HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Neuroimage. 2016 March ; 128: 293–301. doi:10.1016/j.neuroimage.2016.01.003.

Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations

Mathias Scharinger[a,d,e], **Philip J. Monahan**[b,c], and **William J. Idsardi**[d]

[a]Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

[b]Centre for French and Linguistics, University of Toronto Scarborough, Canada

[c]Department of Linguistics, University of Toronto, Canada

[d]Department of Linguistics, University of Maryland, College Park, MD, USA

[e]Biological incl. Cognitive Psychology, Institute for Psychology, University of Leipzig, Germany

Abstract

While previous research has established that language-specific knowledge influences early auditory processing, it is still controversial as to what aspects of speech sound representations determine early speech perception. Here, we propose that early processing primarily depends on information propagated top-down from abstractly represented speech sound categories. In particular, we assume that mid-vowels (as in 'bet') exert less top-down effects than the high-vowels (as in 'bit') because of their less specific (default) tongue height position as compared to either high- or low-vowels (as in 'bat'). We tested this assumption in a Magnetoencephalographic (MEG) study where we contrasted mid- and high-vowels, as well as the low- and high-vowels in a passive oddball paradigm. Overall, significant differences between deviants and standards indexed reliable mismatch-negativity (MMN) responses between 200 and 300 ms post stimulus onset. MMN amplitudes differed in the mid/high-vowel contrasts and were significantly reduced when a mid-vowel standard was followed by a high-vowel deviant, extending previous findings. Furthermore, mid-vowel standards showed reduced oscillatory power in the pre-stimulus beta-frequency band (18–26 Hz), compared to high-vowel standards. We take this as converging evidence for linguistic category structure to exert top-down influences on auditory processing. The findings are interpreted within the linguistic model of underspecification and the neuropsychological predictive coding framework.

Corresponding author: Mathias Scharinger (PhD), Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, 60322, Grüneburgweg 14, Germany, mathias.scharinger@aesthetics.mpg.de.

Keywords

Speech Sound Perception; MEG; Mismatch Negativity; Cortical Oscillations; Beta-Band Power; Underspecification; Predictive Coding

1. Introduction

Neuroimaging methods have been increasingly used to probe the mechanisms that underlie speech sound processing. Recently, a number of studies have demonstrated that linguistic category structure has specific modulatory effects on early stages of auditory perception (Bien and Zwitserlood, 2013; Cornell et al., 2011, 2013; Eulitz and Lahiri, 2004; Friedrich et al., 2008). Linguistic category structure allows speech sound classification according to their acoustic and articulatory properties, often described in terms of a deviation from the neutral, resting position of the mouth. For example, high vowels (e.g., [ɪ] as in 'bit') with a relatively high tongue position during production can be distinguished from low vowels (e.g., [æ] as in 'bat') with a relatively low tongue position during production. Some theories assume that vowels that fall between high and low vowels (e.g. [ɛ] as in 'bet') are neither high nor low, and, being produced with a neutral tongue position, have no descriptive feature for tongue height (Lahiri and Reetz, 2002, 2010; Scharinger and Idsardi, 2014). The production of mid vowels in English does not necessarily lead to a larger spread of individual vowel tokens, but rather to greater overlap with neighboring vowel category tokens (Hillenbrand et al., 1995). These vowels are assumed to be underspecified and may refer to a rather unspecific motor plan regarding their tongue height.

Recently, it has been proposed that less specific, underspecified vowels have less intrinsic "predictive" value compared to more specific, specified vowels (Eulitz and Lahiri, 2004; Scharinger et al., 2012a; Scharinger et al., 2012b). Scharinger et al. (2012b) demonstrated that the unspecific category structure of the American English vowel [ɛ] influenced processing, as indexed by the Mismatch Negativity, an automatic change and prediction error response of the brain (Näätänen and Alho, 1997; Schröger, 2005; Winkler, 2007). In a passive oddball design, the authors contrasted the high- and low-vowels [ɪ] and [æ] in standard position with the low- and high-vowels [æ] and [ɪ] in deviant position. This condition showed a relatively large acoustic distance of the first resonance frequencies (first formant, F1) between the vowels and was compared to a condition in which the acoustic F1 distance was relatively small, i.e., in which the standard was either specific ([æ]) or unspecific ([ɛ]), contrasting with the deviants [ɛ] and [æ]. The results showed similar symmetric mismatch responses in the large F1-distance condition, while the small F1-distance condition showed asymmetric MMN differences: the condition with unspecific [ɛ]-standards yielded significantly reduced MMN amplitudes compared to the condition with specific [æ]-standards. This result is consistent with other electrophysiological studies (Cornell et al., 2011, 2013; Eulitz and Lahiri, 2004). Within the framework of predictive coding (Friston, 2005; Garrido et al., 2009), this pattern was interpreted as evidence for [ɛ] being inherently less predictive, such that the prediction error upon encountering the deviant [æ] was reduced.

While this study suggests that linguistic category structure may indeed influence early auditory processing, generalization to further vowel contrasts was impossible (e.g., between [ɛ] and [ɪ]). Moreover, there was no measure with a closer relation to the assumed top-down propagation of category information (strong for [ɪ] and [æ], weak for [ɛ]). In this regard, recent research suggests that cortical oscillations index directional message passing between different levels of the cortical hierarchy (Arnal and Giraud, 2012; Arnal et al., 2011; Engel and Fries, 2010; Fontolan et al., 2014). In particular, cortical oscillations within the beta-band (15–30 Hz) are assumed to reflect endogenous top-down processes that are interpreted within the predictive-coding framework (Wang, 2010). In this framework, beta-power scales with prediction strength that is propagated downward from representational units to lower processing levels. This mechanism should also operate on speech sound category representations, such that differences in linguistic structure lead to differences in cortical beta-power, which should arise prior to stimulus presentation in an MMN paradigm.

Thus, the current Magnetoencephalography (MEG) study has two primary goals: (1) to examine cortical oscillations as means to further elucidate the mechanisms by which linguistic category structure exerts influence on lower-level auditory processing, and (2) to extend the MMN findings from Scharinger et al. (2012b) to the contrast between the vowels [ɛ] and [ɪ]. We expect (1) beta-power to differ between [ɛ] and [ɪ] presented as standards, where predictions build up (Winkler et al., 1996a) and should most strongly be influenced by linguistic category structure, and (2) the MMN to be reduced or absent if deviant [ɪ] follows the standard [ɛ].

2. Methods

2.1. Participants

Thirteen students, all native speakers of American English, were recruited from the University of Maryland (9 females, 4 males, mean age 21 ± 1.3 years). They had no reported history of hearing or neurological problems and participated for class credit or monetary compensation ($10 per hour). All participants provided informed written consent and tested strongly right-handed (> 80%) on the Edinburgh Handedness Inventory (Oldfield, 1971). The study was approved by the Institutional Review Board of the University of Maryland and in accordance with the declarations of Helsinki.

2.2. Materials

Stimulus material was similar to that used in Scharinger et al. (2012b) and involved 10 renditions of each of the vowels [æ], [ɛ] and [ɪ], produced by a female native speaker of American English, who made a robust three-way height distinction (see Figure 1). All vowels were recorded embedded in the carrier sentence “I will say h__d again”. This was repeated 20 times for each vowel. The phonetically trained speaker ensured that vowels had the quality of short vowels. The speech material was digitized at 44 kHz with 16 bit amplitude resolution using the phonetic sound application PRAAT (Boersma and Weenink, 2011). We then spliced 100 ms out of the steady-state portion of the respective vowels from the carrier sentences and selected a final set of 10 vowels on the basis of intensity and pitch. The first 10 ms of each vowel was multiplied with the first half period of a $(1-\cos(x))/2$

function and the last 10 ms with the first half period of a (1+cos(x))/2 function to reduce acoustic artifacts. Stimulus intensity was normalized to 70 dB within PRAAT, which corresponds to sound-pressure level (SPL, Boersma and Weenink, 2011). We further set up the sound delivery system in the MEG scanner such that participants would hear the stimuli at a level of 60 dB SPL, which was confirmed using a sound pressure level meter in the MEG cabin. Finally, we obtained three independent opinions regarding the perceived loudness of the three vowel types in the scanner. Since no differences in perceived loudness was reported, no further loudness modifications were deemed necessary. Detailed acoustic measures of the vowel stimuli are provided in Table 1 and illustrated in Figure 1.

Spectral analyses of the vowel stimuli involved a linear predictive coding (LPC) formant analysis and estimated the first three resonant frequencies (formants, F1–F3). As the three vowels mainly differ in tongue height, which is inversely correlated with F1 frequency (Stevens, 1998), we had defined that the opposition of [æ] and [ɪ] constitutes the large F1-distance condition, while the opposition of [ɛ] and [ɪ] represents the small F1-distance condition. This definition was corroborated by the Euclidean F1 distances which were larger between [æ] and [ɪ] (491.8 Hz) than between [ɛ] and [ɪ] (269.5 Hz; t (18) = 17.88, $p < 0.001$).

2.3. Design

Vowel stimuli were presented in a passive standard/deviant many-to-one oddball paradigm (Winkler et al., 1999): The vowels [æ]/[ɪ] (large F1 distance) and [ɛ]/[ɪ] (small F1 distance) were distributed over four blocks (the order permutated across participants) in which they occurred in either standard (p = 0.875, N = 700) or deviant position (p = 0.125, N = 100, for details, see Figure 1C). We referred to the direction of standard/deviant presentation as “F1 increasing” if the deviant had a higher F1 (lower tongue position) than the standard (i.e. [ɪ]-[æ]; [ɪ]-[ɛ]) and as “F1 decreasing” if the deviant had a lower F1 (higher tongue position) than the standard (i.e. [æ]-[ɪ]; [ɛ]-[ɪ]). The distribution of vowel stimuli over the factor levels is illustrated in Table 2.

The 10 different vowel renditions for standards and deviants had the same probability of occurrence. Note that using different renditions for standards is beneficial for activating memory traces not bound to a particular phonetic realization but rather referring to an abstract representation (see Phillips et al., 2000). The number of consecutive standards pseudo-randomly varied between 3 and 10, and inter-stimulus intervals (ISIs) were jittered between 500 and 1000 ms (in steps of 1 ms, random selection from a uniform distribution) to prevent participants from entraining to a specific presentation rhythm. A total of 800 vowel stimuli were presented in each block, leading to block durations of approximately 15 minutes and a total experiment duration of about 90 minutes. This included participant preparation and debriefing. Stimulus presentation was controlled by *Presentation* software (Neurobehavioral Systems, Albany, CA); delivery of auditory stimuli into the shielded MEG chamber was achieved by air conduction transduction and non-magnetic earphones (Etymotic Research Inc., IL, USA), resulting in a binaural, comfortable listening level at 60 dB (SPL). Earphones (Etymotic ER3A insert) were calibrated to have a flat frequency

response between 50 Hz and 3100 Hz within the shielded room. This guaranteed an optimal acoustic delivery of the first three vowel formant frequencies (Stevens, 1998).

2.4. MEG recording

MEG activity was recorded from 157 axial gradiometers (whole-head system, Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 500 Hz. Data were online filtered between DC and 200 Hz, together with a notch filter of 60 Hz to reduce ambient electrical noise. Participants lay supine in a magnetically shielded chamber, with two pre-auricular and three pre-frontal electrodes attached to them to account for head movement within the scanner.

Prior to the main experiment, an auditory localizer was performed to independently determine the 10 strongest channels per hemisphere in response to sinusoidal tones. To that end, participants were instructed to silently count a total of 150 high (1000 Hz) and 150 low (250 Hz) tones.

During the main experiment, participants passively listened to vowel stimuli presented in four blocks, as illustrated in Figure 1C, with short breaks in-between each block. Participants viewed a silent movie of their choice to reduce excessive eye movements and to maintain an awake state (Tervaniemi et al., 1999). A projector outside the shielded chamber projected the movie through a mirror system onto a 36 × 24 cm screen mounted approximately 15 cm above the participants.

2.5. MEG preprocessing and analysis

All MEG raw data were analyzed within fieldtrip (Oostenveld et al., 2011), running on Matlab 2009b (Mathworks, Inc., Natick, MA, USA). In a first step, environmental and sensor noise was removed from the MEG raw data by means of a multi-shift PCA noise reduction algorithm (de Cheveigné and Simon, 2007, 2008).

2.5.1. Auditory localizer—Neuromagnetic responses to the sinusoids in the auditory localizer pre-test were epoched from −200 pre-stimulus onset to 500 ms post-stimulus onset. Epochs were band-pass filtered between 0.3 Hz and 30 Hz (Hamming-window digital Butterworth filter). For baseline correction, the mean amplitude of the pre-stimulus window (−200-0 ms) was subtracted from the epoch. Responses to 250 Hz and 1000 Hz tones were averaged separately.

The scalp distribution of the resulting averaged evoked field between 90 and 140 ms post stimulus onset was consistent with the typical N1m source in supra-temporal auditory cortex (Diesch et al., 1996) for all participants and tone types. Magnetic sources and sinks were distributed over posterior and anterior positions with a typical reversal across hemispheres (Figure 2). From the by-participant average of the auditory localizer, the strongest 5 channels per quadrant (left anterior, left posterior, right anterior, right posterior) were selected for subsequent MMN analyses. This was done on the basis of both the 250 Hz and 1000 Hz tones. In rare cases, the strongest channels differed between the two tones in which case the selection was based on the 1000 Hz tone (see Scharinger et al., 2012b), which typically has a larger amplitude and shorter latency.

2.5.2. Passive oddball paradigm evoked responses—Epochs 1000 ms in duration were extracted from the continuous waveform centered around stimulus onset, i.e., −500 ms (pre-stimulus onset) to 500 ms (post-stimulus onset) for all trials, to allow an examination of pre-stimulus effects as well as evoked responses. First, a Hamming-window digital Butterworth low pass filter at 100 Hz was applied to the data. Subsequently, we applied automatic artifact rejection and channel repair to remove outliers and noisy channels. Epochs containing artifacts or epochs with amplitudes exceeding 3 pT (pT, 10^{-12} Tesla) were automatically discarded from further analyses using fieldtrip (Oostenveld et al., 2011). This helped eliminate MEG jump artifacts, muscle artifacts, and extensive eye movements. This affected less than 15% of trials on average per participant. Channel repair involved the detection and interpolation of dead channels. Interpolation was done using the average of the four nearest neighbors for the respective dead channels. No more than two channels per participant had to be interpolated.

For evoked-response MMN analyses, epochs were re-defined to include a 100 ms pre-stimulus and a 500 ms post-stimulus window. Bandpass-filtering was done between 0.3 and 30 Hz using Hamming-window digital Butterworth filter. The mean amplitude of the −100 to 0 ms pre-stimulus window was subtracted from the epoch. Root-mean squared (RMS) amplitudes were calculated over the channels of interest obtained from the auditory localizer, separately for the left and right hemisphere of each participant. For illustration purposes in Figures 2 and 3, epochs were additionally low-pass filtered at 15 Hz using a Hamming-window digital Butterworth filter.

2.5.3. Time-frequency analyses—Time-frequency analyses were done separately for standards and deviants. First, cleaned data (epochs of 1 sec, low-pass filtered at 100 Hz as described above) were down-sampled to 125 Hz, and the linear trend was removed from the data using General Linear Modeling, as implemented in Fieldtrip. Subsequently, single-trial MEG data was decomposed into time-frequency representations with a Morlet wavelet analysis (Bertrand and Pantev, 1994) for each channel centered around each time point. The analysis was carried out using Hanning windows that moved in steps of 10 ms along the temporal dimension (from −500 to 500 ms). In the spectral dimension, we used 1-Hz bins from 1 to 30 Hz. Wavelet widths ranged from 1 to 8 cycles, equally spaced over the 30 frequency bins. For normalization, analysis windows were divided by their number of points. Resulting complex values of each frequency bin were transformed to power values, and mean power values of a pre-stimulus baseline interval (−300 to −200 ms) were subtracted from the epoch. The selection of this time window was based on the following considerations: (a) It should contain a minimum of activity from previous stimuli, and (b) it should leave room for a pre-stimulus window that is not subtracted from the epoch. With the above window selection, only in the shortest ISI that occurred in 0.2% of the standards in a block, would this window contain potential activity from a previous stimulus occurring at latencies > 300 ms. At the same time, it would allow for a 200 ms pre-stimulus window that was not subtracted from the epoch.

Note that due to the single-trial time-frequency transformation, resulting time-frequency representations contain both evoked and induced (total) power. Power latencies used in the statistics and shown in the Figures are based on the middle latencies of the wavelets.

2.6. Data Analysis

For all analyses, the first three standards of each block and the first standard after a deviant were discarded.

2.6.1. MMN responses—Inspection of the grand average waveforms (Figure 3) revealed the strongest mismatch effects between 200 and 300 ms post stimulus onset. Note that we have defined the mismatch response as a difference in the evoked field between the same physical stimulus in deviant and in standard position (identity MMN; Pulvermüller and Shtyrov, 2006; Pulvermüller et al., 2006), based on the RMS as in previous research (see Herholz et al., 2009; Lappe et al., 2013; Scharinger et al., 2012b; "difference of the RMS method"). The above-defined window is comparable with windows used in previous research and also with respect to slightly later MMN peaks to speech sounds compared to sinusoids (Phillips et al., 2000; Scharinger et al., 2012b).

Since the method of defining the MMN by subtracting the RMS of the standard from the RMS of the deviant may be insensitive to topographical differences, we additionally calculated the MMN as the RMS of the deviant-standard difference ("RMS of the difference method"). Differences between deviant and standard responses (RMS averages for each participant in each condition and hemisphere) were tested in a linear mixed-effect model (LMM) using the lme 4 package (Bates et al., 2014) within the statistical software R (R Development Core Team, 2014). Reported F-values are estimated by the lmerTest package (Kuznetsova et al., 2014). The model was comprised of the random effect *participant*, the fixed effects *position* (standard, deviant) and *hemisphere* (left, right), as well as the interaction *position x hemisphere*. The effect of *position* was significant ($F(1,192) = 19.04$, $p < 0.001$), with larger amplitudes for deviants than standards. Neither *hemisphere* nor the interaction *hemisphere x position* reached significance (all $Fs < 1$, $p > 0.60$).

Subsequently, we focused on differences between deviants and standards that were calculated separately for each direction (increasing F1, decreasing F1), distance condition (large F1/small F1), and participant, based on either the *difference of the RMS method* or the *RMS of the difference method*. In both cases, difference values were entered into different linear mixed-effect models. These models involved a random subject effect and the fixed effects *distance* (large F1 distance, small F1 distance), *direction* (increasing F1, decreasing F1) and *hemisphere* (left, right), as well as the interaction between the effects of *distance* and *direction*. Post-hoc analyses were calculated using the multcomp package in R (Bretz et al., 2011) and consisted of Tukey-adjusted *t*-tests with *z*-transformed *t*-values. Post-hoc comparisons were driven by our assumption that responses should not differ between the factor levels of *direction* in the large F1 distance condition (same predictive standard context), while responses should differ between the factor levels of *direction* in the small F1 distance condition (predictive and non-predictive standard context).

2.6.2. Power contrasts—Power analyses focused on standard responses. Grand averages of the time-frequency analyses showed discernible differences between [ɛ] and [ɪ] in the small F1 distance condition. These differences were most pronounced in a beta-cluster (17–27 Hz) from −70 to 30 ms (pre-stimulus beta) and in a theta-cluster (4–7 Hz) from 300 to

400 ms (theta, cf. Figure 4). The definition of these regions was based on visual inspection and on typical beta and theta-frequency findings (Engel and Fries, 2010; Strauß et al., 2014). The by-participant power of these regions was averaged across channels, time- and frequency-bins (separately for the four standards) and subjected to LMMs with the same structure as for the MMN analyses (random factor *subject*, fixed effects *distance* and *direction*, and the interaction *distance* × *direction*). Post-hoc tests were calculated as described above.

To obtain a more conservative measure of power contrasts without prior assumptions of regions of interest, we additionally followed a multilevel statistical approach for time-frequency power comparisons (e.g. Henry and Obleser, 2012; Strauß et al., 2014). At the first level, we calculated independent-samples *t*-tests between the single-trial time-frequency power values of [æ] and [ɪ] (large F1 distance condition) and of [ɛ] and [ɪ] (small F1 distance condition). Uncorrected by-participant *t*-values were obtained for all time-frequency bins of all channels. At the second level, *t*-values were tested against 0 with dependent-sample *t*-tests. A Monte-Carlo nonparametric permutation method with 1000 randomizations as implemented in fieldtrip (Oostenveld et al., 2011) estimated type I-error controlled cluster significance probabilities (at $p < 0.05$).

3. Results

3.1. MMNs

Grand averages of standard and deviant responses for all four conditions are illustrated in Figure 3.

The LMM on the MMN response using the *difference of the RMS method* showed a main effect of *distance* ($F(1,84) = 4.25$, $p < 0.05$), with larger MMN responses in the large F1 distance condition than in small F1 distance condition ($z = 2.93$, $p < 0.01$). Crucially, *distance* interacted with *direction* ($F(1,84) = 4.40$, $p < 0.05$). This interaction effect recapitulates the distinction between predictive and non-predictive contexts from Table 2, with the one non-predictive context eliciting a smaller MMN. Post-hoc comparisons showed that in the large F1 distance condition, MMN responses did not differ between the increasing ([ɪ]-[æ]) and decreasing ([æ]-[ɪ]) direction ($z = 0.03$, $p = 0.98$), while in the small F1 distance condition, MMN responses were smaller in the decreasing ([ɛ]-[ɪ]) than in the increasing ([ɪ]-[ɛ]) direction ($z = -2.92$, $p < 0.01$; see Figure 3). No other effects or interactions were significant (all $Fs < 2$, $p > 0.2$).

The LMM on the MMN response using the *RMS of the difference method* showed a main effect of distance ($F(1,84) = 10.74$, $p < 0.01$), with larger MMN responses in the large F1 distance condition than in small F1 distance condition. Furthermore, there was a trend for a direction x distance interaction ($F(1,84) = 2.80$, $p = 0.09$). Although strictly speaking not legitimate to decompose, we examined the pattern of this interaction and found that in the large F1 distance condition, MMN responses did not differ between the increasing ([ɪ]-[æ]) and decreasing ([æ]-[ɪ]) direction ($z = 1.15$, $p = 0.44$), while in the small F1 distance condition, MMN responses were smaller in the decreasing ([ɛ]-[ɪ]) than in the increasing ([ɪ]-[ɛ]) direction ($z = -3.52$, $p < 0.01$).

3.2. Power

The LMM on total beta power in the pre-stimulus region of interest (−70–30 ms) showed no significant main effects (all Fs < 2), but a significant interaction of *distance* and *direction* ($F(1,48) = 5.18$, $p < 0.05$). Planned comparisons revealed that power values did not differ between the standards [æ] and [ɪ] in the large F1 distance condition ($z = 0.64$, $p = 0.77$, see Figure 4A), but in the small F1 distance condition, beta power in response to standard [ɛ] was significantly smaller than in response to standard [ɪ] ($z = -2.58$, $p < 0.05$, see Figure 4B).

The LMM in the theta regions of interest showed no significant main effects or interaction (all Fs < 3, $p > 0.2$). Note that we also looked at power differences in the deviants (illustrated in Supplementary Figure 1). These analyses revealed power reductions between 15 and 25 Hz at latencies around 300 ms post stimulus onset in all vowels but mid [ɛ]. As a result, deviant [ɛ] showed higher beta-power than deviant [ɪ] in the small F1 distance condition. The difference plot also revealed higher pre-stimulus (−100 to −50 ms) beta for deviant [ɛ] than for deviant [ɪ]; however, LMMs on power values obtained from these regions did not show any significant effects, and only trends for interactions (Fs < 3, $p > 0.09$).

The more conservative multi-level statistical approach confirmed that power in the small F1 distance condition differed between [ɛ] and [ɪ] in a time-frequency region between −50 and 20 ms and between 18 and 26 Hz. This is visible from a negative cluster illustrated in Figure 5B with a right-anterior distribution that survived the cluster-permutation based threshold. Note that power differences in theta band (as suggested by Figure 4) and power differences in the beta band for deviants did not survive this statistical threshold ($t = \pm 2.18$). The multi-level approach also confirmed that there were no power differences between the standards in the large F1 distance condition (Figure 5A).

4. Discussion

The main result of our neuromagnetic study on the processing of vowels with differing linguistic structure is that the mid vowel [ɛ] (as in ‘bet’) consistently resulted in neural patterns that were distinct from the more specific high vowels [ɪ] and [æ]: oscillatory power in the beta-band was reduced even before the onset of [ɛ] in standard position, compared to [ɪ]. Moreover, MMN amplitudes were significantly reduced when standard [ɛ] preceded deviant [ɪ], compared to the reverse case, i.e. when standard [ɪ] preceded deviant [ɛ]. We interpret this as evidence that mid [ɛ] is indeed underspecified in its long-term representation and exerts less predictive top-down effects during processing. The difference in the statistical patterns between the difference of the RMS and the RMS of the difference methods suggest that the effects were accompanied by topographical differences between conditions. This implies different source configurations depending on the vowel stimuli, which is consistent with previous observations of dipole differences as a function of vowel type (e.g. Obleser et al., 2004; Scharinger et al., 2011). These differences in configuration may also underlie the MMN response.

The pre-stimulus oscillatory beta-power effect provides evidence for a (mainly) predictive top-down mechanism that operates before sensory evidence actually becomes available. Our findings are discussed in detail below, and are compatible with both a neuropsychological predictive coding framework (Baldeweg, 2006; Friston, 2005) and a linguistic underspecification account (Lahiri and Reetz, 2002, 2010; Scharinger, 2009).

4.1. Beta oscillations distinguish between speech sounds of differing specificity

Brain oscillatory activity is increasingly used to examine the dynamics and functional coupling and uncoupling of networks involved in cognitive processing (Buzsáki, 2006; Buzsáki and Draguhn, 2004; Mazaheri et al., 2014), as well as in processing speech and language (Arnal et al., 2011; Doelling et al., 2014; Lewis et al., 2015; Luo and Poeppel, 2007; Obleser and Weisz, 2012; Peelle and Davis, 2012). Furthermore, brain oscillatory activity has been studied against the background of predictive processing (Arnal and Giraud, 2012; Arnal et al., 2011; Morillon and Schroeder, 2015; Peelle and Davis, 2012; Stefanics et al., 2011). Regarding our specific assumptions of the top-down propagation of category information onto speech sound processing, a particular role has been ascribed to beta-band (15–30 Hz) oscillations in electrophysiological responses (Arnal and Giraud, 2012; Bidelman, 2014; Buschman and Miller, 2007; Engel and Fries, 2010; Fontolan et al., 2014). These studies suggest that top-down propagation of predictive information is indexed by beta-band synchrony. Fontolan et al. (2014) showed that beta (15–30 Hz) frequencies were dominant in propagating information top-down (from higher levels of processing to lower levels of processing), while gamma (30–80 Hz) frequencies were dominant in propagating information bottom-up (from lower levels to higher levels of processing). This functional separation of frequency channels aligns with assumptions of the predictive coding framework (Friston, 2005), with prediction errors stemming from bottom-up sensory evidence being projected bottom-up, and predictions stemming from representational units being projected top-down (Arnal and Giraud, 2012; Wang, 2010). Given these findings and our interest in top-down propagation of information derived from linguistic category structure, we focused on low-frequency oscillations and examined the oscillatory dynamics during and prior to the processing of standard stimuli. Note that we restricted these analyses to the standards because we assumed that they would show the top-down effect most strongly, since conflicting bottom-up evidence was only encountered with deviants.

We found that pre-stimulus beta-power (18–26 Hz) was significantly reduced for [ɛ] as compared to [ɪ]. No such differences were observed between the standards [æ] and [ɪ]. We interpret this pattern to reflect differences in linguistic category structure, setting apart the mid-vowel [ɛ] from the high vowel [ɪ]. Presented in standard position, top-down propagation of linguistic category structure information appears to have been weaker (or less predictive) for [ɛ] than for [ɪ], because of the less specific representation of [ɛ]. We propose that this top-down effect is indexed by beta-power reduction, consistent with studies suggesting that beta-power scales with prediction strength propagated downward from representational units to lower-level auditory processing (Arnal et al., 2011; Fontolan et al., 2014). Note that the pre-stimulus beta-power difference strengthens the claim that the underlying mechanism is of a predictive nature. Beta-power effects after stimulus onset, for which we found only weak statistical evidence in this study, may be interacting with bottom-up sensory evidence.

Our strong focus on top-down modulations precluded us from looking at gamma oscillations but extending oscillation analyses to higher frequencies is clearly necessary in future work.

It is possible that the pre-stimulus time-interval may have been contaminated by oscillatory activity from preceding standards. While we cannot exclude such a possibility, we would argue that considerable contamination should only occur at the shortest ISI of 500 ms between subsequent standards. In 0.2% of the standards, oscillatory activity from latencies larger than 300 ms of a preceding standard may have spilled over into the −300 to −200 pre-stimulus window used as the baseline; however, even at the shortest ISI, it is unlikely that any oscillatory activity in the first 400 ms after preceding stimulus onset may have contributed to the pre-stimulus oscillatory effect of a subsequent standard. While we acknowledge that the assumed pre-stimulus effect may contain some bottom-up activity, we suspect that it is primarily driven by top-down processing.

Note that regarding the power results of the deviants, no pre-stimulus power difference in the small F1-distance could be observed. Our study cannot adjudicate between two possible interpretations: On the one hand, it is possible that the lack of a pre-stimulus power effect for deviants is related to the fact that deviant trials had a much smaller number than standard trials. On the other hand, a more plausible reason for this finding may have to do with the fact that top-down expectations for subsequent standards actually decrease with increasing numbers of standards. Participants in this experiment implicitly learn that after a couple of standards, it is more and more likely to encounter a deviant, such that the top-down propagation of linguistic information (pertaining to a following standard) is actually reduced prior to the deviants as compared to other positions within the standard sequence. Clearly, these two interpretations are speculative at the moment and require further research.

Apart from a predictive account, differences in beta-power from standard-[ɛ] responses might arise if participants did not interpret the mid-vowel tokens used in this experiment as prototypical category members. This possibility is based on the observation of Bidelman (2014), who found that beta-power was increased in response to prototypical vowel sounds, compared to less prototypical, ambiguous vowel sounds. The authors also showed that beta-power scaled with the slope of the psychometric identification function, suggesting that beta-power codes the strength of categorical percepts. Note that this assumption is not necessarily at odds with the proposed scaling of beta-power with top-down information propagation. It might be possible that speech sounds with less specific representations have more ambiguous realizations. Thus, it may have been more likely to obtain ambiguous tokens for [ɛ] than for [ɪ] or [æ], which in turn caused the reduction in beta-power; however, this interpretation remains speculative as we have no independent measure of whether participants in fact interpreted the vowel stimuli as non-prototypical category members.

A more general interpretation of the observed beta-effects with regard to sensori-motor interactions, as suggested by the link between beta-oscillations and motor activity (Aumann and Prut, 2015; Classen et al., 1998; Engel and Fries, 2010), is beyond the scope of this paper. We speculate that the reduced beta-power in response to [ɛ] might reflect less specific motor plans (and therefore less specific acoustic consequences), because [ɛ] is produced with a neutral tongue position. This contrasts with [ɪ], where the tongue has to reach a

specifically high target position, and with [æ], where the tongue has to reach a specifically low target position. Clearly, future research and experiments with possibilities of source-localizing beta-power effects in speech perception are necessary in order to arrive at more conclusive evidence.

4.2. Asymmetries in MMN responses

A growing body of research suggests that early, pre-attentive sound processing is influenced by the native language of the listener (Dehaene-Lambertz et al., 2000; Näätänen, 2001; Näätänen et al., 1997; Peltola et al., 2003; Pettigrew et al., 2004; Sharma and Dorman, 2000; Winkler et al., 1999). In these studies, the MMN (amplitude and/or latency) is influenced by linguistic information beyond mere acoustic contrasts. Further research has revealed that the MMN is not only sensitive to the statistics of sound sequences (Alexandrov et al., 2011; Bonte et al., 2005), but also to the abstract linguistic structure of the speech sounds themselves (Cornell et al., 2011, 2013; Eulitz and Lahiri, 2004; Scharinger et al., 2012a; Scharinger et al., 2012b). In particular, Scharinger et al. (2012b) set out to test the claim that the mid vowel [ɛ] is less specified than either [ɪ] or [æ] with regard to the vowel's tongue height. In their MEG study, the authors could show that the amplitude of the MMN was significantly reduced if the deviant [æ] occurred in the context of [ɛ], but not if the deviant [ɛ] occurred in the context of [æ]. Here, we complemented this design by examining the effects of the deviant [ɪ] in the context of [ɛ] and found very similar results: The MMN to [ɪ] in the context of [ɛ] was reduced, compared to the MMN to [ɛ] in the context of [ɪ]. By contrast, the opposition of [æ] and [ɪ] yielded symmetrical MMN responses in both directions, again similar to the results of Scharinger et al. (2012b). Note that in the statistical analysis of this study, we accounted for a possible acoustic confound. While the main effect of acoustic distance is in line with the existing evidence for the MMN amplitude to be positively correlated with the acoustic distance between standards and deviants in some studies (see Näätänen et al., 2007, for a review), the interaction of the effects of *distance* and *direction* and the interaction pattern (Figure 3) exclude the possibility that the observed effect is solely of an acoustic origin. Note, however, that we cannot make a strong claim to this effect due to the observed divergence of the MMN results between the difference of the RMS and the RMS of the difference method. Future research is necessary to allow for a more concise separation of acoustic and categorical effects and concomitant topographic differences in the MMN.

These notes of caution aside, the underspecification model assumes that the mid vowel [ɛ] has no long-term memory representation for tongue height because this vowel is produced with a neutral tongue position that is neither low (as [æ]) nor high (as [ɪ]). For this reason, the memory trace activated by the standard [ɛ] lacked tongue height information (for a similar argument, see Eulitz and Lahiri, 2004). The deviant [ɪ], then, did not provide conflicting tongue height information: The resulting MMN was not enhanced by linguistic category information, i.e., was reduced. In the reverse condition, the standard [ɪ] assumedly activated a long-term memory representation for tongue height (i.e., high, and its acoustic correlate, low F1), and the deviant [ɛ] did provide conflicting tongue height information (in its acoustic F1 value being outside of the predicted range induced by the high-vowel standards): The resulting MMN was enhanced by linguistic category information (i.e., was

significant in this experiment). Note that the underspecification account introduced by Lahiri and colleagues (Lahiri and Reetz, 2002, 2010) does not ascribe to a pure acoustic approach that would assume auditory representations to be built on the observed statistics of speech. Neither does it ascribe to a pure articulatory approach that would claim that there is no motor plan pertaining to tongue position for the production of mid vowels. A difficulty for a pure acoustic approach results from the observation that mid vowel realizations are not necessarily more variable (in terms of F1 spreading), but rather more likely to overlap with vowel tokens of neighboring categories (Hillenbrand et al., 1995), and thus, harder to discriminate. Note, however, that Hillenbrand et al. (1995) did not find evidence that the mid vowels were harder to discriminate than the non-mid vowels. By contrast, Peterson and Barney (1952) showed that [ɛ] had one of the lowest accuracy rates of any category. Moreover, [ɛ] and [æ] were the most confused categories. All of this is with the caveat that accuracy was very high in both studies. Thus, it is not necessarily the case that mid-vowels are harder to recognize, but perhaps harder to discriminate from neighbouring vowels in same cases.

A pure articulatory approach is challenged by the observation that high and low vowels are characterized by relatively stable F1 values that are robust against tongue displacements (Perkell, 1996). Thus, the specification of tongue height in high and low vowels does not necessarily mean that articulatory targets are very precise and constrained, while the non-specification of mid vowels does not necessarily mean that acoustic realizations show more F1 variations. While a more thorough discussion of acoustic vs. articulatory approaches is beyond the scope of this article, it is important to note that the underspecification approach and in fact, most theories assuming abstract phonological features, envisages abstract phonological information that represents the interaction between articulatory configurations and acoustic consequences.

Within the predictive coding framework, it is assumed that standard presentations help the suppression of prediction errors while integrating bottom-up sensory information with top-down predictions. These predictions may be derived from linguistic category structure (see Scharinger et al., 2012a; Scharinger et al., 2012b). When the deviant occurs, the sensory bottom-up information fails to meet the top-down prediction. The consequence is the elicitation of an MMN (Bendixen et al., 2012; Näätänen and Winkler, 1999; Winkler et al., 1996b). For the current experiment, we assume that the mid-vowel [ɛ] differs from the high-[ɪ] and low-vowel [æ] in its predictiveness: Since [ɛ] is not specified for tongue height in its long-term memory representation, the top-down support from this representation is relatively weak and leads to a less precise prediction regarding tongue height and the acoustic consequences thereof. As a result, the prediction error elicited by the deviant [ɪ] should be relatively weak, consistent with the reduced MMN. On the other hand, in all other cases, the vowels [ɪ] and [æ] in standard position were more predictive: Their specific tongue height information (high vs. low) allowed stronger top-down support, leading to possibly more precise predictions regarding tongue height and its acoustic consequences. In turn, the respective deviants elicited a strong prediction error that was reflected in significant MMNs. While Scharinger et al. (2012b) did not provide experimental evidence for possible top-down propagations of predictions, the time-frequency analysis in this study allows for a consolidation of the above interpretation.

5. Conclusions

In this neuromagnetic study on the processing of American English vowels, we found that less specific category structure (as exemplified by the mid vowel [ɛ]) resulted in reduced MMN responses and reduced beta-power. Our results are compatible within the predictive coding framework (Friston, 2005) and the underspecification approach (Lahiri and Reetz, 2002, 2010), while pure bottom-up sensory models would not be able to readily account for the observed patterns in our data. This is particularly true for the pre-stimulus beta-power effect that starts before sensory evidence becomes available and therefore provides evidence for reflecting a predictive top-down mechanism. We thus conclude that linguistic category structure exerts specific influences onto lower levels of early auditory processing. These influences primarily depend on the abstract representations of speech sounds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ariane Rhone for helping prepare the stimuli and Max Ehrmann for laboratory assistance. The research for this study was funded by the NIH grant 7ROIDC005660-07 to W.J.I. & David Poeppel. During preparation of the manuscript, MS was supported by a personal grant from the German Science Foundation (DFG) on “Global and local aspects of temporal and lexical predictions for speech processing” (University of Leipzig).

References

Alexandrov AA, Boricheva DO, Pulvermüller F, Shtyrov Y. Strength of word-specific neural memory traces assessed electrophysiologically. PLoS ONE. 2011; 6:e22999. [PubMed: 21853063]

Arnal LH, Giraud AL. Cortical oscillations and sensory predictions. Trends in Cognitive Sciences. 2012; 16:390–398. [PubMed: 22682813]

Arnal LH, Wyart V, Giraud AL. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nature Neuroscience. 2011; 14:797–801. [PubMed: 21552273]

Aumann TD, Prut Y. Do sensorimotor beta-oscillations maintain muscle synergy representations in primary motor cortex? Trends in Neurosciences. 2015; 38:77–85. [PubMed: 25541288]

Baldeweg T. Repetition effects to sounds: Evidence for predictive coding in the auditory system. Trends in Cognitive Sciences. 2006; 10:93–94. [PubMed: 16460994]

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–6. 2014. http://CRAN.R-project.org/package=lme4

Bendixen A, SanMiguel I, Schröger E. Early electrophysiological indicators for predictive processing in audition: A review. International Journal of Psychophysiology. 2012; 83:120–131. [PubMed: 21867734]

Bertrand, O.; Pantev, C. Stimulus frequency dependence of the transient oscillatory auditory evoked response (40 Hz) studied by electric and magnetic recordings in humans. In: Pantev, C.; Elbert, T.; Lütkenhöner, B., editors. Oscillatory Event-Related Brain Dynamics. Plenum Press; New York: 1994. p. 231-242.

Bidelman GM. Induced neural beta oscillations predict categorical speech perception abilities. Brain and Language. 2014; 141C:62–69. [PubMed: 25540857]

Bien H, Zwitserlood P. Processing nasals with and without consecutive context phonemes: Evidence from explicit categorization and the N100. Frontiers in Psychology. 2013:4.10.3389/fpsyg.2013.00021 [PubMed: 23378839]

Boersma, P.; Weenink, D. PRAAT: Doing Phonetics by Computer (ver. 5.2.24). Institut for Phonetic Sciences; Amsterdam: 2011.

Bonte ML, Mitterer H, Zellagui N, Poelmans H, Blomert L. Auditory cortical tuning to statistical regularities in phonology. Clinical Neurophysiology. 2005; 116:2765–2774. [PubMed: 16256430]

Bretz, F.; Hothorn, T.; Westfall, P. Multiple Comparisons Using R. CRC Press; Boca Raton, FL: 2011.

Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science. 2007; 315:1860–1862. [PubMed: 17395832]

Buzsáki, G. Rhythms of the Brain. Oxford University Press; Oxford: 2006.

Buzsáki G, Draguhn A. Neuronal oscillations in cortical networks. Science. 2004; 304:1926–1929. [PubMed: 15218136]

Classen J, Gerloff C, Honda M, Hallett M. Integrative visuomotor behavior is associated with interregionally coherent oscillations in the human brain. Journal of Neurophysiology. 1998; 79:1567–1573. [PubMed: 9497432]

Cornell SA, Lahiri A, Eulitz C. "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. Brain Research. 2011; 1394:79–89. [PubMed: 21549357]

Cornell SA, Lahiri A, Eulitz C. Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. Journal of experimental psychology Human perception and performance. 2013; 39:757–772. [PubMed: 23276108]

de Cheveigné, Ad; Simon, JZ. Denoising based on time-shift PCA. Journal of Neuroscience Methods. 2007; 165:297–305. [PubMed: 17624443]

de Cheveigné, Ad; Simon, JZ. Sensor noise suppression. Journal of Neuroscience Methods. 2008; 168:195–202. [PubMed: 17963844]

Dehaene-Lambertz G, Dupoux E, Gout A. Electrophysiological correlates of phonological processing: A cross-linguistic study. Journal of Cognitive Neuroscience. 2000; 12:635–647. [PubMed: 10936916]

Diesch E, Eulitz C, Hampson S, Ross B. The neurotopography of vowels as mirrored by evoked magnetic field measurements. Brain and Language. 1996; 53:143–168. [PubMed: 8726531]

Doelling KB, Arnal LH, Ghitza O, Poeppel D. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage. 2014; 85(Pt 2):761–768. [PubMed: 23791839]

Engel AK, Fries P. Beta-band oscillations — signalling the status quo? Current Opinion in Neurobiology. 2010; 20:156–165. [PubMed: 20359884]

Eulitz C, Lahiri A. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. Journal of Cognitive Neuroscience. 2004; 16:577–583. [PubMed: 15185677]

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nature Communications. 2014:5.10.1038/ncomms5694

Friedrich C, Lahiri A, Eulitz C. Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34:1545–1559. [PubMed: 19045992]

Friston K. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005; 360:815–815.

Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: A review of underlying mechanisms. Clinical Neurophysiology. 2009; 120:453–463. [PubMed: 19181570]

Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:20095–20100. [PubMed: 23151506]

Herholz SC, Lappe C, Pantev C. Looking for a pattern: An MEG study on the abstract mismatch negativity in musicians and nonmusicians. BMC Neuroscience. 2009; 10:42. [PubMed: 19405970]

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97:3099–3111. [PubMed: 7759650]

Kuznetsova, A.; Bruun Brockhoff, P.; Bojesen Christensen, R. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0–6. 2014. http://CRAN.R-project.org/package=lmerTest

Lahiri, A.; Reetz, H. Underspecified recognition. In: Gussenhoven, C.; Warner, N., editors. Laboratory Phonology. Vol. VII. Mouton de Gruyter; Berlin: 2002. p. 637-677.

Lahiri A, Reetz H. Distinctive features: Phonological underspecification in representation and processing. Journal of Phonetics. 2010; 38:44–59.

Lappe C, Steinsträter O, Pantev C. A Beamformer analysis of MEG data reveals frontal generators of the musically elicited Mismatch Negativity. PLoS ONE. 2013; 8:e61296. [PubMed: 23585888]

Lewis AG, Wang L, Bastiaansen M. Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? Brain and Language. 2015

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54:1001–1010. [PubMed: 17582338]

Mazaheri A, van Schouwenburg MR, Dimitrijevic A, Denys D, Cools R, Jensen O. Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. Neuroimage. 2014; 87:356–362. [PubMed: 24188814]

Morillon B, Schroeder CE. Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. Annals of the New York Academy of Sciences. 2015; 1337:26–31. [PubMed: 25773613]

Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology. 2001; 38:1–21. [PubMed: 11321610]

Näätänen R, Alho K. Mismatch negativity (MMN) - the measure for central sound representation accuracy. Audiology and Neuro-Otology. 1997; 2:341–353. [PubMed: 9390839]

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Ilvonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 1997; 385:432–434. [PubMed: 9009189]

Näätänen R, Winkler I. The concept of auditory stimulus presentation in cognitive neuroscience. Psychological Bulletin. 1999; 125:826–859. [PubMed: 10589304]

Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology. 2007; 118:2544–2590. [PubMed: 17931964]

Obleser J, Lahiri A, Eulitz C. Magnetic brain response mirrors extraction of phonological features from spoken vowels. Journal of Cognitive Neuroscience. 2004; 16:31–39. [PubMed: 15006034]

Obleser J, Weisz N. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cerebral Cortex. 2012; 22:2466–2477. [PubMed: 22100354]

Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. Neuropsychologia. 1971; 9:97–113. [PubMed: 5146491]

Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience. 2011:1–9.10.1155/2011/156869 [PubMed: 21837235]

Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. Frontiers in Psychology. 2012:3.10.3389/fpsyg.2012.00320 [PubMed: 22291676]

Peltola MS, Kujala T, Tuomainen J, Ek M, Aaltonen O, Näätänen R. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. Neuroscience Letters. 2003; 352:25–28. [PubMed: 14615041]

Perkell JS. Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. Journal of Phonetics. 1996; 24:3–22.

Peterson GE, Barney HL. Control methods used in a study of the vowels. Journal of the Acoustical Society of America. 1952; 24:175–184.

Pettigrew CM, Murdoch BE, Ponton CW, Finnigan S, Alku P, Kei J, Sockalingam R, Chenery HJ. Automatic auditory processing of english words as indexed by the mismatch negativity, using a multiple deviant paradigm. Ear and Hearing. 2004; 25:284–301. [PubMed: 15179119]

Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T. Auditory cortex accesses phonological categories: An MEG mismatch study. Journal of Cognitive Neuroscience. 2000; 12:1038–1105. [PubMed: 11177423]

Pulvermüller F, Shtyrov Y. Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. Progress in Neurobiology. 2006; 79:49–71. [PubMed: 16814448]

Pulvermüller F, Shtyrov Y, Illmoniemi RJ, Marslen-Wilson W. Tracking speech comprehension in space and time. Neuroimage. 2006; 31:1297–1305. [PubMed: 16556504]

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: 2014.

Scharinger M. Minimal representations of alternating vowels. Lingua. 2009; 119:1414–1425.

Scharinger M, Poe S, Idsardi WJ. A three-dimensional cortical map of vowel space: Evidence from Turkish. Journal of Cognitive Neuroscience. 2011; 23:3972–3982. [PubMed: 21568638]

Scharinger M, Bendixen A, Trujillo-Barreto NJ, Obleser J. A sparse neural code for some speech sounds but not for others. PLoS ONE. 2012a; 7:e40953. doi:40910.41371/journal.pone.0040953. [PubMed: 22815876]

Scharinger M, Idsardi WJ. Sparseness of vowel category structure: Evidence from English dialect comparison. Lingua. 2014; 140:35–51. [PubMed: 24653528]

Scharinger M, Monahan PJ, Idsardi WJ. Asymmetries in the processing of vowel height. Journal of Speech, Language, and Hearing Research. 2012b; 55:903–918.

Schröger E. The mismatch negativity as a tool to study auditory processing. Acta Acústica. 2005; 91:490–501.

Sharma A, Dorman MF. Neurophysiologic correlates of cross-language phonetic perception. Journal of the Acoustical Society of America. 2000; 107:2697–2703. [PubMed: 10830391]

Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I. Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. The Journal of Neuroscience. 2011; 30:13578–13585. [PubMed: 20943899]

Stevens, K. Acoustic Phonetics. The MIT Press; Cambridge, MA: 1998.

Strauß A, Kotz SA, Scharinger M, Obleser J. Alpha and theta brain oscillations index dissociable processes in spoken word recognition. Neuroimage. 2014; 97:387–395. [PubMed: 24747736]

Tervaniemi M, Radil T, Radilova J, Kujala T, Näätänen R. Pre-attentive discriminability of sound order as a function of tone duration and interstimulus interval: a mismatch negativity study. Audiology and Neuro-Otology. 1999; 4:303–310. [PubMed: 10516390]

Wang XJ. Neurophysiological and computational principles of cortical rhythms in cognition. Physiological Reviews. 2010; 90:1195–1268. [PubMed: 20664082]

Winkler I. Interpreting the Mismatch Negativity. Journal of Psychophysiology. 2007; 21:147–163.

Winkler I, Cowan N, Csepe V, Czigler I, Näätänen R. Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. Journal of Cognitive Neuroscience. 1996a; 8:403–415. [PubMed: 23961944]

Winkler I, Karmos G, Näätänen R. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Research. 1996b; 742:239–252. [PubMed: 9117400]

Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csepe V, Aaltonen O, Raimo I, Alho K, Lang H, Iivonen A, Näätänen R. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cognitive Brain Research. 1999; 7:357–369. [PubMed: 9838192]

Highlights

- We are interested in speech-sound specific top-down influences on speech recognition.
- We measured transient and oscillatory brain activity in a Magnetoencephalogram experiment.
- Less specific speech sounds caused reduced change detection responses.
- Less specific speech sounds displayed decreases in pre-stimulus beta-oscillations.
- Effects are interpreted as reduced top-down influences from less specific speech sounds.

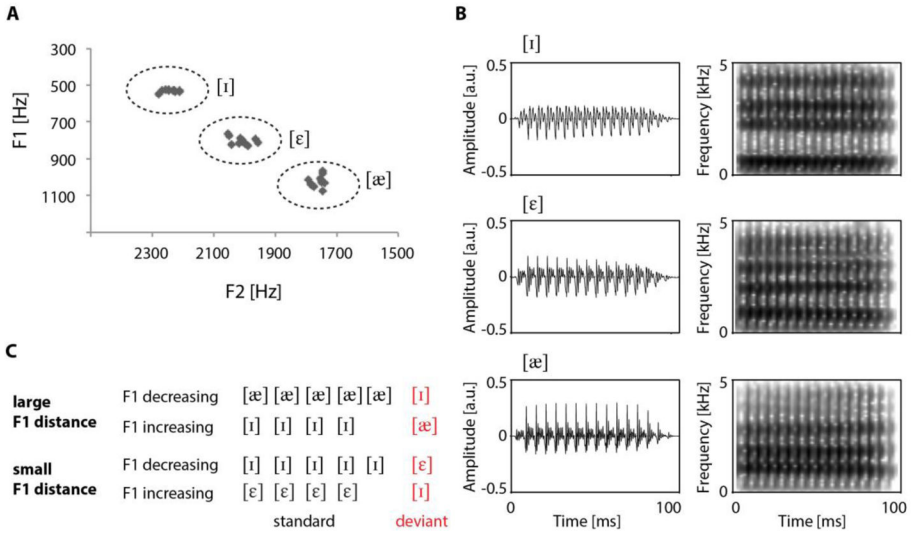

Figure 1.

A. Acoustic properties (first and second formant frequencies; F1/F2) of the vowels as obtained from a linear predictive coding (LPC) analysis. **B.** Waveform and spectrogram for a representative member of each vowel type**. C.** Illustration of passive oddball paradigm. In the large F1 distance condition, low [æ] and high [ɪ] were presented in standard position and thus set up a predictive context, violated by the respective deviants [ɪ] and [æ]. In the small F1 distance condition, high [ɪ] as standard set up a predictive context while [ɛ] created a relatively non-predictive context.

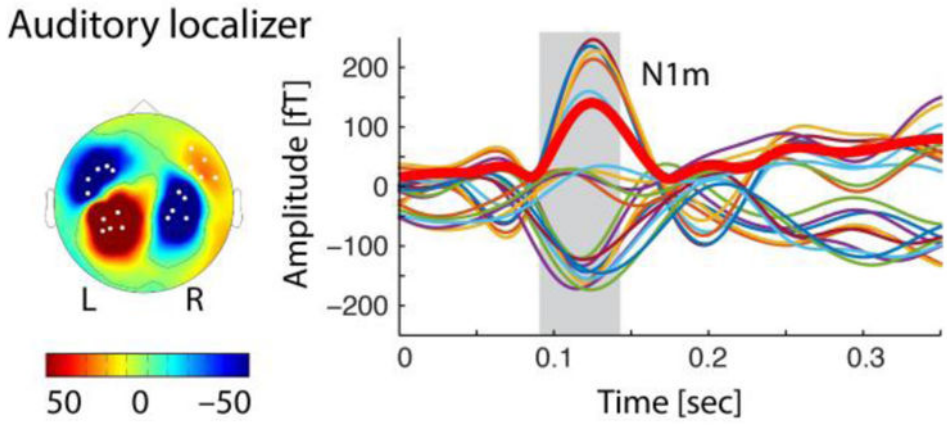

Figure 2.

Auditory localizer results from responses to 1000 Hz tones from a representative participant. The topography (left) of the auditory N1m (between 90 and 140 ms post stimulus onset, right) shows a typical source/sink distribution of the magnetic field across left posterior and anterior sites, with a reversal in the right hemisphere. Amplitudes over time (right) are illustrated for the 20 strongest channels (indicated by white dots in the topography). For illustration purposes, the thick red line represents the root-mean squared (RMS) amplitude over the 20 strongest channels.

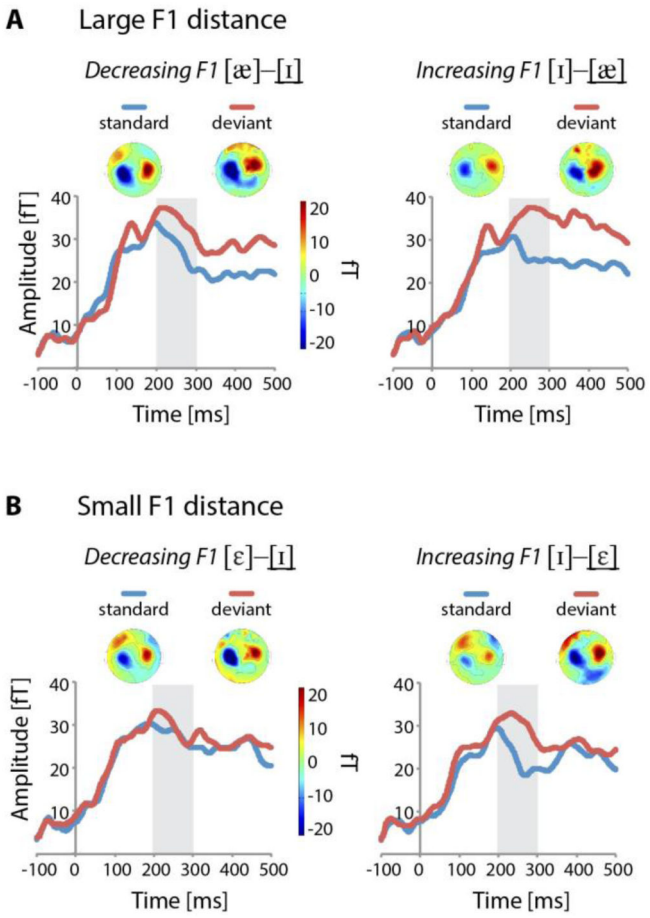

Figure 3.

Grand averaged RMS responses for standards (blue) and deviants (red). **A.** Standard and deviant responses to the vowels [æ] and [ɪ] in the large F1 distance condition. Note that standard and deviant waveforms stem from different experimental blocks as is common in identity-MMN analyses. Vowels in deviant position are underlined in the header lines. For instance, the left panel shows the deviant response to the vowel [ɪ] and the standard response to the same vowel from the reverse (increasing F1) direction. **B.** Standard and deviant responses to the vowels [ɛ] and [ɪ] in the small F1 distance condition. Sensor-level scalp topographies are provided for the time window of 200–300 ms post stimulus onset (marked in grey). RMS-amplitudes are plotted in femto-Tesla (fT, 10^{-15}T).

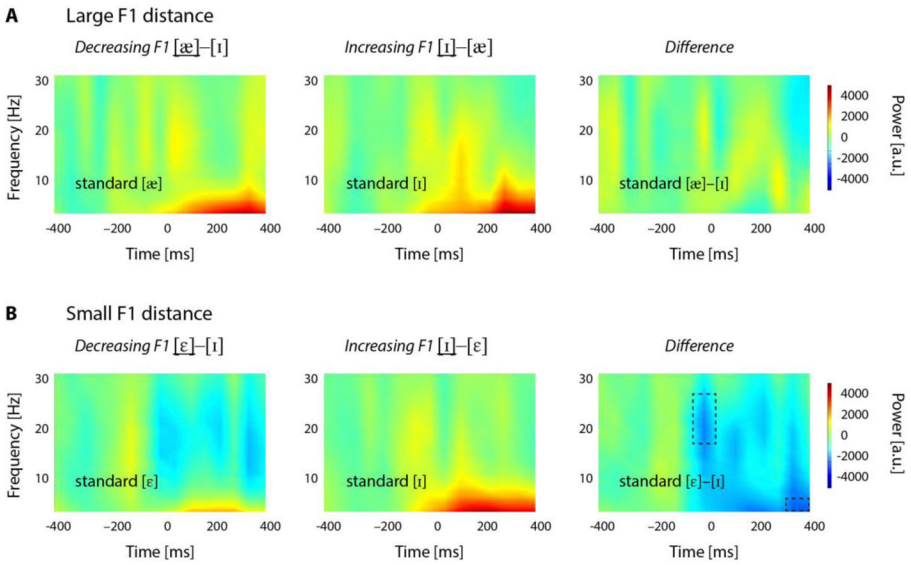

Figure 4.

Time-frequency analyses. Power values are color-coded, with warmer colors for positive power and cooler colors for negative power. **A.** Total power distribution for the standard vowels [æ] (left) and [ɪ] (middle) in the large F1 distance condition. The difference plot (right) indicates no power differences between the standards. Vowels in standard position are underlined in the header lines. **B.** Total power distribution for the standard vowels [ɛ] (left) and [ɪ] (middle) in the small F1 distance condition. The difference plot (right) indicates that there were at least two regions (marked with dashed rectangles) where the power for [ɛ] was lower than the power for [ɪ]. Power values are in arbitrary units (a.u.).

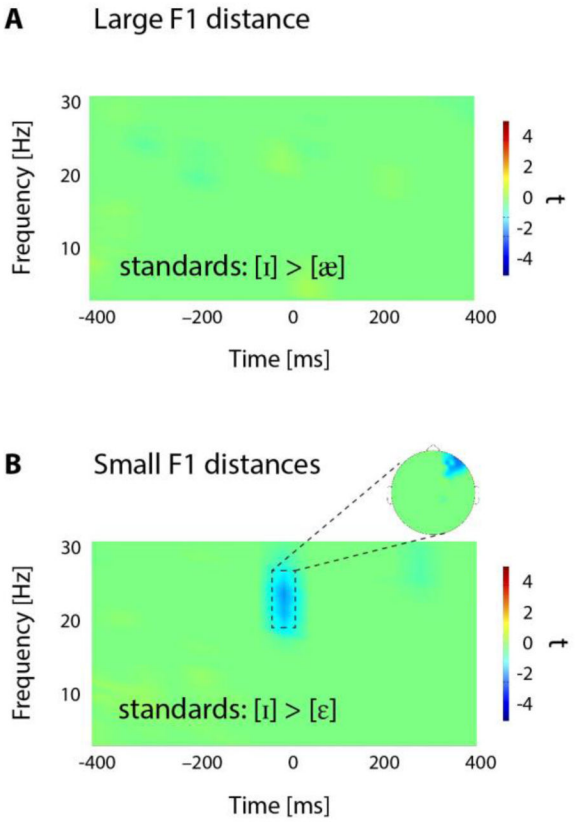

Figure 5.

Results from the multi-level statistical approach on power differences. **A.** No significant time-frequency bins are observed for the standard vowel contrasts in the large F1 distance condition. No time-frequency bin survived the cluster-permutation based statistical threshold (critical *t*-value = ±2.18). **B.** Significant time-frequency bins for the standard vowel contrasts in the small F1 distance condition were found between −50 and +20 ms and between 18 and 26 Hz, indicated by the dashed rectangle. The blue colors shows *t*-values that survived the statistical threshold. In these time-frequency bins, power was significantly lower for [ɛ] than for [ɪ].

Table 1

Acoustic properties of the stimulus materials, averaged across the 10 instances of each vowel. Formant values were obtained from a linear predictive coding (LPC) formant analysis carried out in PRAAT (Boersma and Weenink, 2011). Standard deviations are given in parentheses.

Vowel	Pitch (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
[æ]	171 (3.46)	1023 (31.60)	1761 (19.92)	2713 (126.25)
[ɛ]	177 (2.94)	801 (22.59)	2009 (32.85)	2896 (73.42)
[ɪ]	184 (2.67)	532 (6.88)	2240 (23.51)	3010 (72.49)

Table 2

Distribution of vowel stimuli over the factor levels of distance (large F1, small F1) and direction (increasing F1, decreasing F1) on the basis of standard-deviant relations. The context-column describes whether the standard is assumed to set up a predictive or non-predictive context.

Standard	Deviant	Distance	Direction	Context
[æ]	[ɪ]	large F1	decreasing F1	predictive
[ɪ]	[æ]	large F1	increasing F1	predictive
[ɛ]	[ɪ]	small F1	decreasing F1	non-predictive
[ɪ]	[ɛ]	small F1	increasing F1	predictive