Article

Asymmetries in the Processing of Vowel Height

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

Purpose: Speech perception can be described as the transformation of continuous acoustic information into discrete memory representations. Therefore, research on neural representations of speech sounds is particularly important for a better understanding of this transformation. Speech perception models make specific assumptions regarding the representation of mid vowels (e.g., [ɛ]) that are articulated with a neutral position in regard to height. One hypothesis is that their representation is less specific than the representation of vowels with a more specific position (e.g., [æ]).
Method: In a magnetoencephalography study, we tested the underspecification of mid vowel in American English. Using a mismatch negativity (MMN) paradigm, mid and low lax vowels ([ɛ]/[æ]), and high and low lax vowels ([ɪ]/[æ]), were opposed, and M100/N1 dipole source parameters as well as MMN latency and amplitude were examined.
Results: Larger MMNs occurred when the mid vowel [ɛ] was a deviant to the standard [æ], a result consistent with less specific representations for mid vowels. MMNs of equal magnitude were elicited in the high–low comparison, consistent with more specific representations for both high and low vowels. M100 dipole locations support early vowel categorization on the basis of linguistically relevant acoustic–phonetic features.
Conclusion: We take our results to reflect an abstract long-term representation of vowels that do not include redundant specifications at very early stages of processing the speech signal. Moreover, the dipole locations indicate extraction of distinctive features and their mapping onto representationally faithful cortical locations (i.e., a feature map).

Determining the representational nature of speech sounds is an integral component in furthering our understanding of the set of perceptual and neurophysiological computations that underlie the mapping between the incoming acoustic signal and lexical representations. Given their tractable nature and well-understood spectral properties (Rosner & Pickering, 1994), vowels have been exploited extensively in investigations of these processes. Typically, vowels are classified on the basis of their first two resonant frequencies (peaks of energy in an acoustic spectrum; Ladefoged, 2001; Stevens, 1998), in which the first formant frequency (F1) inversely correlates with articulatory tongue height, and the second formant

frequency (F2) reflects the place of articulation in the horizontal dimension (back to front). Given these dimensions, vowels can be distinguished on the basis of tongue height (high [i] vs. low [a]) and place of articulation (front [e] vs. back [o]; Ladefoged, 2001).

Distinctive feature theories (e.g., McCarthy, 1988; for a neurolinguistic account, see Poeppel, Idsardi, & van Wassenhove, 2008) assume that long-term memory representations of vowels are based on symbolic and abstract labels referring to both acoustic cues (F1, F2) and articulatory configurations (tongue height, place of articulation). They thereby connect audition and articulation (for contrasting perspectives, see Browman & Goldstein, 1989; Stevens, 2002). Rooted in distinctive feature theory, Lahiri and Reetz (2002, 2010) have offered a model of speech perception and lexical access (the featurally underspecified lexicon model) in which speech sounds are described as bundles of feature specifications (e.g., high, front [ɪ]) and are also processed on the basis of their abstract features. Underlying (phonological) representations employ the same feature bundles, but crucially, not every feature is expressed or specified. For example, on the basis of particular assimilatory properties of front (coronal) speech sounds (Avery & Rice, 1989), their place of articulation is assumed to be

[a]University of Maryland, College Park
[b]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[c]Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain

Correspondence to Mathias Scharinger: mscharinger@cbs.mpg.de

unspecified for coronality (underspecified for place of articulation). Note that the assumption of Lahiri and Reetz (2002), which we adopt here, is that *coronal* refers to a macro-category of place of articulation. It subsumes dental, alveolar, and alveo-palatal (or postalveolar) places of articulation. These places contrast with bilabial (labial) on the one hand and with velar and uvular (dorsal) places on the other hand. Further, the term *underspecification* refers to the overall representation of a speech sound and indicates that some feature dimension (e.g., tongue height for [ɛ]) is not specified at all. According to the featurally underspecified lexicon model, a consonant like [t] or a vowel like [ɪ] would not have a long-term memory representation for its *phonological* place of articulation feature, [coronal]. As a consequence, the mapping from acoustic signal to underlying features is asymmetric: For noncoronal sounds, bottom-up signal information is evaluated against more detailed long-term memory representations, whereas for coronal sounds, signal information is mapped onto less detailed (i.e., underspecified) memory representations.

Regarding vowel height, it has been proposed that some tongue heights are not encoded underlyingly; that is, mid [ɛ] is underspecified for height (Lahiri & Reetz, 2002, p. 657). This contrasts with the specification [high] for [ɪ] and [low] for [æ].

In this article, we investigate the representational nature of mid vowels in American English, which have the precarious status of being neither [high] nor [low] in traditional distinctive feature analyses; for example, they are characterized by the conjoined lack of a positive specification (cf. Avery & Idsardi, 2001). We use the brain's magnetic response to acoustic signals derived from contrasts in tongue height as a neurophysiological measure in a classical mismatch negativity (MMN) study (Näätänen, 2001). The aim is to examine long-term vowel representations and processes involved in translating a continuous acoustic signal into a discrete mental sound representation under the assumption that this representation does not yield a one-to-one correspondence to the physical acoustic properties of the corresponding speech sound. Although we tested American English speakers with no hearing deficits, the framework may have important clinical implications, particularly for cochlear implant (CI) patients whose vowel spaces differ in characteristic ways from those of healthy controls (Harnsberger et al., 2001; Löfqvist, Sahlén, & Ibertsson, 2010; Neumeyer, Harrington, & Draxler, 2010).

Neurophysiological Investigations of Vowel Perception

The cortical processing of vowels has been extensively investigated using electro-encephalography (Diesch,

Eulitz, Hampson, & Ross, 1996; Diesch & Luce, 1997a, 1997b; Hill, McArthur, & Bishop, 2004; Jacobsen, Schröger, & Alter, 2004; Poeppel et al., 1997; Roberts, Flagg, & Gage, 2004; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004), that is, the brain's electric response underlying the perception and processing of vowel sounds in human cortex. Because vowels have relatively robust acoustic cues in F1 and F2, they are ideal candidates for the elicitation of early auditory-evoked components, such as the N1 or its magnetic equivalent, the N1m/M100. The M100 is a pronounced peak around 100 ms after the onset of an auditory stimulus that appears to encode basic speech processing in auditory cortices (Diesch et al., 1996; Shestakova et al., 2004). It is considered an index of temporal and topographical coding. Measures on the surface of the scalp (sensor-space) as well as measures of the source estimates producing a particular scalp electric distribution (source-space) are systematically modulated by the spectral characteristics of different vowel stimuli. For (pure) sinusoidal tones, the M100 latency inversely correlates with tone frequency (Poeppel & Marantz, 2000; Roberts, Ferrari, Stufflebeam, & Poeppel, 2000 [and references therein]), whereas for vowels, the latency of the M100 seems to track F1 (Diesch et al., 1996; Poeppel et al., 1997; Roberts et al., 2000, 2004; Tiitinen, Mäkelä, Mäkinen, May, & Alku, 2005). Recent work, however, suggests an additional sensitivity to vowel formant ratios in dense vowel spaces (Monahan & Idsardi, 2010), in which larger ratios between the first and third vowel formant frequencies (F1/F3) elicit earlier M100 latencies.

Source-based parameters of the M100 are also sensitive to spectral characteristics of complex tones, vowels, and consonants (Diesch & Luce, 1997a, 1997b, 2000; Eulitz, Diesch, Pantev, & Hampson, 1995). The latter studies parallel earlier findings that showed a tonotopic organization of auditory cortex (Pantev et al., 1988; Pantev, Hoke, Lutkenhöner, & Lehnertz, 1989), that is, a spatial representation of frequency differences (frequency maps). M100 dipole locations reflect a categorical distinction of speech sounds on the basis of their formant frequencies (Mäkelä, Alku, & Tiitinen, 2003; Obleser, Elbert, Lahiri, & Eulitz, 2003; Shestakova et al., 2004). Spectral acoustic distances, as determined by Euclidean distances on the basis of two or three vowel formants, were paralleled in spatial dipole locations in a number of studies (Obleser, Elbert, et al., 2003; Shestakova et al., 2004). The crucial finding was that if two stimuli had a large spectral acoustic distance, the underlying sources of cortical activity to these stimuli had a relatively large spatial separation in auditory cortex. Conversely, the cortical sources of stimuli with a small acoustic distance were more closely collocated. Moreover, it has been found that M100 source dipoles differed along the anterior–posterior dimension according to differences in place of articulation, where [coronal]

vowels reliably localized to more anterior positions in auditory cortex than [dorsal] vowels, seemingly parallel to their oral articulation (Obleser, Lahiri, & Eulitz, 2003, 2004). This pattern was replicated with CV syllables in German that contained either a dorsal ([o]) or a coronal vowel ([ø]), irrespective of the initial consonant (Obleser, Lahiri, & Eulitz, 2003). Finally, absolute dipole locations were modulated when attention was shifted away from the linguistic characteristics of the stimulus (Obleser, Elbert, & Eulitz, 2004), whereas the relative anterior/posterior distinction was maintained. Taken together, these findings suggest that M100 source parameters are modulated by acoustic and, in particular, by linguistically relevant acoustic characteristics of speech sounds.

Neurophysiological assessments of speech sound underspecification have primarily been made on the basis of components temporally subsequent to the M100, such as the MMN, an automatic and preattentive brain response to acoustic change or rule violation (Näätänen, 2001; Näätänen & Alho, 1997; Näätänen, Paavilainen, Rinne, & Alho, 2007; Winkler, 2007). The response is usually elicited in passive oddball paradigms, during which participants listen to frequent (standard) stimuli interspersed with infrequent (deviant) stimuli. The MMN is a fronto-temporal negativity in the difference waveform obtained by subtracting standard responses from deviant responses to the same acoustic stimuli and peaks between 150 and 250 ms poststimulus onset.

Designs testing coronal underspecification exploit the opposition of frequently occurring standard stimuli and rarely occurring deviant stimuli on the basis of the following logic: Standard repetitions activate memory traces, corresponding to underlying representations of speech sounds (Näätänen et al., 1997), affixes (Shtyrov & Pulvermüller, 2002a), or words (Pulvermüller et al., 2001; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004; Shtyrov & Pulvermüller, 2002b) in long-term memory. Standard repetitions set up expectations regarding feature specifications that might be violated by rarely occurring deviants with conflicting features (Eulitz & Lahiri, 2004; Winkler et al., 1999). Conflicting features are those that are mutually exclusive, such as [high] versus [low] and [coronal] versus [dorsal]. It should be noted that acoustic differences between standards and deviants will always yield some electrophysiological responses (Eulitz & Lahiri, 2004) given a minimum acoustic distance (cf. Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). However, if acoustic differences are accompanied by featural differences (i.e., featural mismatches), the resulting MMN is expected to be larger than if there are no featural deviances. That is, the observed MMN is the result of a combination of differences from acoustic (bottom-up) information and from more abstract (top-down) distinctions.

Testing German vowels, Eulitz and Lahiri (2004) found that a coronal deviant [ø] that is preceded by a dorsal standard [o] elicited earlier and larger MMN responses than the reverse presentation, that is, a dorsal deviant [o] preceded by a coronal standard [ø]. These findings are predicted if the dorsal standard activates its fully specified place of articulation feature [dorsal], whereas the underspecified deviant fails to satisfy the prediction that a place of articulation feature is specified. Conversely, a coronal standard activates an underspecified representation in long-term memory, for which the dorsal deviant does not provide a featural mismatch. Further, underspecified standards do not generate specific predictions as to the specification of place of articulation. Thus, the observed MMN reflects a combination of these two difference measures, one pertaining to the acoustic difference between [o] and [ø] (and equivalently vice versa, [ø] and [o]) and the other one to the phonological difference between the two vowels (i.e., specified vs. underspecified). The MMN elicited by the deviant [o] was relatively small because it only reflected the acoustic dissimilarity to the standard, whereas the larger MMN elicited by the deviant [ø] additionally reflected a featural mismatch (see Bergelson & Idsardi, 2009, for similar results from music perception; see Hwang, Monahan, & Idsardi, 2010, for asymmetries in voicing of consonants).

The similar featural mismatch logic should hold for tongue height differences in English front vowels. We follow Lahiri and Reetz (2002) in the assumption that mid vowels are entirely underspecified for height, that is, never provide mismatching features to their low or high vowel counterparts. However, as before, a fully specified low or high vowel in standard position should generate a strong expectation regarding tongue height specification that might be violated if the deviant to this standard sequence is an underspecified mid vowel. Hence, in an MMN design, the mid vowel [ɛ], preceded by the low vowel [æ], should elicit a larger MMN response than in the reverse case, that is, if the low vowel [æ] is preceded by the mid vowel [ɛ]. In the former (but not the latter) case, the standard generates a strong prediction regarding its tongue height specification that is then violated by the deviant. Further, assuming a low vowel standard and a high vowel deviant would lead to a featural mismatch between [low] and [high]. This mismatch, however, should be observed in the reverse case as well.

The hypotheses discussed above are compatible with the *model adjustment hypothesis* (Näätänen & Winkler, 1999; Winkler, Karmos, & Näätänen, 1996) as well as the *predictive coding* approach to the neural source of MMN generation at frontal and temporal cortical locations (Baldeweg, 2006; Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009). Note that the model adjustment hypothesis is part of the predictive coding

account. In both approaches, the MMN results from violations of inference predictions on the basis of the repetitious standard presentations. The predictive coding account, combining the model adjustment approach with an adaptation approach of MMN generation (Garrido et al., 2009), proposes that standard presentation resembles perceptual learning during which hierarchical sensory levels receive bottom-up sensory input from lower levels and receive top-down predictions from higher levels. As a result of the repetition of standard presentations, prediction errors are reduced by repetitive suppression or adaptation. A deviant presentation then leads to a violation of bottom-up prediction that is reflected in MMN generation. Applied to vowel underspecification, underspecified standards (such as mid [ɛ]) make weaker predictions with regard to their tongue height features. As a result, the violation of such an expectation from deviant sensory information should be less severe, and the MMN ought to be reduced compared with the reverse condition, in which the underlying, specified representation of the standard can generate a stronger featural prediction.

We tested these predictions in a magnetoencephalography (MEG) MMN experiment with the American English short (lax) front vowels [æ], [ɛ], and [ɪ]. The choice of these vowels was determined by the observation that short front vowels in American English are less likely to be diphthongized (Hillenbrand, Getty, Clark, & Wheeler, 1995), such that our featural assumptions are more straightforward. For instance, it is not clear how to specify the diphthong [ai] in terms of tongue height features: It could subsume [low] and [high], [low] only, [high] only, or neither feature specification.

Testing Vowel Height Differences in American English
Rationale

Using an MMN design with MEG enables us to test our featural mismatch predictions and to simultaneously investigate M100 source parameters. These measures enable us to examine the neural bases of speech sound category access during the transformation of an acoustic signal into a mental representation. Note that MEG allows us to estimate the underlying source dipole of the observed magnetic field at the scalp with greater accuracy than EEG (e.g., Leahy, Mosher, Spencer, Huang, & Lewine, 1998; Lopes da Silva, Wieringa, & Peters, 1991; Nakasatp et al., 1994).

Materials

Twenty acoustically distinct exemplars of each of the American English vowels [æ], [ɛ], and [ɪ] were produced by a female native speaker of American English with phonetic training. This dialect and this speaker did not merge any of these vowels (cf. Labov, Ash, & Boberg, 2006) and made a robust three-way height distinction. Detailed spectral acoustic measures are provided in Table 1.

Note that the motivation of using multiple acoustic vowel exemplars is twofold: First, we want to guarantee a more natural listening situation. Second, using single vowels without any acoustic variation may result in response patterns that solely depend on the particular acoustic quality of the stimulus. In contrast, multiple exemplars in a many-to-one oddball paradigm increase the likelihood that the participants will form more abstract memory traces against which the deviants can be compared (Phillips et al., 2000; Winkler et al., 1999).

All stimuli were recorded and digitally sampled at 44.1 kHz with an amplitude resolution of 16 bits within the phonetic sound application PRAAT (Boersma & Weenink, 2009). Ten vowels of each category with similar pitch and intensity were selected as experimental stimuli. Their F1 and F2 values confirmed the three-way height distinction (see Figure 1). Stimuli intensity was normalized and corresponded to a presentation level of 60 dB (sound pressure level). An onset and offset ramp (cosine-square) of 15 ms was applied to all vowels (duration: 100 ms) that reduced sharp onsets and offsets.

Table 1. Spectral acoustic characteristics of the three vowel types used in the mismatch negativity experiment.

Vowel	Pitch (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
[æ]	171.25 (3.46)	1023.30 (31.60)	1760.60 (19.92)	2712.60 (126.25)
[ɛ]	177.19 (2.94)	801.00 (22.59)	2008.80 (32.85)	2895.80 (73.42)
[ɪ]	184.31 (2.67)	531.50 (6.88)	2239.90 (23.51)	3009.50 (72.49)

Note. Pitch and formant values were obtained from a linear predictive coefficient formant analysis within the sound application PRAAT (Boersma & Weenink, 2009). The table shows the averaged values across the 10 variable tokens per vowel category. Standard deviations are given in parentheses. F1 = first formant frequency; F2 = second formant frequency; F3 = third formant frequency.

Figure 1. Acoustic characteristics of test stimuli. Formant (resonance) frequencies of the first two components (first formant frequency [F1], second formant frequency [F2]) are given in Bark, a nonlinear scale that better matches the psycho-acoustic perception of frequency. All vowel types clearly differ in F1.
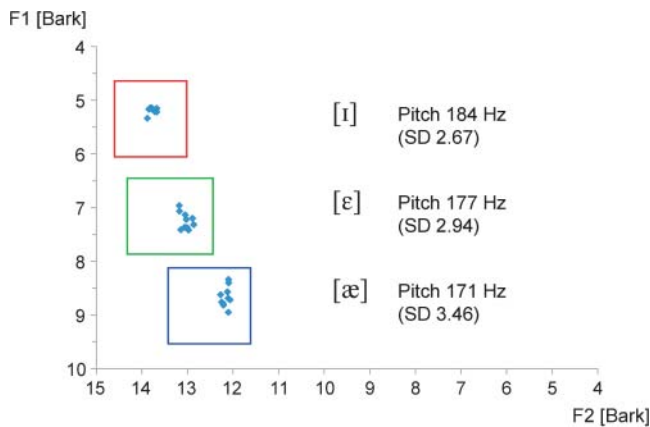
Design

Stimuli were organized in a passive standard/deviant many-to-one oddball paradigm (Phillips et al., 2000; Winkler et al., 1999). The crucial mid vowel comparison was made between the low vowel [æ] and mid vowel [ɛ]. As a control, we included the standard–deviant pair of the low vowel [æ] and the high vowel [ɪ], both of which are underlyingly specified for height. The vowels [æ] and [ɛ], as well as [æ] and [ɪ], were distributed over 2×2 blocks in which they occurred in either a standard ($p = .875, N = 700$) or a deviant ($p = .125, N = 100$; see Table 2) position. The different category exemplars were pseudo-randomly drawn from the corresponding sets. We ensured that all exemplars occurred equally often. The

number of standards between two deviants varied randomly in each block. The interstimulus intervals pseudorandomly varied between 500 and 1,000 ms to prevent participants from entraining to a specific presentation rhythm that may confound our event-related potential measures of interest. Each block contained a total of 800 trials and lasted approximately 15 min. Block order was permutated across participants. Trials were presented binaurally at a comfortable listening level (~60 dB SPL) with the software package Presentation (Neurobehavioral Systems, 2012).

Expectations

Following Eulitz and Lahiri (2004), we predict asymmetric MMN responses to the deviants [ɛ] and [æ]. This prediction is based on the assumption that [ɛ] is underspecified for tongue height, whereas [æ] is not. As laid out above, underspecified deviants after specified standards ought to elicit a larger MMN response than in the reverse case. In contrast, the test conditions with [æ] and [ɪ] should show similar MMN responses because the standard/deviant opposition always includes the specified features [low] and [high], or [high] and [low]. Furthermore, we also expect the MMN response to be modulated by the size of the acoustic differences as well, such that the larger formant distance between [low] and [high] vowels should enhance their MMN response.

On the basis of previous M100 findings, we additionally expect source- and sensor-space differences between the M100 responses of all three vowels (e.g., Obleser, Lahiri, & Eulitz, 2003). This would provide further evidence for an early extraction of linguistically relevant acoustic information to access discrete long-term memory representations of speech sounds (here: vowels).

Participants and Procedure

Fifteen students from the University of Maryland (9 women, 6 men; mean age = 21.1 years, SD = 3.2; all native speakers of American English), without any reported history of hearing or neurological problems, participated for class credit or monetary compensation. All participants provided informed written consent and tested strongly right-handed (>80%) on the Edinburgh Handedness Inventory (Oldfield, 1971). Participants' head shapes were digitized before the experiment with a POLHEMUS 3 Space Fast Track system. Together with localization data from two preauricular and three prefrontal electrodes, these data allowed us to perform dipole localization analyses, as reported in the Data Analysis section.

During MEG recording and stimulus presentation, participants lay supine in a magnetically shielded

Table 2. Design of experiment and distribution of test stimuli.

Block	Standard	Deviant	Feature opposition	Expectancy violation
1	[æ]	[ɛ]	[low]–[—] specified–unspecified	✔
2	[ɛ]	[æ]	[—]–[low] unspecified–specified	✗
3	[æ]	[ɪ]	[low]–[high] specified–specified	✔
4	[ɪ]	[æ]	[high]–[low] specified–specified	✔

Note. The right-most column indicates whether the expectation (derived from the standard) is violated in the deviant. Note that the features [high] and [low] in Blocks 3 and 4 violate the expectancy by a featural mismatch (cf. Lahiri & Reetz, 2002).

chamber. Magnetic fields were recorded by a whole-head, 157 axial-gradiometer MEG system (Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 500 Hz. Auditory stimuli were delivered binaurally via Etymotic ER3A insert earphones. Earphones 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 resonance frequencies (Stevens, 1998). An online, 200-Hz, low-pass filter and a 60-Hz notch filter were applied to the raw data.

Prior to the main experiment, participants took part in a tone perception pretest that served as a control for dipole estimations as well as the basis for determining regions of interest on the surface of the scalp. Note that pure sinusoidal tones have been shown to engage neural circuitry in the auditory cortices, reflecting magnetic activity to the left and right temporal scalp areas. Further, the (base) frequency of sinusoidal tones parametrically determines the center of neural activity because of the tonotopic principle of auditory cortex (Pantev et al., 1988, 1989, 1995).

During the tone pretest, participants were instructed to silently count high (1000 Hz) and low (250 Hz) sinusoidal tones (~300 total) presented over headphones in a pseudorandom order. The scalp distribution of the resulting averaged evoked M100 field was consistent with the typical M100 source in the supratemporal auditory cortex (Diesch et al., 1996). Only participants with a reliable bilateral M100 response were included in further analyses, resulting in the exclusion of one participant.

For the main experiment, participants passively listened to vowel stimuli presented in blocks, as illustrated in Table 2. The use of acoustically variable standard vowels (10 tokens of each vowel) ensured that standards activated more abstract representations, such that a pure acoustic explanation of the resulting MMN can be excluded (Phillips et al., 2000). Block order was counterbalanced between subjects, and there was always a short break after each block. Participants viewed a silent movie during the passive listening task to reduce excessive eye movements and to maintain an awake state (Tervaniemi et al., 1999). The movie was projected onto a screen approximately 15 cm above the participants. Together with the head shape construction and the tone pretest, the entire experiment lasted approximately 90 min.

Data Analysis

Environmental and scanner noise were removed from the MEG raw data using a multishift, principal components analysis, noise-reduction algorithm (de Cheveigné & Simon, 2007, 2008). Epoch averaging used a 100-ms prestimulus interval and a 500-ms poststimulus

interval. This allowed us to look at the first 500-ms poststimulus onset during which we expected the two event-related potential measures of interest: M100 and MMN. Epochs were baseline corrected using the 100-ms prestimulus epoch, guaranteeing that effects were truly stimulus based and not elicited by events prior to stimulus presentation. Although we used a denoising algorithm prior to averaging to reduce environmental magnetic influences, participants always produced artifacts by eye movements or muscle activity that masked our components of interest. For this reason, we rejected certain artifacts by visual inspection on the noise-reduced continuous data. Epochs were rejected if amplitudes were higher than 3 pico-Tesla (10^{-12} Tesla) or contained more than three consecutive eye blinks. Because of excessive noise and artifacts in the raw data, leading to exclusion rates of >15% of standards and deviants, one participant was excluded from further analyses (leaving a total of 13 participants for the analyses). Otherwise, no more than 15% of standards or deviants were excluded in any of the remaining participants. Averaged data were baseline corrected as elucidated above and were band-pass filtered by a Hamming-window digital filter with frequency cutoffs at 1 Hz and 30 Hz. The 10 strongest channels from the tone pretest were selected for subsequent amplitude and peak latency analyses separately for each hemisphere. Channels stemmed from scalp areas where the magnetic field was oriented toward the cortex (sink) and areas where the magnetic field was oriented emerging from the cortex (source). Following the usual time course of auditory-evoked MEG components (Ackermann, Hertrich, Mathiak, & Lutzenberger, 2001) and a visual inspection of the grand average waveform, several time windows with a length of 50 ms were selected. The first window (80–130 ms) covered the M100, whereas the three later windows (150–200, 200–250, and 250–300 ms) accounted for the MMN (see Figure 2). These windows were selected on the basis of previous findings (Näätänen et al., 2007; Winkler, 2007) regarding an MMN latency range between 150 and 300 ms and a particular latency difference in a window from 150 to 200 ms, which has been found to be crucial for testing underspecification (Eulitz & Lahiri, 2004).

Root-mean-squared magnetic field strengths over the selected 10 left- and right-hemispheric channels were analyzed separately for each time window of interest in mixed effect models with *subject* as random effect (Baayen, 2008; Pinheiro & Bates, 2000). M100 latencies were determined by visual inspection of the averaged epochs per condition and subject. MMN latencies were calculated from the difference waveforms of acoustically identical stimuli in deviant and standard positions (i.e., deviant – standard). Because of the few-to-many design of the oddball paradigm, we randomly selected 100 standard

Figure 2. Time course of auditory-evoked components and their topography elicited by a vowel deviant from a representative participant. MMN = mismatch negativity; fT = femto Tesla.
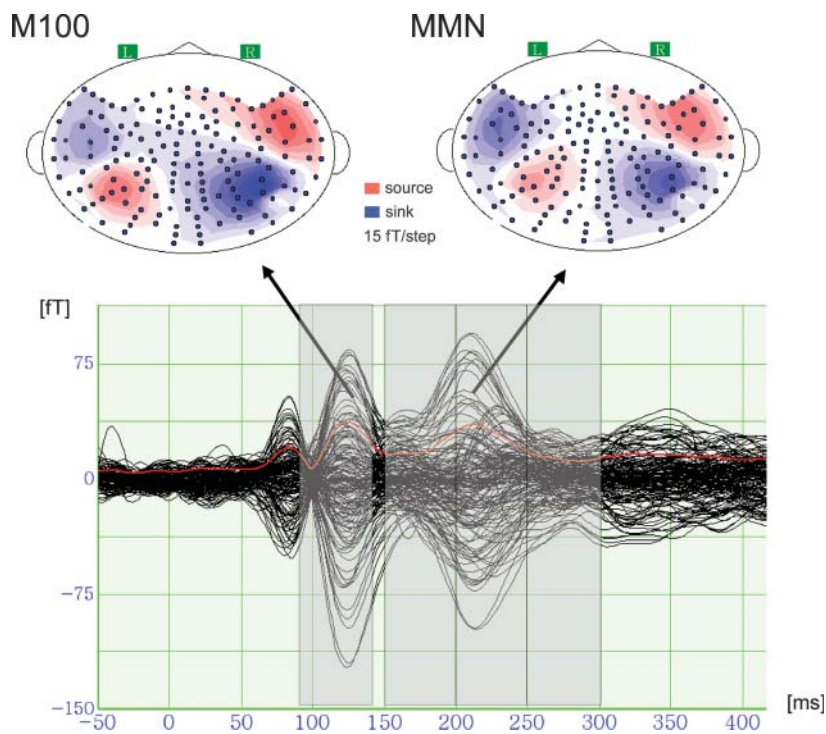
epochs to guarantee that the root-mean-squared values did not involve unequal variance between the standard and deviant stimuli.

Dipole Fitting

The fitting of equivalent current dipoles (ECDs) followed the procedure described in Obleser, Lahiri, and Eulitz (2004). An ECD is the simplified cortical source of the M100 response. It corresponds to the activity of some 10,000 neurons in auditory cortex, producing a local field potential and accompanying magnetic activity, and it can be considered the center of cortical activity in response to an auditory stimulus. To determine the cortical location of ECDs elicited by our vowel stimuli, we first defined an orthogonal left-handed head frame, on the basis of the dimensions x and z; x projected from the inion through to the nasion, and z projected through the center-midline location according to the 10–20 system. The x coordinates defined the lateral–medial dimension, the y coordinates defined the anterior–posterior dimension, and the z coordinates defined the superior–inferior dimension. Then, a sphere, whose center position and radius were calculated in head frame coordinates, was fit for each participant covering the entire surface of his/her digitized head shape. A single ECD model in a spherical volume conductor was used for source modeling analysis

of the neuromagnetic data (cf. Diesch & Luce, 1997a; Sarvas 1987). Left- and right-hemispheric dipoles were modeled separately (Sarvas, 1987). Tone and vowel source parameters were calculated from the median of the five best ECD solutions (minimally 90% goodness-of-fit) on the rising slope of the M100. Fittings after the respective peaks were not included (cf. Scherg, Vajsar, & Picton, 1990).

Results

Tones

The latency analysis (in ms) for the tone pretest included the effects *hemisphere* (left/right), *frequency* (250 Hz/1000 Hz), and the *Hemisphere × Frequency* interaction. There was a main effect of *hemisphere,* $F(1, 36) = 11.06, p < .001$, and *frequency,* $F(1, 36) = 20.91, p < .001$. The M100 in the right hemisphere peaked approximately 5 ms earlier than left-hemispheric M100, and 1000-Hz tones elicited a 6-ms earlier M100 than 250-Hz tones. Both findings are consistent with previous results (e.g., Roberts, Ferrari, & Poeppel, 1998; Roberts & Poeppel, 1996). The peak amplitude analysis (in femto Tesla [fT]; 10^{-15} Tesla), using the same model, showed no significant effects (all Fs < 2).

Tone source parameters included location in the lateral–medial, anterior–posterior, and superior–inferior

dimensions (in millimeters). ECDs distinguished the high-frequency tone with a more medial location from the low-frequency tone with a more lateral ECD location (cf. Pantev et al., 1989). The spatial difference of approximately 4 mm was marginally significant, $F(1, 34) = 3.73$, $p = .06$, and was consistent with the expected spacing based on Pantev et al. (1989).

Vowels

M100 analysis. The M100 analyses were composed of the fixed effects *vowel* ([æ]$_{/ɛ/}$, [æ]$_{/ɪ/}$, [ɛ]$_{/æ/}$, and [ɪ]$_{/æ/}$; the subscript denoting the context provided by the standard), *position* (standard/deviant), and *hemisphere* (left/right) in a full-factorial design. M100 amplitudes showed a main effect of *position*, $F(1, 180) = 18.21$, $p < .001$, reflecting larger amplitudes for deviants than for standards. Additionally, there was a marginally significant *Vowel × Position* interaction, $F(3, 180) = 2.31$, $p = .07$. In the low/mid comparison, M100 amplitudes differed between deviant and standard only for [ɛ] ($t = 3.20$, $p < .01$) but not for [æ] ($t = 0.14$, $p = .89$). Put differently, if [ɛ] was a deviant preceded by the standard [æ], the M100 amplitude was significantly larger compared with the M100 elicited by [ɛ] in standard position. In contrast, the deviant [æ] preceded by [ɛ] did not elicit a larger M100 compared with the corresponding M100 of [æ] as a standard.

First MMN window (150–200 ms). Turning now to the analysis of the latency and amplitude of the MMN component, amplitudes did not differ in the first MMN window (150–200 ms). The MMN peak latency analysis with the effects *hemisphere* (left/right) and *opposition* (low/high vs. low/mid) showed a significant effect of *opposition*, $F(1, 88) = 13.71$, $p < .01$, reflecting earlier MMN peak latencies (~10 ms) for the low/high contrast as compared with the low/mid contrast. Within the low/mid condition, latencies were earlier for the deviant [ɛ] than for the deviant [æ], but this difference (~6 ms) did not reach statistical significance, $t(84) = 1.49$, $p = .13$.

Second MMN window (200–250 ms). The amplitude analysis for the second MMN window (200–250 ms) revealed a significant effect for *position*, $F(1, 180) = 34.96$, $p < .001$, and a *Position × Vowel* interaction, $F(3, 180) = 3.16$, $p < .05$. On average, deviant amplitudes were 12 fT higher than standard amplitudes (see Figure 3). The *Position × Vowel* interaction was driven by larger differences between standards and deviants if [æ] was the standard and [ɛ] was the deviant than in the reverse case, that is, if [ɛ] was the standard and [æ] was the deviant ($t = 2.44$, $p < .01$). This difference did not hold for either [æ] or [ɪ] in the low/high contrast ($t = 1.26$, $p = .21$); that is, the MMN was of similar size for the low/high contrast in both directions. Peak latencies did not differ in the second MMN window (all Fs < 1).

Third MMN window (250–300 ms). The third MMN window (250–300 ms) showed a similar amplitude pattern. Deviants elicited a 10 fT larger amplitude than standards, $F(1, 180) = 44.87$, $p < .001$. Again, *position* interacted with *vowel*, $F(3, 180) = 3.00$, $p < .05$, reflecting the amplitude asymmetry between the low/mid condition, in which standard–deviant differences were significantly larger if [ɛ] was the deviant than if [æ] was the deviant ($t = 2.78$, $p < .01$). No asymmetry was observed for the low/high conditions ($t = 0.18$, $p = .86$). As before, peak latencies did not differ in this MMN window. The results remain consistent when the second and third MMN windows are collapsed (i.e., 200–300 ms).

ECD Source Analysis

ECD source parameters for the M100 responses across participants in the lateral–medial dimension showed a significant *Position × Hemisphere* interaction, $F(2, 93) = 13.88$, $p < .05$. In the left hemisphere, deviants elicited ECDs that were 6 mm more medial to their standards. Dipoles in the anterior–posterior dimension differed between vowels in deviant position, $F(2, 93) = 3.66$, $p < .05$. In particular, [ɪ] elicited a dipole that was located 7.5 mm more anterior to the dipole of either [ɛ] or [æ] ($t = 2.29$, $p < .05$). Dipole locations for the deviants [ɛ] and [æ] did not differ ($t = 0.60$, $p = .55$). The superior–inferior dimension analysis showed a main effect of *vowel*, $F(2, 93) = 6.13$, $p < .01$. The dipoles for [ɪ] were approximately 7 mm more superior to the dipoles for [æ] ($t = 3.42$, $p < .001$). Locations between [ɪ] and [ɛ] ($t = 1.74$, $p = .09$) and between [ɛ] and [æ] ($t = 1.30$, $p = .20$) did not differ.

Dipole orientations differed in the horizontal dimension (i.e., deviating from the lateral–medial axis). This was reflected in a main effect of *hemisphere*, $F(1, 93) = 7.65$, $p < .01$. Dipoles on the right hemisphere were oriented more toward frontal areas than on the left hemisphere. Additionally, we found a significant *Vowel × Position* interaction, $F(3, 93) = 4.41$, $p < .05$. Interestingly, ECDs were oriented more toward frontal regions for the deviant [æ] if preceded by the standard [ɛ], whereas the deviant [ɛ] did not differ from deviant [ɪ].

Figure 4 illustrates ECD locations in the left hemisphere. Because the M100 source estimates appear to reflect the vowels' acoustic properties, we were interested in relating these measures. Therefore, we calculated Euclidean distances in the three-dimensional vowel space on the basis of the first three formants and related this measure to Euclidean distances in cortical space on the basis of lateral–medial, anterior–posterior, and inferior–superior locations. To obtain a correlation measure, we used a fixed effect model with the effects *hemisphere* (left/right), *position* (standard/deviant), and *acoustic distance* in a fully factorial design. The fixed effect *acoustic distance* was significant, $F(1, 31) = 4.21$, $p < .05$, showing

Figure 3. Root-mean-squared grand average waveforms for standards and deviants (across the left and right hemispheres). For the purpose of illustration, a low-pass filter of 15 Hz (instead of 30 Hz) has been applied to the averaged data. Topographies are plotted on the basis of the difference waveforms between deviants and standards at their maximum amplitudes (~180–230 ms).
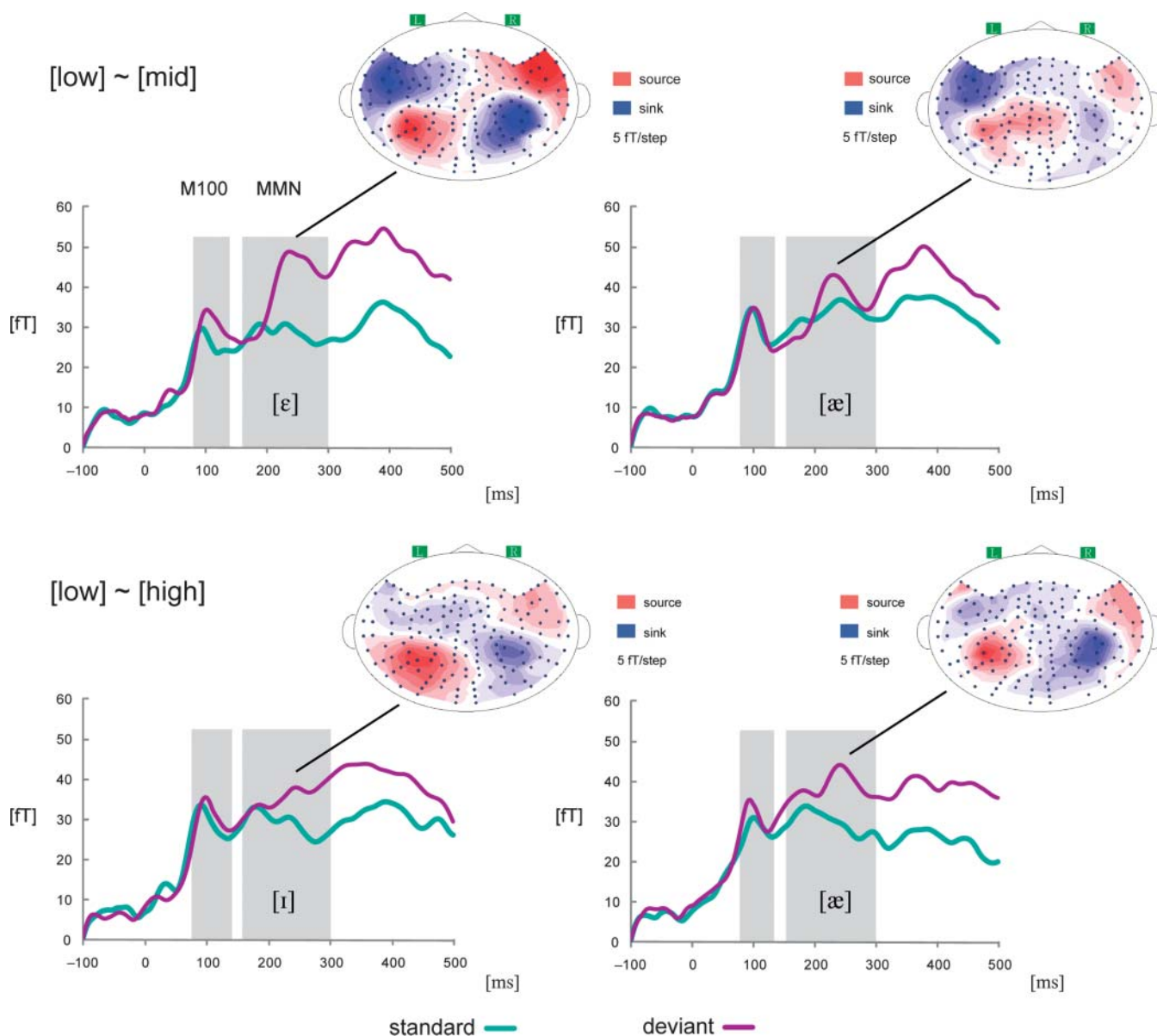
that acoustic vowel distances were reflected in cortical ECD distances, as reported previously (e.g., Obleser, Lahiri, & Eulitz, 2004). The relation of feature distance, dipole distance, and acoustic distance is illustrated in Table 3.

On the basis of statistical model comparison and model criticism (cf. Pinheiro & Bates, 2000), we calculated a second fixed effect model to which we added the fixed effect *feature difference* (1 vs. 2; see Table 3). Adding additional fixed effects to such models may result in a better fit to the data. The goodness-of-fit is usually indicated by

an information-theoretic score (such as the Akaike information criterion [AIC] or the Bayesian information criterion [BIC]). A reduction in these scores means that the addition or removal of a fixed effect or interaction results in a better fit to the data beyond what would normally be expected by adding additional parameters. Our second model with the additional effect *feature difference* resulted in lower AIC and BIC scores (AIC: 337.69 vs. 326.97; BIC: 355.06 vs. 345.82). For this reason, we conclude that the second model (with the additional featural effect) should be preferred. Adding featural phonological

Figure 4. Equivalent current dipole locations, illustrated for deviants in the left hemisphere. *Panel A:* Locations in the two-dimensional plane determined by the lateral–medial and inferior–superior axes. *Panel B:* Locations in the two-dimensional plane determined by the posterior–anterior and inferior–superior axes. Standard errors of the means (±) are indicated by whiskers.
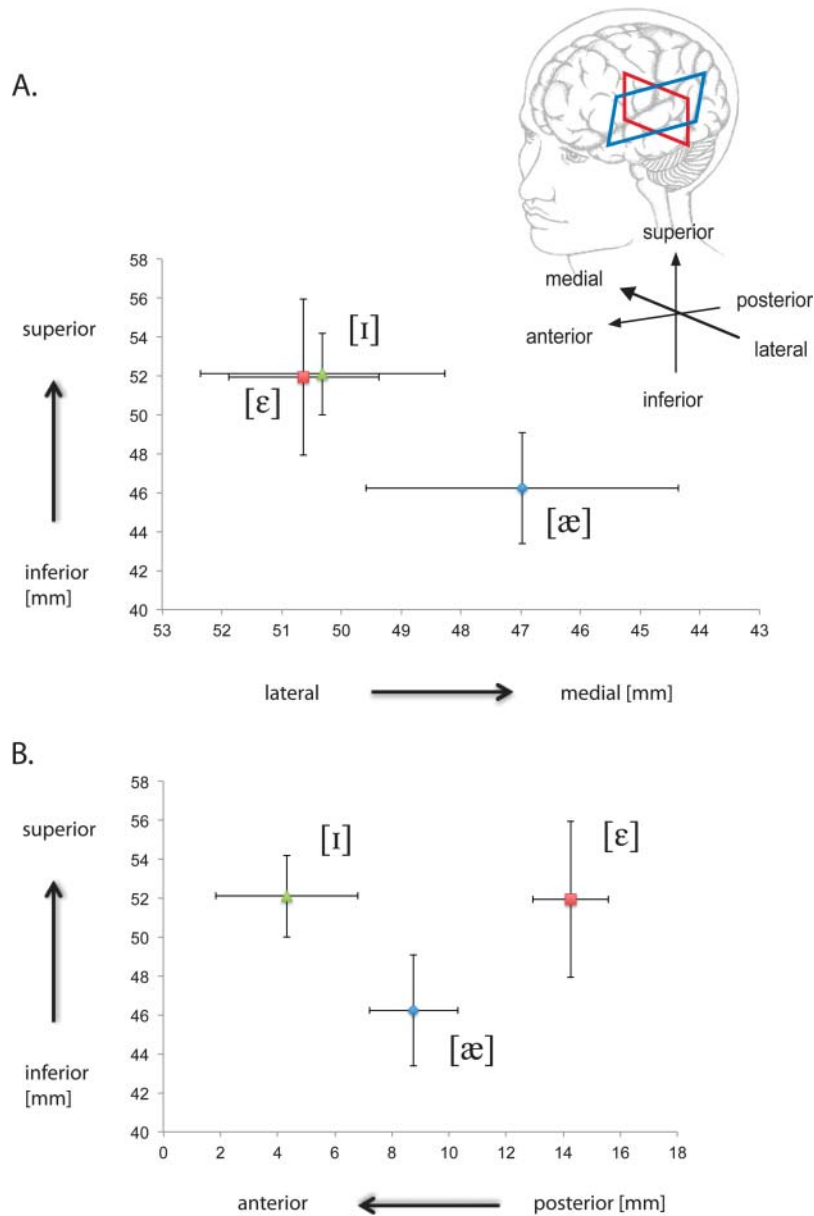

Table 3. Overview of feature difference, dipole distance, and acoustic distance.

Feature difference	Dipole distance (mm)	Acoustic distance (Hz)	Example
1	11.03 (0.96)	375.81 (17.69)	[æ]~[ɛ]
2	14.04 (1.28)	748.16 (15.65)	[æ]~[ɪ]

Note. Low–high (e.g., [æ] vs. [ɪ]) equals a feature distance of 2, whereas low–mid (e.g., [æ] vs. [ɛ]) and mid–high (e.g., [ɛ] vs. [ɪ]) equal a feature distance of 1. Standard errors of the means are given in parentheses.

information to the model results in a more accurate fit to the data, suggesting that ECD distances are determined by acoustic properties that bear specific importance for the corresponding long-term featural memory representations of the vowels. That is, even early cortical responses are composite, deriving from both information directly from the incoming signal and the influence of expectations based on long-term memory representations.

General Discussion

From our neuromagnetic experiment on the representational nature of front mid vowels in American English, we report two main findings:

1. The M100 as index of cortical processing of acoustic stimuli provided us with source- and sensor-space measures that suggest an early processing of vowels on the basis of their spectral properties that are particularly relevant for linguistic categorization.

2. The MMN as automatic change detection response of the brain supported our claim for mid vowel underspecification in showing asymmetric change responses between low/mid and low/high vowels.

With respect to our M100 findings, the early extraction of specific acoustic structures even in variable contexts has been known for quite some time (e.g., Saarinen, Paavilainen, Schröger, Tervaniemi, & Näätänen, 1992). The physiology of auditory cortex is particularly adept at encoding acoustic distinctions by spatially distinct centers of activation (Pantev et al., 1988, 1989), as measured by M100/N1 dipole source modeling. In speech, this so-called *tonotopic* principle extended to formant frequency distinctions (Diesch et al., 1996; Diesch & Luce, 1997a; Eulitz et al., 1995; Mäkelä et al., 2003; Shestakova et al., 2004), particularly for F2, encoding place of articulation (Obleser, Lahiri, & Eulitz, 2003, 2004). Obleser and colleagues (Obleser, Elbert, & Eulitz, 2004; Obleser, Lahiri, & Eulitz, 2003, 2004) found that ECDs, underlying the surface field pattern of the M100, spatially distinguished coronal (front) and dorsal (back) vowels and consonants. The difference in centers of activation was found in the anterior–posterior dimension: Front segments elicited dipoles anterior to the dipoles elicited by back segments. This is in line with Ohl and Scheich (1997) and Diesch and Luce (1997a), who have argued that this alignment is a F2–F1 difference map perpendicular to the tonotopic gradient along the medial–lateral axis (Pantev et al., 1989). Obleser, Lahiri, and Eulitz (2004) have argued that the ECD location pattern in their study does not support a linear F2–F1 M100 mapping. It rather suggests that the mapping is better accounted for by the assumption of more abstract, yet acoustically based, features. Vowel distances appear to be warped toward perceptually salient properties and not pure acoustic F2–F1 differences. This

is also reminiscent of Stevens's quantal theory of speech production and perception, in which distinctive features express stable acoustic patterns based on salient articulatory configurations (Stevens, 1989; Stevens & Blumstein, 1978). Note that saliency of articulatory configurations appears to be particularly relevant for CI users. Whereas it has been observed that the vowel space of CI users is commonly reduced (as measured in the F2/F1 plane; cf. Harnsberger et al., 2001; Löfqvist et al., 2010), Neumeyer et al. (2010) observed that this reduction mainly affects acoustic dimensions for which there are no clearly visible (hence, salient) articulator positions. The authors found that the range of F2 (expressing place of articulation differences) was smaller in CI users compared with healthy controls, whereas the range of F1 (expressing tongue height differences) did not show significant differences between the two participant groups. They argued that CI users could still rely on visible jaw height positions for F1 in maintaining category differences, whereas there was less of a visual cue accompanying F2 differences.

Our study adds to previous findings, illustrating the correlation between acoustic distance and ECD location in the auditory cortex. The high vowel [ɪ] differed from the other two vowels along the anterior–posterior and the superior–inferior axes. Note that the anterior–posterior distinctions follow the observation that more front vowels elicit more anterior dipoles. The vowel height differences, on the other hand, seem to be expressed by superior–anterior distinctions, although the directionality of our stimuli was not the same as found in Obleser, Elbert, et al. (2003). Although we cannot provide direct evidence that our dipole locations are better accounted for by abstract features, we still show that dipole distances are statistically better explained by feature differences than by acoustic distances. Furthermore, the fact that no single spatial dimension directly correlated with one of the first formant frequencies of our stimuli provides additional evidence for the view that the auditory cortex is particularly sensitive to more complex interactions of frequency components (e.g., Ohl & Scheich, 1997). Again, we would like to propose that these interactions are the acoustic bases for distinctive features. Localizing M100 sources obtained from the CI user promises to shed more light onto the interaction of spectral-acoustic and articulator-configuration information during vowel category formation in perception. Here, we would expect a stronger influence of articulator positions on the localizations of cortical vowel sources compared with normal-hearing controls. A model comparison as reported in this article could provide the necessary statistical means by which this difference may be established. Again, this is a promising study for future work, the results of which are potentially very useful for the design and improvement of CIs as well as for possible strategies to retain perceptually robust vowel categories.

Our second main finding relates to the MMN data that provide evidence for the underspecification of tongue height in the vowel [ɛ]. Crucially, we found the MMN in response to the mid vowel deviant [ɛ] after the standard vowel [æ] to be significantly larger than vice versa, that is, when the mid vowel [ɛ] was the standard and [æ] was the deviant. We assume that the low vowel [æ] in standard position activated its specified featural tongue height representation and set up a relatively high expectation for the subsequent mid vowel deviant to be specified as well. Its tongue height underspecification, however, violated this expectation and led to a larger MMN than one would assume on the basis of acoustic differences between standards and deviants alone. It is important to bear in mind that the MMN response will always be elicited in response to a detectable acoustic change. Our claim is that the size of the MMN furthermore depended on the featural oppositions between standard and deviant vowels. Importantly, the MMN for [ɛ] in [æ] context was larger than if [ɛ] was the standard and [æ] was the deviant. In this case, we assume that the mid vowel [ɛ] in standard position did not activate a tongue height specification and, consequently, created a low expectancy as to the tongue height of the deviant. Necessarily, the [low] specification of the deviant was a less severe violation. We would expect similar results for the deviant [ɪ]; however, because of duration limitations of MEG experiments with regard to participants' comfort levels, we could not include this condition in our current study. Clearly, the testing of the latter condition will be a necessary enterprise for future work.

The detected asymmetry in the mid/low vowel opposition was not found in the control condition that contrasted the low vowel [æ] with the high vowel [ɪ]. This important finding enables us to reject alternative explanations for the low/mid contrast asymmetry. Here, MMNs did not differ and were of similar magnitude for the deviant [æ] and the deviant [ɪ]. This pattern is expected on the basis of mismatching features. Both low and high vowels in standard position activate their fully specified representations for which the corresponding deviants provide a conflicting feature, that is, resulting in mutually exclusive feature oppositions of [low] versus [high], and [high] versus [low]. The fact that the MMN responses in the low/high conditions did not differ from MMN responses in the low/mid conditions furthermore suggests that the MMN amplitude is not solely driven by the acoustic distance between the standard and the deviant. This is in line with previous MMN research on vowels that suggests an interaction of auditory-sensory and phonological-categorical processing (Winkler et al., 1999), once again confirming that the size of the MMN does not linearly correlate with acoustic differences as measured as distance in formant space.

The finding that MMN latencies in the 150–200 ms window were shorter for the low/high conditions than for the low/mid conditions is consistent with earlier observations indicating that feature mismatches in passive oddball paradigms are not only reflected by higher MMN amplitudes but also by earlier MMN peak latencies (Eulitz & Lahiri, 2004; Scharinger, Eulitz, & Lahiri, 2010).

Our proposed correlation of standard–deviant feature opposition and magnitude of the resulting MMN is compatible with both the model adjustment hypothesis (Näätänen & Winkler, 1999; Winkler et al., 1996) and the predictive coding approach that incorporates the former (Baldeweg, 2006; Friston, 2005; Garrido et al., 2009). According to the model adjustment hypothesis, the MMN results from the need to update an acoustic model of the environment to incorporate (or assimilate) the respective deviant. This acoustic model, consisting of auditory memory, is instantiated by a sequence of standard sounds and generates inferences regarding future sound events, that is, a continuation of the standard sequence. Importantly, the model can have high or low inference values, which means that a deviant is less readily or more readily incorporated in the model (Winkler et al., 1996). Put differently, the model can be more or less confident in inferring future sound events and, consequently, will show larger mismatch responses if a highly confident inference is violated. We suggest that featurally specified vowels provide relatively high inference values, that is, high expectations that the next vowel in the standard sequence is also specified for the same feature. Encountering a deviant with a different feature extracted from the acoustic signal will then require a model update and therefore will elicit an MMN. In contrast, an underspecified vowel has a low inference value—that is, the expectation for a future sound to be specified is lower. Consequently, a deviant is more readily incorporated in the model and therefore will elicit a smaller MMN.

The predictive coding approach combines the *adaption mechanism* with the model adjustment hypothesis discussed above. According to the adaption mechanism, the MMN is an emergent phenomenon by subtracting adapted responses to standards from nonadapted responses to deviants (Jääskeläinen et al., 2004). Within the predictive coding approach, standard presentations help the suppression of prediction errors while integrating bottom-up sensory information with top-down predictions. Neurally, this is achieved by plastic changes in synaptic connections (adaptation). Upon encountering the deviant, the sensory bottom-up information fails to meet the top-down prediction. The consequence is the elicitation of an MMN (Näätänen & Winkler, 1999; Winkler et al., 1996). Note that the model adjustment hypothesis suggests two mechanisms for the elicitation of the MMN. One mechanism with generators at temporal locations

signals deviation from the learned regularity through the standards and involves a comparison of the sensory auditory input with the memory trace of the previous stimuli. The other mechanism with generators at frontal locations relates to an automatic (involuntary) attention switching process (Escera, Alho, Winkler, & Näätänen, 1998; Escera, Schröger, & Winkler, 2000; Giard, Perrin, Pernier, & Bouchet, 1990) that modulates deviance detection (Escera, Yago, Corral, Corbera, & Nunez, 2003).

Our results are also compatible with the predictive coding approach. In a similar way to that argued above within the model adjustment hypothesis, underspecified standards—by virtue of their less specified status—should evoke weaker bottom-up predictions. In particular, underspecified [ɛ] should not evoke specific predictions regarding the tongue height information supported from the signal. As a consequence, the deviant [æ] provides a relative weak predictive error and, consequentially, elicits a weak MMN. In contrast, specified [æ] in standard position makes strong and specific bottom-up predictions, such that the deviant [ɛ] supplies a stronger predictive error than in the reverse case. On the other hand, underspecified [ɛ] in standard position yields the only case in which there is no tongue height expectation, such that the low deviant [æ] provides novel information regarding a specific level of tongue height (low). The fact that we found a more frontal orientation of the M100 dipole for this deviant suggests that there might be more frontal processing already before the onset of the MMN. This would be in line with the involuntary attention switching process. Previous research has shown that the strength of the attention switching process depends on the novelty of the deviant (Escera et al., 1998; Rinne, Särkkä, Degerman, Schröger, & Alho, 2006). Thus, it could be the case that on the basis of the predictive model of the underspecified standard, the extraction of a discrete tongue height value from the deviant resulted in a stronger involuntary attention switch than in the other conditions in which the fully specified standards provided a predictive model in which the extraction of a discrete tongue height feature was more predicted. This attention switch, furthermore, might already be visible in a different M100 source configuration as reflected in a more frontal ECD orientation. Although these considerations have to remain speculative at this point, they can be taken as valuable hypotheses for future research.

Other MMN studies on vowel height differences suggest that early MMN effects are actually driven by acoustic stimulus properties. For example, Hill et al. (2004) reported that MMNs did not differ between 100 and 200 ms poststimulus onset between a vowel condition with [ɛ] and [ɪ] and a tone condition matched for the vowel's F1. The vowel and tone conditions diverged only between 200 and 350 ms after stimulus onset.

Although our crucial amplitude interaction in a later time window (200–250 ms) seems to support this view, care has to be taken with the interpretation of their results because they used synthetic vowels, whereas we employed spoken vowels with natural variation. Nevertheless, their results conform to our hypothesis according to which neural vowel processing is first subject to a specific acoustic analysis that focuses on formant (resonance) frequency regions and is later characterized by acoustic–phonetic feature integration and evaluation against long-term memory representations of discrete (vowel) sounds. Tone and vowel conditions in Hill et al.'s study differ at later latencies because these latencies probably reflect the comparison between acoustic–phonetic information and memory representations that differ between tones and vowels.

Hill et al.'s (2004) findings are challenged by the work of Jacobsen and colleagues (Jacobsen, Schröger, & Alter, 2004; Jacobsen, Schröger, & Sussman, 2004), who have provided evidence that language-relevant information is extracted even from complex nonspeech sounds. Jacobsen, Schröger, and Alter (2004) showed that complex tones with speech-like formant structure reliably elicited MMNs in the 100–200-ms time range after stimulus onset. In contrast, similarly complex tones without speech-like formant structure failed to elicit significant MMNs. Thus, the authors provided evidence that speech-relevant acoustic information is extracted preattentively from the incoming signal. Because the speech-relevant acoustic information, as expressed by formant structures, determines or constitutes distinctive features, we take these findings together with the results of our experiment as evidence for the brain's early and preattentive sensitivity to abstract phonetic features.

Again, our MMN findings may speak to the observation of reduced vowel spaces in CI users. Within underspecification theory, the higher confusability of vowels in CI users—as, for example, found by Löfqvist et al. (2010)—can be modeled by an increase in underspecified vowels. Possibly, specified vowels such as low [æ] can become underspecified if the spectral cue (F1) is consistently degraded or shows greater variation. Note that underspecified vowels are supposed to be more confusable because their long-term memory representation is compatible (i.e., not mismatching) with all the vowels that have no other mismatching feature dimensions. For example, [ɛ] may be confused with both [æ] and [ɪ] (underspecified vs. specified), but [æ] will be less confused with [ɪ] (both specified). CI users might counteract the emerging confusability of vowels by focusing on feature dimensions (e.g., F1) for which an alternative (visual) cues exists. Eventually, the visual cue can restore the full representation of a vowel's tongue height. Again, this is a speculative remark at the moment, but it

provides a possible starting point for fruitful future research.

Conclusion

Taken together, the MMN pattern of our experiment supports the idea of distinctive features that determine the representation of vowels and guide the perception of speech sounds. We found support for tongue height underspecification in American English mid vowels and showed that underspecification predictions for the MMN are in line with current models accounting for its underlying mechanisms. The M100 results, on the other hand, suggest that the auditory cortex categorically deals with acoustic cues to these features very early during spoken sound processing.

Acknowledgments

References

Ackermann, H., Hertrich, I., Mathiak, K., & Lutzenberger, W. (2001). Contralaterality of cortical auditory processing at the level of the M50/M100 complex and the mismatch field: A whole-head magnetoencephalography study. *NeuroReport, 12,* 1683–1687.

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 41–70). Berlin, Germany: Mouton de Gruyter.

Avery, P., & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology, 6,* 179–200.

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, England: Cambridge University Press.

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences, 10,* 93–94.

Bergelson, E., & Idsardi, W. J. (2009). A neurophysiological study into the foundations of tonal harmony. *NeuroReport, 20,* 239–244.

Boersma, P., & Weenink, D. (2009). PRAAT: Doing phonetics by computer (Version 5.1.0) [Computer software]. Amsterdam, the Netherlands: Institut for Phonetic Sciences.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201–251.

de Cheveigné, A. d., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience Methods, 165,* 297–305.

de Cheveigné, A. d., & Simon, J. Z. (2008). Sensor noise suppression. *Journal of Neuroscience Methods, 168,* 195–202.

Diesch, E., Eulitz, C., Hampson, S., & Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language, 53,* 143–168.

Diesch, E., & Luce, T. (1997a). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology, 34,* 501–510.

Diesch, E., & Luce, T. (1997b). Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research, 116,* 139–152.

Diesch, E., & Luce, T. (2000). Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex. *Journal of Cognitive Neuroscience, 12,* 878–893.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience, 10,* 590–604.

Escera, C., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology, 5,* 151–166.

Escera, C., Yago, E., Corral, M. J., Corbera, S., & Nunez, M. I. (2003). Attention capture by auditory significant stimuli: Semantic analysis follows attention switching. *European Journal of Neuroscience, 18,* 2408–2412.

Eulitz, C., Diesch, E., Pantev, C., & Hampson, S. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *Journal of Neuroscience, 15,* 2748–2755.

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

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

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology, 120,* 453–463.

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology, 27,* 627–640.

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., & Meyer, T. A. (2001). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *The Journal of the Acoustical Society of America, 109,* 2135–2145.

Hill, P. R., McArthur, G. M., & Bishop, D. V. M. (2004). Phonological categorization of vowels: A mismatch negativity study. *NeuroReport, 15,* 2195–2199.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97,* 3099–3111.

Hwang, S.-O. K., Monahan, P. J., & Idsardi, W. J. (2010). Underspecification and asymmetries in voicing perception. *Phonology, 27,* 205–224.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., . . . Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences, USA, 101,* 6809–6814.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41,* 654–659.

Jacobsen, T., Schröger, E., & Sussman, E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research, 20,* 473–479.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English.* Berlin, Germany: Mouton de Gruyter.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages.* Malden, MA: Blackwell.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 637–677). Berlin, Germany: Mouton de Gruyter.

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

Leahy, R. M., Mosher, J. C., Spencer, M. E., Huang, M. X., & Lewine, J. D. (1998). A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalography and Clinical Neurophysiology, 107,* 159–173.

Löfqvist, A., Sahlén, B., & Ibertsson, T. (2010). Vowel spaces in Swedish adolescents with cochlear implants. *The Journal of the Acoustical Society of America, 128,* 3064–3069.

Lopes da Silva, F. H., Wieringa, H. J., & Peters, M. J. (1991). Source localization of EEG versus MEG: Empirical comparison using visually evoked responses and theoretical considerations. *Brain Topography, 4,* 133–142.

Mäkelä, A. M., Alku, P., & Tiitinen, H. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neuroscience Letters, 353,* 111–114.

McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica, 43,* 84–108.

Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Process, 25,* 808–839.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38,* 1–21.

Näätänen, R., & Alho, K. (1997). Mismatch negativity (MMN): The measure for central sound representation accuracy. *Audiology and Neuro-Otology, 2,* 341–353.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., . . . Alho, K. (1997, January 30). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

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

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). Primitive intelligence in the auditory cortex. *Trends in Cognitive Sciences, 24,* 283–288.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus presentation in cognitive neuroscience. *Psychological Bulletin, 125,* 826–859.

Nakasatp, N., Levesque, M. F., Barth, D. S., Baumgartner, C., Rogers, R. L., & Sutherling, W. W. (1994). Comparisons of MEG, EEG, and ECoG source localization in neocortical partial epilepsy in humans. *Electroencephalography and Clinical Neurophysiology, 91,* 171–178.

Neumeyer, V., Harrington, J., & Draxler, C. (2010). An acoustic analysis of the vowel space in young and old cochlear-implant speakers. *Clinical Linguistics & Phonetics, 24,* 734–741.

Neurobehavioral Systems. (2012). Presentation [Computer software]. Retrieved from www.neurobs.com/.

Obleser, J., Elbert, T., & Eulitz, C. (2004). Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neuroscience, 5,* 1–24. doi:10.1186/1471-2202-5-24

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research, 15,* 207–213.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage, 20,* 1839–1847.

Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience, 16,* 31–39.

Ohl, F. W., & Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences, USA, 94,* 9440–9444.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9,* 97–113.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology, 94,* 26–40.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., & Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography & Clinical Neurophysiology, 69,* 160–170.

Pantev, C., Hoke, M., Lutkenhöner, B., & Lehnertz, K. (1989, October 27). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science, 246,* 486–488.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., . . . Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12,* 1038–1105.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Berlin, Germany: Springer-Verlag.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society. B: Biological Sciences, 363,* 1071–1086.

Poeppel, D., & Marantz, A. (2000). Cognitive neuroscience of speech processing. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 29–50). Cambridge, MA: MIT Press.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P. L., & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters, 221,* 145–148.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., . . . Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage, 14,* 607–616.

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology, 41,* 106–112.

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,* 135–143.

Roberts, T. P., Ferrari, P., & Poeppel, D. (1998). Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *NeuroReport, 9,* 3265–3269.

Roberts, T. P., Ferrari, P., Stufflebeam, S. M., & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology, 17,* 114–129.

Roberts, T. P., Flagg, E. J., & Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *NeuroReport, 15,* 1679–1682.

Roberts, T. P., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport, 7,* 1138–1140.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production.* Oxford, England: Oxford University Press.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport, 3,* 1149–1151.

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine & Biology, 32,* 11–22.

Scharinger, M., Eulitz, C., & Lahiri, A. (2010). Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics, 23,* 383–399.

Scherg, M., Vajsar, J., & Picton, T. W. (1990). A source analysis of the late human auditory evoked field. *Journal of Cognitive Neuroscience, 1,* 336–355.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research, 21,* 342–350.

Shtyrov, Y., & Pulvermüller, F. (2002a). Memory traces for inflectional affixes as shown by mismatch negativity. *European Journal of Neuroscience, 15,* 1085–1091.

Shtyrov, Y., & Pulvermüller, F. (2002b). Neurophysiological evidence of memory traces for words in the human brain. *Cognitive Neuroscience and Neuropsychology, 13,* 521–526.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–45.

Stevens, K. N. (1998). *Acoustic phonetics* (Vol. 30). Cambridge, MA: MIT Press.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America, 111,* 1872–1891.

Stevens, K., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America, 64,* 1358–1368.

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

Tiitinen, H., Mäkelä, A. M., Mäkinen, V., May, P. J. C., & Alku, P. (2005). Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory N1m response of the human brain. *BMC Neuroscience, 6,* 62. doi:10.1186/1471-2202-6-62

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology, 21,* 147–163.

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742,* 239–252.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., . . . Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357–369.

Asymmetries in the Processing of Vowel Height

Mathias Scharinger, Philip J. Monahan, and William J. Idsardi

This information is current as of June 11, 2012

This article, along with updated information and services, is
located on the World Wide Web at:
http://jslhr.asha.org/cgi/content/full/55/3/903

AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION