

A Comprehensive Three-dimensional Cortical Map of Vowel Space

Mathias Scharinger^{1,2}, William J. Idsardi¹,
and Samantha Poe¹

Abstract

■ Mammalian cortex is known to contain various kinds of spatial encoding schemes for sensory information including retinotopic, somatosensory, and tonotopic maps. Tonotopic maps are especially interesting for human speech sound processing because they encode linguistically salient acoustic properties. In this study, we mapped the entire vowel space of a language (Turkish) onto cortical locations by using the magnetic N1 (M100), an auditory-evoked component that peaks approximately 100 msec after auditory stimulus onset. We found that dipole locations could be structured into two distinct maps, one for vowels produced with the tongue positioned toward the front of the mouth (front

vowels) and one for vowels produced in the back of the mouth (back vowels). Furthermore, we found spatial gradients in lateral–medial, anterior–posterior, and inferior–superior dimensions that encoded the phonetic, categorical distinctions between all the vowels of Turkish. Statistical model comparisons of the dipole locations suggest that the spatial encoding scheme is not entirely based on acoustic bottom–up information but crucially involves featural–phonetic top–down modulation. Thus, multiple areas of excitation along the unidimensional basilar membrane are mapped into higher dimensional representations in auditory cortex. ■

INTRODUCTION

Research in cognitive neuroscience has uncovered several one- and two-dimensional maps in mammalian cortices that spatially parallel sensory dimensions in somatosensory (Tanriverdi, Al-Jehani, Poulin, & Olivier, 2009; Pons, Garraghty, Friedman, & Mishkin, 1987), visual (Wandell, Brewer, & Dougherty, 2005; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; Inouye, 1909), and auditory perception (Ohl & Scheich, 1997; Romani, Williamson, & Kaufman, 1982). In somatosensory cortex, distinct areas represent every part of the body. Furthermore, spatial relations between body part locations are reflected in spatial locations of cell assemblies, e.g., for tongue positions (Picard & Olivier, 1983). In visual cortex, centers of neuronal receptive fields encode the topographic structure of the visual field. They represent cortical maps of the visual field projected onto the retina and are accordingly called retinotopic maps (Wandell et al., 2005). In audition, the tonotopic structure of auditory cortex is well known and has been studied extensively in animals and humans (Romani et al., 1982). The frequency maps first appear at the level of the cochlea and are maintained through the ascending auditory pathways up to the cortex. Thus, tonotopy can be considered a basic principle of the information processing in the auditory system.

Spatial coding schemes in neuroscience such as those mentioned above are particularly interesting when applied to speech perception, because speech sounds relate acoustic and articulatory information (Stevens, 1998; Diehl, 1992), that is, involve a mapping of acoustics to body part positions (the tongue in relation to the oral cavity). Although linguistic theories differ as to whether speech sounds should be defined predominantly on the basis of their acoustic properties or predominantly on the basis of their underlying articulatory configuration (Stevens, 2002; Löfqvist, 1990), magnetic brain mapping studies on speech sounds (mostly vowels) showed that their cortical representation conforms to tonotopic maps in auditory cortex (Obleser, Lahiri, & Eulitz, 2003, 2004; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004; Obleser, Elbert, Lahiri, & Eulitz, 2003; Diesch & Luce, 1997; Diesch, Eulitz, Hampson, & Ross, 1996). However, previous studies have been limited to relatively small subset of vowel categories in any particular language, such that the tonotopic projection may only represent the relative spectral differences of the set of stimuli but not of the entire vowel system. Because somatosensory and visual maps represent the entire set of body parts and visual field locations, we assume that tonotopic maps can similarly encode entire sets of language-specific vowels. We propose that because of the multidimensional spectral properties of vowels involving several characteristic resonance frequencies (Stevens, 1998), the mapping in auditory cortex does not only use one but rather several spatial dimensions, mapping

¹University of Maryland, ²Max Planck Institute of Human Cognitive and Brain Sciences, Leipzig, Germany

multiple areas of excitation in the basilar membrane into higher dimensional maps in auditory cortex. Furthermore, given the articulatory aspect of speech sounds, we hypothesize that speech sound maps are not solely determined by bottom-up acoustic information but are simultaneously modulated by top-down information relying on abstract featural information relating to articulator positions. We tested our assumptions by examining the entire vowel space of Turkish, which provides us with an ideal system of eight vowels that differ along three articulatory dimensions forming a $2 \times 2 \times 2$ system. We illustrate this system in more detail in the Materials section.

Studies on the tonotopy of speech sounds in auditory cortex have focused on a dominant auditory evoked component, the N1, a negative vertex deflection in the electroencephalographic signal between 70 and 150 msec after stimulus onset and its magneto-encephalographic equivalent, the N1m/M100 (Näätänen & Picton, 1987). Its amplitude correlates with stimulus intensity (Elberling, Bak, Kofoed, Lebech, & Særmærk, 1981), whereas its latency reflects the (perceived) pitch (or other spectrally prominent frequencies) of acoustically presented stimuli (Roberts, Ferrari, & Poeppel, 1998; Roberts & Poeppel, 1996). Although the N1 is not an unitary single component (Näätänen & Picton, 1987), its underlying source is commonly considered to be well modeled by a single dipole in each hemisphere, the location of which differs according to acoustic stimulus spectral frequencies (Pantev et al., 1988, 1995; Tiitinen et al., 1993; Pantev, Hoke, Lutkenhoner, & Lehnertz, 1989; Romani et al., 1982). In particular, it has been shown that higher-frequency tones elicit dipoles with more medial sources in auditory cortex than lower frequency tones. Further research has provided evidence for a tonotopic gradient that is not restricted to the medial-lateral dimension (Langner, Sams, Heil, & Schulze, 1997). Together with other support for cognitive processing in the auditory cortex (Irvine, 2007), researchers also found a separate, specific, and orderly representation of speech sounds in this cortical area (Shestakova et al., 2004; Mäkelä, Alku, & Tiitinen, 2003; Diesch & Luce, 1997; Diesch et al., 1996; Eulitz, Diesch, Pantev, & Hampson, 1995; Huotilainen et al., 1995), contrasting with earlier findings that suggested identical processing of speech sounds and tones (Woods & Elmasian, 1986; Lawson & Gaillard, 1981). Several studies have shown that N1 dipole locations reflect a categorical distinction of speech sounds on the basis of their resonance (formant) frequencies (Obleser, Lahiri, et al., 2003, 2004; Shestakova et al., 2004; Mäkelä et al., 2003). Acoustic differences between vowels were paralleled in spatial dipole locations in a number of studies (Shestakova et al., 2004; Obleser, Elbert, et al., 2003). In particular, M100 source dipoles differed along the anterior-posterior axis according to differences in place of articulation (front or back in the mouth): Front vowels (e.g., [i]) reliably localized to more anterior positions than back vowels did (e.g., [u]; Obleser, Lahiri, et al.,

2003, 2004). Further research showed that this pattern also held for place of articulation differences in consonants but crucially depended on the intelligibility of the stimuli (Obleser, Scott, & Eulitz, 2006). Finally, the categorical anterior-posterior distinction of front and back speech sounds was not affected when attention was shifted away from the linguistic characteristics of the stimulus (Obleser, Elbert, & Eulitz, 2004). These findings suggest that M100 dipole sources emerge from spectral speech sound properties and are modulated by abstract, linguistically relevant categories with a potential reference to articulator positions.

On the basis of these previous findings, we employ M100 latency and source localization as neuromagnetic diagnostics of the cortical structuring of the Turkish vowel inventory. We assume that because of the complex acoustic informational structure of vowels, containing at least three characteristic resonance frequencies (F1–F3), cortical maps are not only based on one spatial dimension but use the available three-dimensional space to encode three separable dimensions of information. We further try to assess whether the expected vowel map is a pure bottom-up reflection of acoustic differences between vowels or whether it is additionally “warped” by linguistic categories (similar to the perceptual magnet effect; cf. Feldman, Griffiths, & Morgan, 2009; Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Guenther & Gajda, 1996; Kuhl, 1991). To that end, we employed statistical model comparison techniques that allowed us to determine whether the M100 latency and source patterns are better accounted for by acoustic gradient predictors or by phonetic discrete predictors.

METHODS

Materials

Turkish has a vowel inventory with eight different vowel phonemes. Relating to the articulator-based dimensions tongue height (vertical tongue position in the mouth, i.e., high/low), place of articulation (horizontal tongue position in the mouth, i.e., front/back), and lip rounding (rounded/unrounded), Turkish symmetrically distinguishes between high/nonhigh, back/front, and rounded/unrounded vowels (Kiliç & Ögüt, 2004). The articulator configurations have specific consequences for the acoustic vowel space of Turkish (Figure 1). The main acoustic cue for vowel height is the first resonance frequency (first formant, F1), whereas the main acoustic cue for place of articulation is the second resonance frequency (second formant, F2). Lip rounding, on the other hand, has an effect on all formant frequencies and is particularly evident on F2 and the third formant, F3 (Stevens, 1998; Fant, 1960).

We recorded 10 exemplars of each of the eight vowels from a native Turkish speaker who produced the stimuli in the Turkish equivalent of the carrier sentence, “I will say _ again.” He was instructed to pronounce the vowels

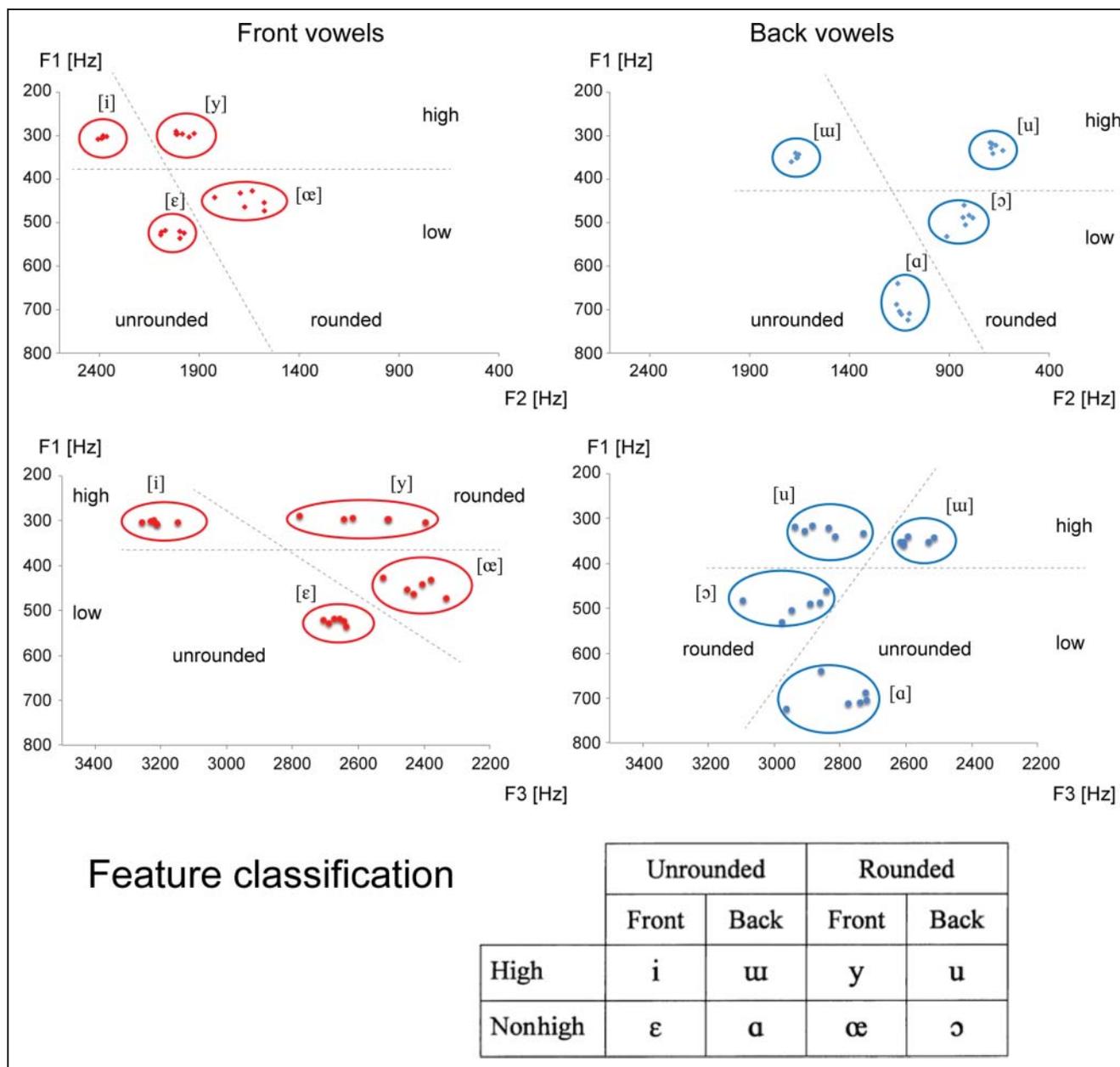


Figure 1. Acoustic vowel space of the vowel exemplars used in this study, separated into front (top and middle left) and back vowels (top and middle right). The top displays the vowel locations in the formant space of F1 and F2, and as the middle shows vowel locations in the F1 space. Corresponding phonetic feature classifications are summarized at the bottom and follow Kiliç and Ögüt (2004).

slowly and clearly, with an audible pause between the preceding and following context. All vowel exemplars were digitized at 44.1 kHz, with an amplitude resolution of 16 bit, using a Røde NT1-A microphone. Subsequently, six exemplars with similar pitch contours were selected, and steady-state portions of the vowels (200 msec) were excised from the original recordings using the phonetic speech analysis software PRAAT (Boersma & Weenink, 2009). Onsets and offsets were faded by 25 msec cosine-squared ramps. Average intensities were normalized to 70 dB (based on the mean power in Pa²/sec) to guarantee reliable auditory delivery at 60 dB SPL in the magnetoencephalogram scanner over the nonmagnetic Etymotic earphones.

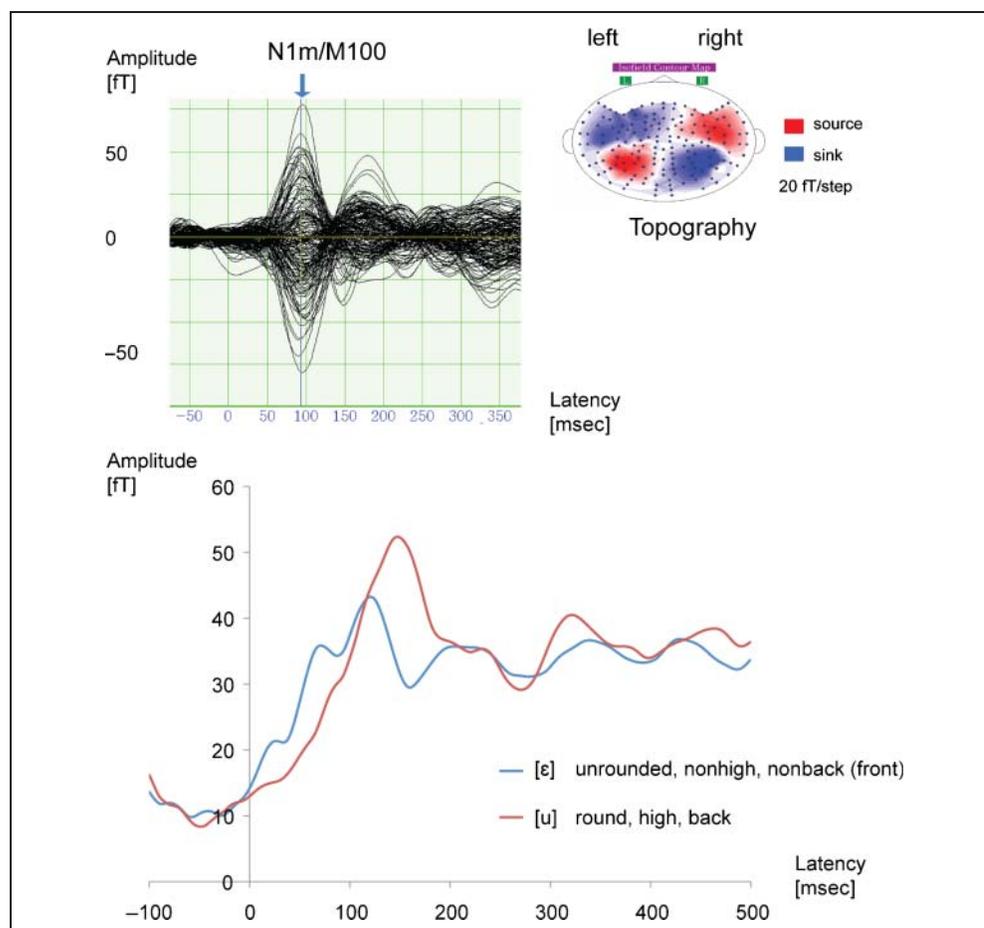
Experimental Design

All six vowel exemplars of each of the eight Turkish vowels were presented in random order with varying interstimulus intervals (400–900 msec). Frication noise with the same duration as the vowels was interspersed randomly ($p = .11$), and participants were asked to respond to the noise by pressing a button to ensure their attention during the experiment. Each vowel was presented at least 100 times (Figure 2). The occurrence of different vowels and frication noise was equiprobable. Stimuli were binaurally transmitted to the subject's ear by means of Etymotic ER3A insert earphones. Frequency bands were equalized with a

selection was necessary to balance between an over- and underfitting of the solution to the inverse problem (Sarvas, 1987). Dipoles were fitted on the rising slope of the M100 without taking solutions at or after the peak to avoid potentially changed source locations that may occur right after the M100 peak (Scherg, Vajsar, & Picton, 1990). Dipole modeling was done separately for each hemisphere (Sarvas, 1987).

Sensor and source data were analyzed with mixed effect models (Baayen, 2008; Pinheiro & Bates, 2000) that are preferable for data with empty cells and can include random effect terms as well as random slopes. In our models, we used Subject as random effect and included random slopes for Hemisphere. Multiple comparison used Tukey-corrected p values. We also performed model comparison analyses, as described in Baayen (2008) and Pinheiro and Bates (2000). The rationale of these model comparisons is to find the model that provides the best fit for the observed data without over- or underfitting. Model comparisons are made on the basis of the Akaike Information Criterion and the Bayes Information Criterion, essentially reflecting the entropy of the model. This entropy should be minimal, and models with a significantly lower Akaike Information Criterion or Bayes Information Criterion are evaluated using likelihood (L) ratios that are associated with a p value.

Figure 3. M100 responses in sensor space. A representative M100 waveform and topography is shown at the top. Latency and RMS amplitude differences between two vowels differing in all three feature dimensions, height, place, and rounding are illustrated at the bottom.



RESULTS

M100 latency, amplitude, and dipole source location data were fitted to acoustic and binary phonetic feature models. Acoustic models comprised the fixed effects $F1$, $F2$, and $F3$ (i.e., the first three formant values), whereas phonetic feature models used the binary fixed effects Height (high/nonhigh), Round (rounded/unrounded), and Place (back, front). The rationale of comparing acoustic and phonetic feature models is to assess whether acoustic or phonetic predictors yield the better fit and, in turn, whether an observed pattern is more likely to be explained by acoustic or by phonetic measures.

Latencies

The acoustic model showed a main effect of $F1$ ($F(1, 153) = 14.57, p < .001, \eta^2 = 0.12$) and Hemisphere ($F(1, 153) = 10.07, p < .01, \eta^2 = 0.01$). M100 latencies increased with decreasing $F1$ values, that is, M100 responses to high vowels (e.g., [i] with low $F1$) occurred later than M100 responses to low vowels (e.g., [a] with high $F1$). The model comparison between the acoustic and the binary feature model showed that the feature model had a better fit (L ratio = 21.07, $p < .001$). This indicates that simple binary classifications offer better model fits than the continuous acoustic

measures (cf. Roberts, Flagg, & Gage, 2004, for a similar finding). To ensure that we would not neglect potentially relevant acoustic characteristics aside from the first three formant frequencies, we calculated alternative acoustic mixed-effect models with different parameters. In one of these models, we included the vowels' fundamental frequency (F0), and in another one, we added F0, F1, F2, and F3 to a quasidiscrete acoustic measure. We also calculated a model with F1, F2, and combined F2 and F3. All alternative models were similarly worse than the discrete feature model (F0 model: L ratio = 106.7, $p < .001$; quasidiscrete model: L ratio = 122.7, $p < .001$; combined F2/F3 model: L ratio = 160.02, $p < .001$). It is not surprising that adding F0 to the acoustic model does not improve the model fit, because F0 usually covaries with F1, that is, does not add much information (Lehiste & Peterson, 1961; Peterson & Barney, 1952). Finally, because some of the formant frequencies strongly correlate with some of discrete features, it was not possible to consistently add discrete feature variables to acoustic models with F1, F2, and F3. For these reasons, we restricted our acoustic models to the first three formant frequencies and compare them to the feature models including three feature dimensions.

In the feature model for M100 latencies, we found a main effect for Height ($F(1, 153) = 7.50, p < .01, \eta^2 = 0.07$), Round ($F(1, 153) = 10.31, p < .01, \eta^2 = 0.1$), Place ($F(1, 153) = 3.78, p = .05, \eta^2 = 0.03$), and Hemisphere ($F(1, 153) = 11.10, p < .01, \eta^2 = 0.12$). Round vowels (e.g., [u]) peaked approximately 5 msec later than their unrounded counterparts (e.g., [ʊ]); high vowels (e.g., [i]) showed peaks on average 4 msec later than nonhigh vowels (e.g., [a]). Back vowels peaked 3 msec earlier than front vowels, and right hemispheric responses were 4.5 msec earlier than left hemispheric responses.

Amplitudes

RMS mean amplitudes were calculated over the 10 selected channels in each hemisphere. The acoustic model had main effects for all three formants (F1: $F(1, 180) = 22.38, p < .001, \eta^2 = 0.05$; F2: $F(1, 180) = 6.87, p < .01, \eta^2 = 0.02$; F3: $F(1, 180) = 16.93, p < .001, \eta^2 = 0.04$). M100 amplitudes increased with decreasing F1 and F2 values, that is, the largest M100 amplitudes were observed for high back vowels (e.g., [u]). In contrast, M100 amplitudes decreased with decreasing F3 values.

Again, the model comparison showed that the feature model fitted the data better than the acoustic model (L ratio = 19.78, $p < .001$), confirming that the extra precision of continuous acoustic values is not statistically warranted. The feature model revealed a main effect Height ($F(1, 180) = 16.88, p < .001, \eta^2 = 0.11$) and a marginal effect Round ($F(1, 180) = 3.49, p = .06, \eta^2 = 0.02$). High vowels (e.g., [u]) elicited larger M100 amplitudes (42 fT) than nonhigh vowels (e.g., [a]; 35 fT). The interaction of Place and Round ($F(1, 180) = 18.66, p < .001, \eta^2 = 0.12$) was based on higher amplitudes for rounded

back vowels (e.g., [u]) than for rounded front vowels (e.g., [y]; $t = 3.21, p < .05$), whereas unrounded vowels showed the opposite pattern ($t = 2.90, p < .05$).

Because source space amplitudes are based on more exact generator distance from the MEG sensors and on precise head position, we also calculated a mixed-effect model with the strength of the ECDs in each condition. Here, we found a significant interaction of Height, Place, Round, and Hemisphere ($F(1, 95) = 5.70, p < .05, \eta^2 = 0.13$), allowing us to look at the two hemispheres separately. Although there was no main effect in the right hemisphere, we found a main effect for Height ($F(1, 46) = 5.01, p < .05, \eta^2 = 0.25$) and a marginal effect of Round ($F(1, 46) = 3.05, p = .08, \eta^2 = 0.15$) in the left hemisphere. As in sensor space, high vowels elicited higher ECD amplitudes (difference, on average = 5 nAm).

In summary, M100 latencies and amplitudes were modulated by the spectral shape of the vowels in ways observed before (Roberts et al., 2004; Poeppel & Marantz, 2000; Roberts, Ferrari, Stufflebeam, & Poeppel, 2000; Roberts et al., 1998; Roberts & Poeppel, 1996). A particularly strong predictor of M100 latency and amplitude was the first formant frequency of the respective vowel and the related articulatory feature Tongue height. Furthermore, our model comparison suggests that feature-based predictors provide a better fit to the data than acoustic predictors; that is, the vowels of Turkish appear to be organized into binary acoustic oppositions rather than taking advantage of the full available acoustic frequency gradients.

Dipole Locations

Dipole models compared locations along the lateral–medial, anterior–posterior, and inferior–superior dimension. Again, we calculated acoustic and feature models, comprising the same effects as before.

The acoustic model for the dipole coordinates in the lateral–medial dimension showed a main effect of F2 ($F(1, 95) = 10.30, p < .01, \eta^2 = 0.13$). Vowels with higher F2 were located at more medial positions. This interacted with F1 ($F(1, 95) = 8.44, p < .01, \eta^2 = 0.02$). In the model comparison, the feature model provided a better fit of the coordinate data (L ratio = 25.03, $p < .001$). Here, we found a main effect of Round ($F(1, 95) = 7.54, p < .01, \eta^2 = 0.27$). Rounded vowels were located at more lateral positions than unrounded vowels (approximately 6-mm difference). Note that roundedness entails a spectrally lower center of gravity, conforming to the findings of Pantev et al. (1989), who showed that low-frequency tones elicited dipoles that were at more lateral locations than dipoles of high-frequency tones.

In the anterior–posterior dimension, we found a main effect of F2 ($F(1, 95) = 10.23, p < .01, \eta^2 = 0.01$) that paralleled a main effect of Place in the feature model ($F(1, 95) = 4.56, p < .05, \eta^2 = 0.02$). Crucially, front vowels (with a high F2) were located on average 3-mm anterior

to back vowels (with a low F2). This parallels previous findings by Obleser, Lahiri, et al. (2003, 2004). Again, the feature model provided a better fit of the location data (L ratio = 14.41, $p < .01$).

The acoustic model on the inferior–superior dimension of dipole locations did not show any significant effects and provided a worse fit than the feature model (L ratio = 17.51, $p < .001$). In the feature model, there was a significant interaction of Round \times Hemisphere ($F(1, 95) = 6.46$; $p < .05$, $\eta^2 = 0.08$). In the left but not in the right hemisphere, dipoles of round vowels were located 6.5 mm inferior to dipoles of nonround vowels (left: $t = 2.73$, $p < .05$; right: $t = 0.84$, $p = .40$).

Because there is interindividual variance in the structure of the temporal lobes (Shapleske, Rossell, Woodruff, & David, 1999), the averaged ECD locations may be blurred by this difference. For this reason, we calculated Euclidean distances between ECD locations separately for each subject, condition and hemisphere. We then calculated a mixed-effect model with the effect Distance type (28 distances between all vowels) and Hemisphere and found a main effect of Distance type ($F(27, 1140) = 5.78$, $p < .01$, $\eta^2 = 0.23$), although there was no hemispheric difference. Importantly, distances were larger between vowels that differed in more feature dimensions. For instance, the distance between [a] and [œ] (differing in rounding and place) was significantly larger than between [a] and [o] (differing in rounding only; $t = 5.08$, $p < .01$). Similarly, the distance between [u] and [ε] (differing in rounding, height, and place) was significantly larger than between [u] and [i] (differing in rounding and place; $t = 3.06$, $p < .05$). We looked in more detail at the number of differing features and found that Feature distance ($F(2, 1186) = 3.14$, $p < .05$, $\eta^2 = 0.04$), but not acoustic distance (Euclidean distance in the F1/F2/F3 space; $F(1, 1192) = 1.81$, $p = .18$, $\eta^2 = 0.01$), was a significant main effect in the expected direction in the corresponding model. Crucially, vowels that differed in less features elicited more collocated ECDs than vowels that

differed in more features. The feature model provided a better fit to the distance data than the acoustic model (L ratio = 42.31, $p < .001$). The acoustic model, on the other hand, yielded a significant interaction of Acoustic distance \times Hemisphere ($F(1, 1192) = 13.90$, $p < .01$, $\eta^2 = 0.04$). Acoustic distance was positively correlated with ECD distance in the right, but not in the left hemisphere.

Pooled coordinates across both hemisphere revealed that vowel dipole positions can be projected onto two cortical maps, each of which preserved the overall topology of the acoustic and articulatory vowel space (Figure 4). A map in the vertical plane contained the front vowels [i, y, ε, œ], and a separate, intersecting map in the horizontal plane contained the back vowels, [u, u, a, ɔ]. This map is positioned perpendicular to the front vowel plane. We, therefore, calculated separate analyses for front and back vowel dipole locations.

With regard to the lateral–medial dimension, front vowels showed F3 and F2 effects in the acoustic and a Round effect in the feature analysis (F2: $F(1, 43) = 4.10$, $p < .05$, $\eta^2 = 0.09$; F3: $F(1, 43) = 4.82$, $p < .05$, $\eta^2 = 0.11$; Round: $F(1, 43) = 5.66$, $p < .01$, $\eta^2 = 0.26$). Rounded vowels (with lower F2) were located at more lateral positions than unrounded vowels (with higher F2; difference = 5 mm, on average). In the inferior–superior dimension, vowels with high F1 were located superior to those with low F1 ($F(1, 43) = 4.28$, $p < .05$, $\eta^2 = 0.16$). This effect was paralleled by a Height effect in the feature analysis ($F(1, 43) = 4.53$, $p < .05$, $\eta^2 = 0.17$), reflecting that high vowel dipoles were located approximately 4.5 mm inferior to low vowel dipoles. For both dimensions, feature models provided better fits than acoustic models (lateral–medial: L ratio = 24.87, $p < .001$; inferior–superior: L ratio = 15.88, $p < .05$).

In back vowels, medial–lateral positions mapped F2 in the same way as in the front vowel plane ($F(1, 40) = 5.15$, $p < .05$, $\eta^2 = 0.08$), paralleling a Round effect in the feature model ($F(1, 40) = 4.63$, $p < .05$, $\eta^2 = 0.25$). Again, round vowel dipoles (with a lower F2) were located at more lateral positions than nonround vowel dipoles (with

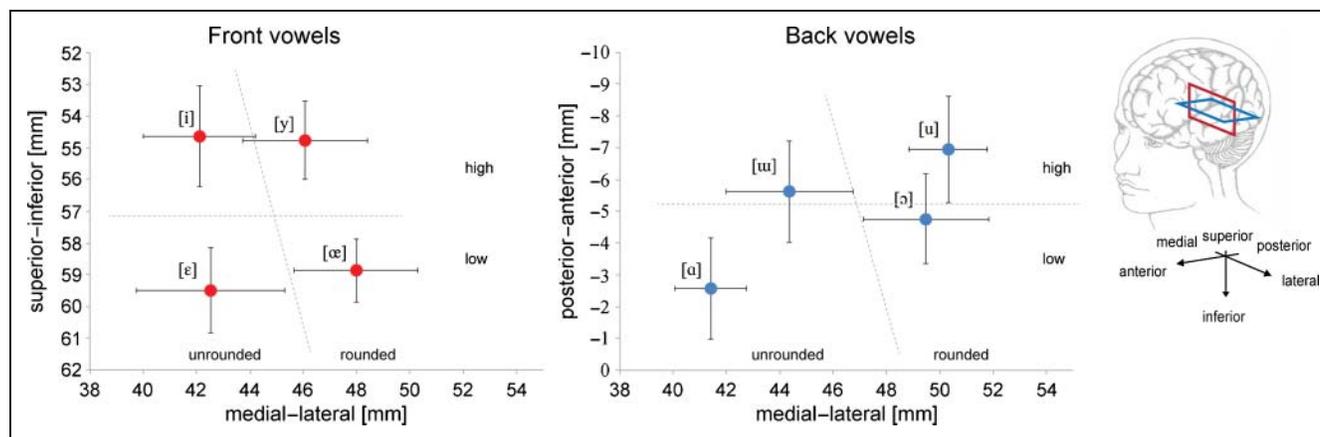


Figure 4. Dipole locations in spherical coordinates, pooled across both hemispheres. A distinct vertical front vowel map parallels acoustic and articulatory vowel distinctions in superior–inferior and medial–lateral dimensions. A horizontal back vowel map, orthogonal to the front vowel map, reflects acoustic and articulatory differences in posterior–anterior and medial–lateral dimensions. SEM (\pm) are indicated by whiskers.

higher F2, average difference = 6.5 mm). The analysis for anterior–posterior dipole locations revealed an interaction of F1 × Hemisphere ($F(1, 40) = 3.92, p = .05, \eta^2 = 0.03$). In the right hemisphere, F1 correlated with the anterior–posterior location ($F(1, 14) = 6.67, p < .05, \eta^2 = 0.07$). Although there was no Height effect in the feature model ($F(1, 40) = 2.20, ns, \eta^2 = 0.01$), feature models overall provided a better fit than acoustic models in both dimensions (lateral–medial: L ratio = 9.99, $p < .05$; inferior–superior: L ratio = 9.96, $p < .05$).

In summary, the dipole analysis provided us with the following spatial gradients: (1) an anterior–posterior gradient for place of articulation (front/back); (2) a medial–lateral gradient for roundedness (rounded/unrounded); (3) an inferior–superior gradient for tongue height in front vowels (high/low).

Front and back vowel dipoles populate distinct and orthogonal maps in the vertical and horizontal plane. Crucially, each spatial map faithfully represents acoustic and articulator-based (featural) distinctions.

DISCUSSION

Cortical maps reflecting spatial coding schemes have provided important insights into how sensory information is structured in the mammalian brain. Spatial dimension between body parts or locations in the visual field seem to be faithfully projected onto cortical surfaces (Tanriverdi et al., 2009; Wandell et al., 2005; Hadjikhani et al., 1998; Picard & Olivier, 1983; Inouye, 1909). Nonspatial differences between tones and speech sounds, on the other hand, have found to be similarly represented by spatial gradients in auditory cortex (Shestakova et al., 2004; Obleser, Elbert, et al., 2003; Diesch & Luce, 1997; Ohl & Scheich, 1997; Huotilainen et al., 1995; Pantev et al., 1988, 1989, 1995; Tiitinen et al., 1993; Romani et al., 1982). In this study, we provide further evidence for these gradients and for the first time report the cortical mapping of an entire vowel space by calculating source localizations to the auditory evoked M100. Our main findings support the assumption that multidimensional acoustic relations are mapped onto three-dimensional cortical space described by lateral–medial, anterior–posterior, and inferior–superior axes. Furthermore, our statistical model comparisons showed that, whereas cortical vowel maps reflect acoustic properties of the speech signal, articulator-based and featural speech sound information additionally warps the acoustic space toward linguistically relevant categories (Obleser, Lahiri, et al., 2004; Jacquemot et al., 2003). We leave it open for further research to address the question of whether this warping provides direct evidence for a neural coding of articulator configurations or articulator movements (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Liberman & Mattingly, 1985) or whether acoustically defined categories suffice to explain the observed warping. On the basis of our model comparisons that yield better fits for models with discrete, categorical feature variables, we suggest that con-

tinuous, precategorical input representations as indexed by the M100 are already shaped according to native sound category representations. We are aware that, for a stronger claim toward dissociating continuous and categorical processing and representation, we would need stimuli with greater acoustic variation and less correlation between acoustic parameters (e.g., F1) and feature dimensions (e.g., tongue height). When considering ECD distances, however, the superiority of the feature models is substantiated. Here, we found that the collocations of ECDs strongly correlated with the number of differing feature dimensions but not with the Euclidean distances in the F1/F2/F3 acoustic space. In particular, vowels that differed in one feature dimension elicited more collocated ECDs than vowels that differed in more than one feature dimension. Finally, some of our model comparisons involving sensor- and source-space data suggest that feature-based variables are better predictors for the left, and acoustics-based variables are better predictors for the right hemisphere (cf. Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; Alho et al., 1998) consistent with well-known claims that language processing is more strongly represented in the left hemisphere. Although ECD modeling makes the strong assumption that N1 sources can be described as a single point source, therefore, estimating centers of activity over widespread auditory areas, the importance of our findings is that these centers of activity follow well-established tonotopic gradients within the temporal cortices. Furthermore, the relative ECD positions correlating with phonetic feature distinctions have been replicated several times, using different speech sounds and slightly different experimental designs (Obleser et al., 2006; Obleser, Lahiri, et al., 2003, 2004; Obleser, Elbert, et al., 2003).

In what follows, we will discuss the significance of each of the three gradients with respect to previous N1 studies.

Lateral–Medial Gradient

The lateral–medial gradient is the original dimension along which the tonotopic organization of auditory cortex was discovered (Pantev et al., 1988, 1989, 1995; Romani et al., 1982). Pure tones with high frequencies were found to elicit dipoles at more medial locations (i.e., deeper in the brain) compared with tones with low frequencies that elicited dipoles at more lateral locations (i.e., less deep in the brain). Although initially it appeared that the lateral–medial gradient reflected pitch differences independently of the acoustic stimulus complexity (Pantev et al., 1989), subsequent studies showed that pitch (fundamental frequency) and timbre (higher frequency components) followed different, orthogonal gradients (Langner et al., 1997). With regard to speech sounds, the lateral–medial gradient does not seem to map pitch (Poeppl et al., 1997) but rather higher-frequency components, such as the first formant (F1). Diesch and Luce (1997) found that stimuli with a high-frequency formant elicited N1 dipoles with a more medial location than stimuli with a low-frequency

formant. This follows from the original observation that higher frequencies correspond to deeper N1 sources. In our experiment, we replicated this general pattern. Overall, we observed that vowels with higher F2 had dipoles at more medial locations. An even better predictor for this pattern was Round, a phonetic feature that indicates whether vowels are articulated with rounded or unrounded lips. Lip rounding lowers all formant frequencies (Stevens, 1998), such that unrounded vowels usually have higher formant frequencies than rounded vowels. This acoustic–phonetic phenomenon was reflected in the lateral–medial gradient: Unrounded vowels with higher formant frequencies elicited dipoles medial to those elicited by rounded vowels.

Anterior–Posterior Gradient

Dipole locations were observed to differ along the anterior–posterior gradient for tones (Langner et al., 1997), but particularly for synthesized vowels (Diesch & Luce, 2000), natural vowels, and consonants (Obleser et al., 2006; Obleser, Lahiri, et al., 2003, 2004; Obleser, Elbert, et al., 2003). The initial finding was that higher formant frequencies resulted in more anterior M100 dipoles. Obleser and colleagues showed that the anterior–posterior gradient is particularly robust for the phonetic features front versus back, encoding place of articulation (referring to the horizontal tongue position in the mouth). The main acoustic cue for this feature is the second formant frequency (F2). Obleser, Lahiri, et al. (2004) showed that front vowels (e.g., [i]) elicited M100 dipoles anterior to the back vowel dipoles (e.g., [u]). A comparison of this pattern to the dipole location of [a] with a rather central place of articulation on the basis of F2 suggested that the anterior–posterior gradient was based on the feature Place of articulation and less so on the second formant. Further support for this view was obtained from a study on consonants, where place of articulation differences are less transparent with respect to F2. Obleser et al. (2006) provided evidence that the anterior–posterior gradient differentiated between front and back consonants (e.g., [t] vs. [k]) to the same extent it differentiated between front and back vowels (e.g., [i] vs. [u]). In our study, we found the same pattern: Front vowels elicited more anterior dipoles than back vowels, and the feature-based model provided a better fit to this pattern than the acoustic model with formant values. For our separation into front and back vowel maps, this gradient means that the center of gravity of the front vowel map is located anterior to the center of gravity of the back vowel map (Figure 4).

Inferior–Posterior Gradient

An inferior–posterior gradient has been reported previously; however, only in relation to an anterior–posterior gradient of tonotopic organization (Langner et al., 1997). To our knowledge, no one has suggested that it might represent a single phonetic feature gradient. In our study, we

found a left hemispheric inferior–superior gradient for the phonetic feature Round: Round vowels with lower formant frequencies corresponded to dipoles with more inferior locations, whereas unrounded vowels with higher formant frequencies corresponded to dipoles with more superior locations. In the front vowel map, low vowel dipoles were superior to high vowel dipoles. Again, this seems to correlate with higher formant (F1) frequencies for superior dipole positions and lower formant (F1) frequencies for inferior dipole positions.

Although the phonetic dimension of tongue height did not correspond to a single cortical spatial dimension across all vowels, it did correspond to M100 latencies. To this end, our data conform to earlier findings that showed earlier N1/M100 peak latencies for low vowels whose first formant frequency approaches 1 kHz (Roberts et al., 2004; Diesch & Luce, 2000; Poeppel & Marantz, 2000; Roberts et al., 2000; Roberts et al., 1998; Poeppel et al., 1997; Roberts & Poeppel, 1996; Eulitz et al., 1995). However, similar to the M100 dipole location pattern in our study (and consistent with Roberts et al., 2004), the M100 latencies were better explained by discrete featural variables than by acoustic predictors.

Conclusion

In mapping the entire vowel space of a language (Turkish) onto cortical locations, we could show that auditory spatial coding schemes use particular spectral information and warp the spatial arrangement of neuronal sources according to phonetic categories. This is an important extension of the previously discovered tonotopic organization of auditory cortex to the extent that human speech uses auditory cortical architecture shared with other mammals in a way that is beneficial to the encoding of categorical information from different sources (auditory and articulatory). Although we can only speculate whether phonetic features emerge from these articulatory-warped acoustic maps in unifying articulator and acoustic information into a more abstract featural representation, it is a promising way to pursue research on the cortical bases of speech perception. Because languages differ widely in their inventories for vowels (Maddieson, 1984; Liljencrants & Lindblom, 1972) with some languages having as few as two vowels and others having as many as two dozen, the mapping of vowel systems of additional languages promises to shed new light on which acoustic properties are the most important for the cortical processes underlying speech.

Acknowledgments

We want to express our special thanks to the following persons who helped us improving this article and who inspired us with invaluable discussions: Ellen F. Lau, David Poeppel, Baris Kabak, and the members of the UMD PFNA group. We also thank Max Ehrmann for laboratory support. The research for this study

was funded by the NIH grant 7ROIDC005660-07 to W. I. & David Poeppel.

Reprint requests should be sent to Dr. Mathias Scharinger, Max Planck Institute for Human Cognitive & Brain Sciences Stephanstr. 1A-04103 Leipzig, Germany, or via e-mail: mscharinger@cbs.mpg.de.

REFERENCES

- Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huottilainen, M., Virtanen, J., et al. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience Letters*, *258*, 9–12.
- Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Boersma, P., & Weenink, D. (2009). *PRAAT: Doing phonetics by computer (ver. 5.1.0)*. Amsterdam: Institut for Phonetic Sciences.
- 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.
- Diehl, R. L. (1992). *Articulatory phonology* (Vol. 49). Basel: Karger.
- 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. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology*, *34*, 501–510.
- 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.
- Elberling, C., Bak, C., Kofoed, B., Lebech, J., & Særmark, K. (1981). Auditory magnetic fields from the human cortex. Influence of stimulus intensity. *Scandinavian Audiology*, *10*, 203–207.
- 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.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*, 2608–2611.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton de Gruyter.
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*, 752–782.
- Guenther, F. H., & Gajda, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*, 1111–1121.
- Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V 8. *Nature Neuroscience*, *1*, 235–241.
- Huottilainen, M., Tiitinen, H., Lavikainen, J., Ilmoniemi, R. J., Pekkonen, E., Sinkkonen, J., et al. (1995). Sustained fields of tones and glides reflect tonotopy of the auditory cortex. *NeuroReport*, *6*, 841–844.
- Inouye, T. (1909). *Die Sehstörungen bei Schussverletzungen der kortikalen Sehphäre nach Beobachtungen an Verwundeten der letzten japanischen Kriege* [Visual impairments resulting from bullet wounds in visual cortex according to observations on injured of the last Japanese wars]. Leipzig: W. Engelmann.
- Irvine, D. R. F. (2007). Auditory cortical plasticity: Does it provide evidence for cognitive processing in the auditory cortex? *Hearing Research*, *229*, 158–170.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *23*, 9541–9546.
- Kiliç, M. A., & Ögüt, F. (2004). A high unrounded vowel in Turkish: Is it a central or back vowel? *Speech Communication*, *43*, 143–154.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*, 93–107.
- Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *181*, 665–676.
- Lawson, E. A., & Gaillard, A. W. K. (1981). Evoked potentials to consonant-vowel syllables. *Acta Psychologica*, *49*, 17–25.
- Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*, 419–425.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862.
- Löfqvist, A. (1990). Speech as audible gestures. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 289–322). Dordrecht: Kluwer Academic Publishers.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.
- 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.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.
- Obleser, J., Elbert, T., & Eulitz, C. (2004). Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neuroscience*, *5*, 1–24.
- 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.
- Obleser, J., Scott, S. K., & Eulitz, C. (2006). Now you hear it, now you don't: Transient traces of consonants and their unintelligible analogues in the human brain. *Cerebral Cortex*, *16*, 1069–1076.

- Ohl, F. W., & Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences, U.S.A.*, *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., et al. (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., Lutkenhoner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science*, *246*, 486–488.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.
- Picard, C., & Olivier, A. (1983). Sensory cortical tongue representation in man. *Journal of Neurosurgery*, *59*, 781–789.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.
- Poeppl, D., & Marantz, A. (2000). Cognitive neuroscience of speech processing. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 29–49). Cambridge, MA: MIT Press.
- Poeppl, 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.
- Pons, T. P., Garraghty, P. E., Friedman, D. P., & Mishkin, M. (1987). Physiological evidence for serial processing in somatosensory cortex. *Science*, *237*, 417–420.
- 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., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport*, *7*, 1138–1140.
- Roberts, T. P. L., 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. L., Flagg, E. J., & Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *NeuroReport*, *15*, 1679–1682.
- Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*, 1339–1340.
- Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine & Biology*, *32*, 11–22.
- 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.
- Shapleske, J., Rossell, S. L., Woodruff, P. W., & David, A. S. (1999). The planum temporale: A systematic, quantitative review of its structural, functional and clinical significance. *Brain Research, Brain Research Reviews*, *29*, 26–49.
- Shetakova, 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., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *Neuroimage*, *12*, 657–663.
- Stevens, K. N. (1998). *Acoustic phonetics* (Vol. 30). Cambridge, MA: The 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.
- Tanriverdi, T., Al-Jehani, H., Poulin, N., & Olivier, A. (2009). Functional results of electrical cortical stimulation of the lower sensory strip. *Journal of Clinical Neuroscience*, *16*, 1188–1194.
- Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R. J., Simola, J., & Näätänen, R. (1993). Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. *Psychophysiology*, *30*, 537–540.
- Wandell, B. A., Brewer, A. A., & Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philosophical Transactions B*, *360*, 693–707.
- Woods, D. L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology*, *65*, 447–459.