: An event-related potential study. *Journal of Neurodevelopmental Disorders*, *6*(1), 1–15. https://doi.org/10.1186/1866-1955-6-21

Heilbron, M., & Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

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

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

Hill, P. R., McArthur, G. M., & Bishop, D. V. M. (2004). Phonological categorization of vowels: A mismatch negativity study. *NeuroReport*, *15*(14), 2195–2199. https://doi.org/10.1097/00001756-200410050-00010

Hong, T., Shuai, L., Frost, S. J., Landi, N., & Pugh, K. R. (2018). Cortical responses to Chinese phonemes in preschoolers predict their literacy skills at school Age. *Developmental Neuropsychology*, *43*(4), 356–369. https://doi.org/10.1080/87565641.2018.1439946

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., & Winkler, I. (2007). MMN or no MMN: No magnitude of deviance effect on the MMN amplitude. *Psychophysiology*, *45*(1), 60–69. https://doi.org/10.1111/j.1469-8986.2007.00599.x.

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), 723–727. https://doi.org/10.1111/1469-8986.3840723

Jacobsen, T., & Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology*, *114*(6), 1133–1143. https://doi.org/10.1016/S1388-2457(03)00043-9

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

Joanisse, M. F., Robertson, E. K., & Newman, R. L. (2007). Mismatch negativity reflects sensory and phonetic speech processing. *NeuroReport*, *18*(9), 901–905. https://doi.org/10.1097/WNR.0b013e3281053c4e

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103

Kenstowicz, M. (1994). *Phonology in generative grammar*. *Blackwell*.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*(3), 332–339. https://doi.org/10.1006/brln.2000.2426

Korpilahti, P., Lang, H., & Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography and Clinical Neurophysiology*, *95*(4), P96. https://doi.org/10.1016/0013-4694(95)90016-G

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review*, *23*(6), 1681–1712. https://doi.org/10.3758/s13423-016-1049-y

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., & Friston, K. J. (2013). Correction: A neurocomputational model of the mismatch negativity. *PLoS Computational Biology*, *9*(11), e1003288. https://doi.org/10.1371/annotation/ca4c3cdf-9573-4a93-9542-3a62cdbb8396

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

Mah, J., Goad, H., & Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]. *Frontiers in Psychology*, *7*(OCT), 1–14. https://doi.org/10.3389/fpsyg.2016.01469

Martynova, O., Kirjavainen, J., & Cheour, M. (2003). Mismatch Negativity and Late Discriminative Negativity in Sleeping Human Newborns. *Neuroscience Letters*, *340*(2), 75–78. https://doi.org/10.1016/S0304-3940(02)01401-5

May, P., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Miglietta, S., Grimaldi, M., & Calabrese, A. (2013). Conditioned allophony in speech perception: An ERP study. *Brain and Language*, *126*(3), 285–290. https://doi.org/10.1016/j.bandl.2013.06.001

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

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(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*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Neuhoff, N., Bruder, J., Bartling, J., Warnke, A., Remschmidt, H., Müller-Myhsok, B., & Schulte-Körne, G. (2012). Evidence for the late MMN as a neurophysiological endophenotype for dyslexia. *PLoS ONE*, *7*(5), e34909–8. https://doi.org/10.1371/journal.pone.0034909

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*(1), 224–230. https://doi.org/10.1121/1.390399

Paavilainen, P., Simola, J., Jaramillo, M., & Näätänen, R. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, *38*(2), 359–365. https://doi.org/10.1111/1469-8986.3820359

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, *118*(1), 177–185. https://doi.org/10.1016/j.clinph.2006.09.001

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*(6), 1038–1055. https://doi.org/10.1162/08989290051137567

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics*, *18*(3), 375–394. https://doi.org/10.1016/S0095-4470(19)30380-8

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

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*(1), 49–71. https://doi.org/10.1016/j.pneurobio.2006.04.004

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

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, *1077*(1), 135–143. https://doi.org/10.1016/j.brainres.2006.01.043

Rodrigues Silva, D. M., Melges, D. B., & Rothe-Neves, R. (2017). N1 response attenuation and the mismatch negativity (MMN) to within- and across-category phonetic contrasts. *Psychophysiology*, *54*(4), 591–600. https://doi.org/10.1111/psyp.12824

Rodrigues Silva, D. M., & Rothe-Neves, R. (2020). Context-dependent categorisation of vowels: A mismatch negativity study of positional neutralisation. *Language, Cognition and Neuroscience*, *35*(2), 163–178. https://doi.org/10.1080/23273798.2019.1638948

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics*, *7*(4), 393–402. https://doi.org/10.1016/S0095-4470(19)31072-1

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

Rubin, J., Ulanovsky, N., Nelken, I., & Tishby, N. (2016). The representation of prediction error in auditory cortex. *PLOS Computational Biology*, *12*(8), e1005058–28. https://doi.org/10.1371/journal.pcbi.1005058

Ruusuvirta, T. (2021). The release from refractoriness hypothesis of N1 of event-related potentials needs reassessment. *Hearing Research*, *399*, 107923. https://doi.org/10.1016/j.heares.2020.107923

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. https://doi.org/10.3758/BF03202653

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

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

Schiff, S., Mapelli, D., Vallesi, A., Orsato, R., Gatta, A., Umiltà, C., & Amodio, P. (2006). Top-down and bottom-up processes in the extrastriate cortex of cirrhotic patients: An ERP study. *Clinical Neurophysiology*, *117*(8), 1728–1736. https://doi.org/10.1016/j.clinph.2006.04.020

Schluter, K., 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/j.1469-8986.1995.tb03406.x

Schröger, E., & Wolff, C. (1996). Mismatch response of the human brain to changes in sound location. *NeuroReport7*, *7*, 3005–3008. https://doi.org/10.1097/00001756-199611250-00041

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

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, *107*(5), 2697–2703. https://doi.org/10.1121/1.428655

Shestakova, A., Brattico, C. A. E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R. J., & Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *Cognitive Neuroscience and Neuropsychology*, *13*(0), 1–5.

Soli, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. *The Journal of the Acoustical Society of America*, *73*(6), 2150–2165. https://doi.org/10.1121/1.389539

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

Tanner, D., Morgan-short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and

cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Theodore, R. M., Miller, J. L., & Desteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, *125*(3974), 3974–3982. https://doi.org/10.1121/1.3106131

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90–92. https://doi.org/10.1038/372090a0

Todd, J., Heathcote, A., Whitson, L. R., Mullens, D., Provost, A., & Winkler, I. (2014). Mismatch negativity (MMN) to pitch change is susceptible to order-dependent bias. *Frontiers in Neuroscience*, *8*, 180. https://doi.org/10.3389/fnins.2014.00180

Todorovic, A., Van Ede, F., Maris, E., & De Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *Journal of Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization. *Psychological Science*, *21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

van Hessen, A. J., & Schouten, M. E. H. (1992). Modeling phoneme perception. II: A model of stop consonant discrimination. *The Journal of the Acoustical Society of America*, *92*(4), 1856–1868. https://doi.org/10.1121/1.403842

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Walter, M. A., & Hacquard, V. (2004). MEG Evidence for Phonological Underspecification. *14th Annual International Conference on Biomagnetism*, 292–293.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*(2), 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Winkler, I., Kujala, T., Tiitinen, H., Alku, P., Lehtokoski, A., Ilmoniemi, R. J., Sivonen, A., Czigler, I., Csépe, V., & Näätänen, R. (1999a). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638–642. https://doi.org/10.1111/1469-8986.3650638

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999b). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, *7*(3), 357–369. https://doi.org/10.1016/S0926-6410(98)00039-1

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of Voice Onset Time and F0 in Stop Consonant Voicing Perception: Effects of Masking Noise and Low-Pass Filtering. *Journal of speech, language, and hearing research : JSLHR*, *56*(4), 1097–1107. https://doi.org/10.1044/1092-4388(2012/12-0086

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

Yu, Y. H., Shafer, V. L., Sussman, E. S., & Monahan, P. J. (2017). *Frontiers in Neuroscience*, *11*(March), 1–19. https://doi.org/10.3389/fnins.2017.00095

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: Early and late mismatch negativity to tone patterns. *NeuroReport*, *16*(18), 2015–2019. https://doi.org/10.1097/00001756-200512190-00009