

## Research report

# Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies

Jonas Obleser<sup>a,\*</sup>, Thomas Elbert<sup>a</sup>, Aditi Lahiri<sup>b</sup>, Carsten Eulitz<sup>a,b</sup><sup>a</sup>University of Konstanz, Department of Clinical Psychology, P.O. Box D25, 78457 Konstanz, Germany<sup>b</sup>University of Konstanz, Department of Linguistics, Konstanz, Germany

Accepted 20 May 2002

---

**Abstract**

We studied neuromagnetic correlates of the processing of German vowels [a], [e] and [i]. The aim was (i) to show an influence of acoustic/phonetic features on timing and mapping of the N100m component and (ii) to demonstrate the retest reliability of these parameters. To assess the spatial configuration of the N100m generators, Euclidean distances between vowel sources were computed. Latency, amplitude, and source locations of the N100m component differed between vowels. The acoustically most dissimilar vowels [a] and [i] showed more distant source locations than the more similar vowels [e] and [i]. This pattern of results was reliably found in a second experimental session after at least 5 days. The results suggest the preservation of spectral dissimilarities as mapped in a  $F_1$ – $F_2$  vowel space in a cortical representation.

© 2002 Elsevier Science B.V. All rights reserved.

*Theme:* Sensory systems*Topic:* Auditory systems: central physiology*Keywords:* Auditory cortex; Auditory evoked potential; Central auditory system; Auditory evoked field

---

**1. Introduction**

As infant brains are exposed to the speech of their native language they develop a blueprint of its sound inventory. At the same time they gradually lose their ability to discriminate foreign speech sounds. This has been shown consistently [2,13,14,17,25]. But surprisingly little is known about how and where this representation of speech sounds is implemented in the human brain [6,30]. The present study further addresses and tests the idea of an orderly phonemotopic representation of vowels in human auditory cortex.

The functional organization of the auditory cortex has been studied extensively in animals [10,24,28,29] as well as in humans [16,20,33] for which tonotopic maps comprise the most robust finding. In the representation of speech sounds, i.e. vowels, consonants and CV-syllables,

tonotopy as well as various other mapping principles have been proposed and tested: in an awake monkey, Steinschneider et al. [29] found the maximal cortical activity in response to three different CV-syllables to be closely linked to the absolute spectral peak of these syllables and interpreted this in favor of a tonotopic phoneme representation that mirrors the maximal spectral peak. In contrast, Schreiner [28], investigating cats, reports distinct cortical patches for the formants  $F_1$ – $F_3$  that corresponded to locations of the formant frequencies along a tonotopic gradient. Using magnetoencephalography to study vowel representation in human subjects, Poeppel et al. [22], Eulitz et al. [8] and Diesch et al. [3] did not find main effects of vowel type along the medial–lateral axis (which reflects tonotopic organization in humans, cf. [16,20]). However, this does not rule out a formant extraction principle as proposed by Schreiner [28]: if the perceptually relevant [9,18,21] formant frequencies  $F_1$  and  $F_2$  are extracted and mapped onto a tonotopically structured cortical patch, this does not necessarily lead to extracranially detectable differences in generator location.

---

\*Corresponding author. Tel.: +49-7531-88-4605; fax: +49-7531-88-4601.

E-mail address: [jonas.obleser@uni-konstanz.de](mailto:jonas.obleser@uni-konstanz.de) (J. Obleser).

Furthermore, a variety of tonotopically organized maps with differentially ordered isofrequency areas have been shown in the auditory cortex [10,24]. This rather rules out differences between speech sound representations only along one spatial dimension according to their spectral properties.

Diesch and Luce [4] also demonstrated that tonotopic organization is insufficient to explain phoneme representation in the human cortex: when applying magnetic source imaging to the N100m component in response to two-formant vowels and to their decomposed formants alone, they did not find the vowel source location to be a linear superposition of formant source locations. Here, and in a following study [5], the authors suggested reciprocal formant inhibition as an important mapping principle: vowel formants  $F_1$  and  $F_2$  may interact at one or more early stages of auditory processing. As neurons along the auditory neural pathway respond best to narrow bands of preferred frequencies and feed narrow-band information to higher processing stages, cortical mapping of complex stimuli might mirror interactions of the extracted prominent peaks in the frequency spectrum. Ohl and Scheich [18] collected further evidence for such a mapping of formant inhibition when they stimulated gerbils with four vowels differing in the  $F_1$ – $F_2$  distance. The size of the activated cortical patch was largest for the vowel [i], which was characterized by a  $F_1$ – $F_2$  distance of more than 2000 Hz and smallest for [o] with a respective distance of less than 500 Hz. Furthermore, the authors found this activation extended along isofrequency stripes and orthogonal to a tonotopic gradient – again consistent with Diesch and Luce's suggestion.

Previous research on the functional organization of speech sounds in the human auditory cortex has suffered from several methodological problems: invasive animal research working with human speech sounds lacks external validity and cannot be related to perceived qualities of the stimulus material, whereas noninvasive research in human subjects suffers from lack of spatial precision and failure to deal with an obviously high interindividual variability. Diesch et al. [3] addressed the latter problem by calculating Euclidean distances between vowel sources on an individual basis (see also [7] for an application of this method in the somatosensory modality). As a result, the distance between sources representing [u] and [i] was larger than the distance between [a] and [ε] in more than 60% of all experimental conditions (i.e. tested hemispheres, components and stimulus lengths). This can be considered indicative for a cortical map related to spectral dissimilarities of speech input. But, still, the main finding of this study was a high variability across subjects.

If extraction of spectral information, tonotopically organized cortical patches with different preferred stimulus characteristics and reciprocal inhibition of spectral peaks coincide along the neural pathway, an orderly representation of phonemes is most likely. It is well-known from

other sensory modalities as well as from lower levels of the auditory modality that such maps exist, that they differ in their degree of abstractness, and that they are shaped by experience [1]. Kohonen and Hari [12] were able to demonstrate that self-organizing systems that model cortical functioning, when trained with Finnish speech, build up a phonemotopic representation that reflects spectral dissimilarities of the input signals. It is then not surprising that an animal study that used species-specific vocalizations [32] reported spatially distributed neuronal discharge patterns in marmosets' primary auditory cortex in response to behaviorally relevant vocalizations.

By and large, evidence suggests that phonemotopic organization in the perisylvian regions of the brain is a hypothetical concept that is worth further examination. The present experiment was set up to test this hypothesis, by locating the cortical representational centers of three different unrounded German vowels, selected to have largest and smallest differences in their spectral and phonological characteristics. If the phonemotopic concept holds, larger spectral and phonological dissimilarities should be reflected by larger Euclidean distances between neural generators activated by the vowels. In this study the N100m component of the brain's neuromagnetic response to acoustically presented vowels was used. Every subject participated in the same experimental session twice, in order to assess test/retest reliability and thus to detect intraindividual stability of features in the functional organization of vowel representation.

## 2. Materials and methods

### 2.1. Subjects

Twelve subjects (six females) participated in the experiment, with a mean age of  $25.6 \pm 2.3$  years ( $M \pm S.D.$ ). None reported a history of neurological, psychiatric, or otological illness. Only right-handers were included, as ascertained by the Edinburgh Handedness Questionnaire [19]. Individual hearing thresholds for both ears were determined individually and for each stimulus. Subjects gave written informed consent and were paid €25 for participation.

### 2.2. Stimulation

The cortical representations of the three German vowels [a], [e] and [i] were explored. [a] and [i] span a great distance in the  $F_1$ – $F_2$ -vowel space, [e] has an intermediate position (Fig. 3b). It was expected that this relationship in vowel space is reflected in the location of the respective cortical representational foci. A fourth vowel, [ə] was selected to assure that subjects attended constantly to the stimuli: in a target detection task, subjects listened to pseudorandom sequences of the four synthetic realizations

[11] of the German vowels [a] (as in ‘father’), [e] (similar to ‘bait’ or ‘bay’ but not diphthongized), [i] (as in ‘beat’), and a long schwa-like vowel [ə]. When the three nontarget vowels were presented every subject easily identified them as the corresponding German vowel.

Stimulus duration was set at 600 ms and all stimuli had a fundamental frequency  $F_0$  of 129 Hz, falling linearly to 119 Hz. Stimuli differed in formant frequencies  $F_1$  to  $F_3$ , as depicted in Table 1.  $F_4$  (3900 Hz),  $F_5$  (4700 Hz) and  $F_6$  (5100 Hz) were held constant across stimuli, as were on- and offset characteristics (50 ms gaussian onset ramp, 150 ms gaussian offset ramp).

The vowel sequences were presented binaurally at 50 dB SPL above respective hearing threshold via a nonmagnetic and echo-free stimulus delivery system (with almost linear frequency characteristics in the critical range of 200–4000 Hz). Each of the three sequences consisted of 520 stimuli presented with a randomized stimulus onset asynchrony of  $2.0 \pm 0.2$  s and a target probability of 7%. Subjects were instructed to press a button with their right index finger when detecting the target vowel. Subjects watched silent videos in order to maintain constant arousal and to reduce excessive eye movements.

In every subject the experimental session was repeated after 1 week with the sensor-to-head position and the resting position held constant across sessions. Measurements were obtained from a supine position in order to minimize movement artifacts. Recordings were made in a sound-attenuated and magnetically shielded room (Vacumschmelze).

### 2.3. Data acquisition and reduction

Auditory magnetic fields (AEFs) evoked by the three different vowel stimuli were recorded simultaneously from both hemispheres using a whole head neuromagnetometer (Magnes 2500, 4D Neuroimaging). Epochs of 1200-ms duration (including a 200-ms pretrigger baseline) were recorded with a bandwidth from 0.1 to 200 Hz and a 678.17-Hz sampling rate. If the peak-to-peak amplitude exceeded 3.5 pT or the co-registered EOG-signal was larger than 100  $\mu$ V in one of the channels, epochs were rejected. Epochs that contained button presses were also excluded. Between 250 and 480 artifact-free epochs that remained for every subject and nontarget vowel were averaged after off-line noise correction. A 20-Hz lowpass

filter (Butterworth 12 dB/oct, zero phase shift) was subsequently applied to the average.

Further analysis was confined to the rising slope and peak of the N100m component defined as the prominent waveform deflection in the time range between 90 and 150 ms. Isofield contour plots of the magnetic field distribution were inspected visually.

N100m peak amplitude was calculated as the maximum root mean square (RMS) over 34 magnetometer channels selected to include the field extrema over the left and the right hemisphere, respectively. Peak latency was defined as the sampling point by which the RMS reached its maximum. Using the same set of channels, an equivalent current dipole (ECD) in a spherical volume conductor (fitted to the shape of the regional head surface) was modeled at every sampling point separately for the left and the right hemisphere.

The N100m source parameters were determined as the median of  $16 \pm 5$  ECD solutions in the latency range of 30 ms before the RMS peak. To be included in this calculation, single ECD solutions had to meet the following criteria: (1) goodness-of-fit greater than 0.90, (2) ECD location larger than 1.5 cm in medial–lateral direction from the center of the brain and 3–8 cm in superior direction, measured from the connecting line of the preauricular points.

N100m RMS peak latency and amplitude, ECD source strength and location of the N100m were submitted to a 3-factorial analysis of variance (ANOVA) with the repeated measurement factors vowel type ([a], [e], [i]), hemisphere (left, right) and session (first, second).

To elucidate relative source configurations, we calculated Euclidean distances between the three vowel ECD sources separately for each hemisphere, each experimental session and each subject. The distance [a]–[i] was statistically compared to the distance [e]–[i] in a 3-factorial ANOVA with the repeated measures factors hemisphere (left, right), session (first, second) and distance ([a]–[i], [e]–[i]). Significant effects were further explored by means of contrast analyses. Where appropriate,  $P$  values were Greenhouse–Geisser corrected.

## 3. Results

### 3.1. Peak latency, amplitude and source strength

Peak latency, amplitude and source location of the N100m component differed systematically between the three different vowels, supporting the notion of vowel-specific cortical responses (Figs. 1 and 2a).

Statistical analysis of the peak latency of the N100m component revealed a main effect of vowel type [ $F(2,20)=13.91$ ,  $\epsilon=0.81$ ,  $P<0.001$ ]: The vowel [a] elicited the fastest N100m response (121.4 ms). As certified

Table 1  
Frequencies of the first three formants of all vowels are shown. Details of fixed frequencies  $F_0$ ,  $F_4$ – $F_6$  are given in the text

|     | $F_1$ (Hz) | $F_2$ (Hz) | $F_3$ (Hz) |
|-----|------------|------------|------------|
| [a] | 780        | 1250       | 2600       |
| [e] | 370        | 2250       | 2800       |
| [i] | 250        | 2700       | 3400       |
| [ə] | 350        | 1400       | 2500       |

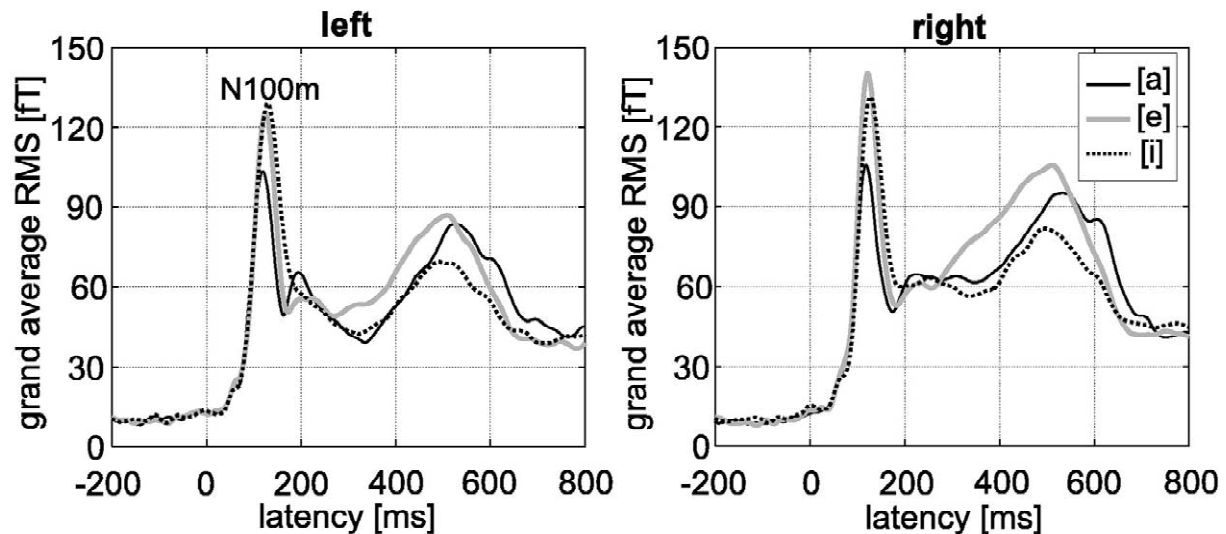


Fig. 1. Grand average amplitudes (RMS) for the vowels [a], [e] and [i] are shown for the left and right hemisphere, respectively. Analyses reported here were confined to rising slope and peak of the N100m deflection.

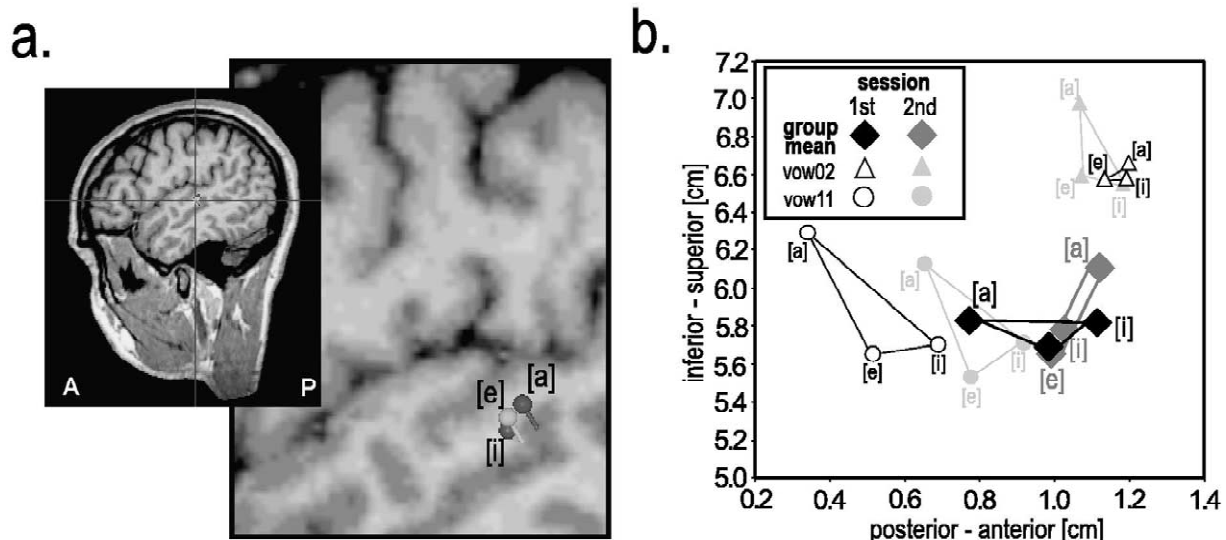


Fig. 2. (a) Source locations for the vowels [a], [e] and [i] (red, green and blue dots, respectively) projected onto a standardized brain are shown for the left auditory cortex of a single subject. (b) Vowel source maps in the sagittal plane of the left hemisphere (first session black, second session gray) are shown for two representative subjects (circles and triangles) and the group mean (diamonds).

by contrast analyses, this was earlier than [i] (130.1 ms,  $P < 0.001$ ) and [e] (122.6 ms, at trend-level  $P = 0.07$ ).

RMS peak amplitude of the N100m component also yielded a main effect of vowel type [ $F(2,20) = 12.58$ ,  $\epsilon = 0.89$ ,  $P < 0.001$ ]: N100m amplitudes in response to [a] (116 fT) were significantly attenuated compared to [e] (141 fT,  $P < 0.001$ ) and [i] (138 fT,  $P < 0.01$ ) (Fig. 1).

This amplitude difference was reproduced by statistical analysis of the ECD sources fitted to the N100m component: the source strength  $|Q|$  showed a main effect of vowel type [ $F(2,20) = 5.08$ ,  $\epsilon = 0.81$ ,  $P = 0.02$ ]. It was strongest for [e]-responses (20.3 nAm) and differed significantly from [a] source strength (16.6 nAm,  $P = 0.03$ ). Mean source strength of [i] responses amounted to 18.0

nAm and did not differ from [a] and [e] in contrast analyses.

### 3.2. Source locations

For eleven subjects, statistical analysis of source locations was performed<sup>1</sup>. Group data and data for two individual cases are illustrated in Fig. 2b. Source locations

<sup>1</sup>Due to a technical error in head shape acquisition, one female subject had to remain excluded from statistical analysis of source location parameters that included the factor experimental session (but not from calculation of Euclidean distances).

along the medial–lateral axis did not show any significant effects of vowel type, hemisphere or experimental session.

Analyzing source locations, a main effect of vowel type on inferior–superior source location was found [ $F(2,18)=4.48$ ,  $\epsilon=0.74$ ,  $P=0.04$ ]: The [a] source was located most superior (5.77 cm), significantly superior to the [e] source (5.54 cm,  $P=0.04$ ), whereas the [i] source (5.63 cm) was located in-between.

A number of error sources add to the variance in absolute source location. The systematic errors can be reduced when comparing relative distances. Therefore, individual maps of vowel sources that reflect relative distances between the different vowel representations prove to be more reliable than absolute locations within the head frame.

### 3.3. Euclidean distances

Euclidean distances between individual [a], [e] and [i] source locations were submitted to a repeated measures ANOVA with factors hemisphere (left, right), experimental session (first, second) and source distance type ([a]–[i], [e]–[i]).

The representational centers of the two vowels with the greatest distance in  $F_1$ – $F_2$  space, i.e. [a] and [i], showed the largest Euclidean distances ( $M \pm S.D.$   $0.86 \pm 0.59$  cm) (Fig. 3a). [e] and [i] with a smaller distance in a  $F_1$ – $F_2$  space had also closer representational foci ( $0.65 \pm 0.47$  cm) in the cortex [ $F(1,10)=5.8$ ,  $P<0.05$ ] (Fig. 3a). This difference in generator distances was also significant when only the projections on the sagittal plane were entered into the analysis [ $F(1,10)=6.8$ ,  $P<0.05$ ]. This effect did not interact with the hemisphere tested and was equally present in both experimental sessions: The retest reliability of distances between the locations of vowel representation in the sagittal plane was strong and statistically significant across experimental sessions ( $0.62 < r_{\text{SPEARMAN}} < 0.87$ , all  $P<0.05$ ) (Fig. 3b). That is, subjects reproduced a highly

similar pattern of vowel source distances in the second experimental session even when deviating from the expected pattern.

## 4. Discussion

Main rationale of our study was to further test topographic mapping of vowels on the cortical surface by means of magnetic source imaging. Specifically, we tested the hypothesis that larger spectral and phonological dissimilarities would be mirrored by larger Euclidean distances between dipole sources fitted to the N100m component of the brain magnetic response.

Our results yielded main effects of vowel type on RMS peak latency, RMS peak amplitude, dipole source strength and inferior–superior dipole location. The peak latency findings are consistent with earlier work [8,22,26]. In particular, Roberts et al. argued that a fast-peaking N100m in response to the vowel [a] might reflect the vowel's typical spectral pattern with two formant peaks around 1 kHz – the frequency band to which the human auditory system is most sensitive. An attenuated amplitude in response to [a], as presently observed, has not been reported previously, but when applying principles of formant inhibition as proposed by Ohl and Scheich [18] as well as Diesch and Luce [5], the closely neighbored formant peaks  $F_1$  and  $F_2$  in [a] could be expected to elicit an extracranially weaker amplitude than vowels like [i] with large  $F_1$ – $F_2$  differences.

The present source analysis indicates that the distances in phonological space are mirrored in the arrangement of the cortical representational centers. A number of reasons may have contributed to the present finding: first, a relatively large number of trials (up to 480 epochs) were averaged. Second, alertness and arousal were well controlled and identical for the different vowel presentations. Both effects improve the signal-to-noise ratio. Third, the

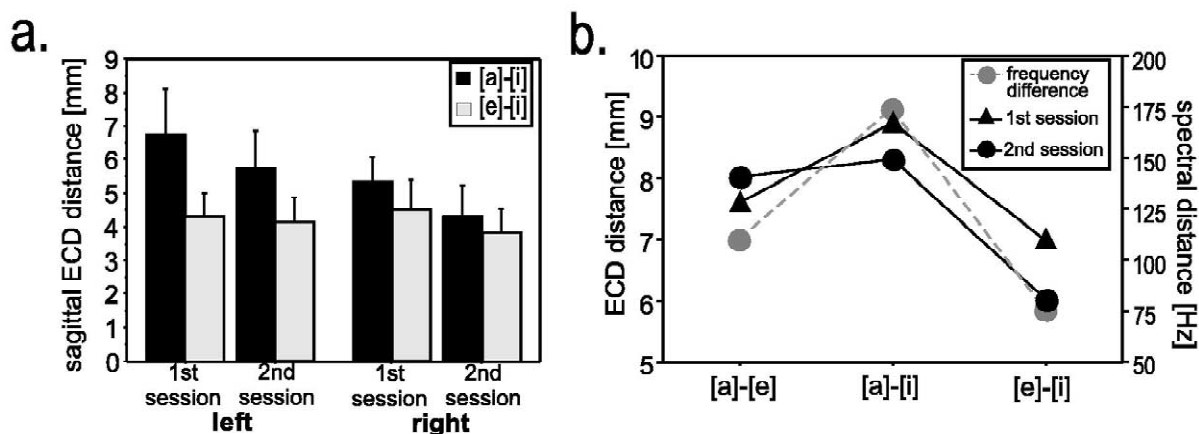


Fig. 3. (a) Mean Euclidean distances ( $\pm$ S.E.M.) of [a]–[i] and [e]–[i] in the sagittal plane are displayed separately for both hemispheres and experimental sessions. (b) Comparison of spectral dissimilarity (gray dashed line) of the stimuli and measured ECD distances for both experimental sessions.

analysis of Euclidean distances looks at the distances between cortical centers irrespectively of their orientation with respect to the anatomy, i.e. it might yield results even if the cortical phonemotopic map is rotated with varying degrees in different subjects [3]. While the results are consistent with the notion of a phonemotopic arrangement of vowel representations in the human cortex, they, of course, do not fully prove the existence of such a map. However, if these maps exist, they may differ individually in their orientation with respect to cranial landmarks.

Euclidean distances account for such individual variations and consequently, the distance between the cortical representations of the dissimilar vowels [a] and [i] was found to be statistically larger than between the representations of the phonologically more similar vowels [e] and [i]<sup>2</sup>.

While the particular spatial configurations of cortical vowel representations vary across subjects, a high retest reliability of this map suggests an individually stable phonemotopy that reappears at different days of testing. Various sources of noise (biological as well as technical) contribute to variations in single ECD locations (see Fig. 2b). As some of the error is systematic (like determining the exact positioning of the cranium), differences in ECD locations are less reliable than relative distance patterns.

The representation of distance patterns (great vs. small spectral dissimilarities) may allow for the high variance in the speech signal a listener encounters: changes in loudness, pitch or speaker identity influence absolute topography and timing of the auditory brain response [16,27,31]. A specific phoneme can therefore not be expected to activate always a circumscribed hard-wired cortical patch. In contrast, preserved relative distances between cortical centers could enable identification of invariant characteristics and further processing of speech sounds in a vast variety of acoustic conditions.

This view is supported by the work of Ohl and Scheich [18] implying that the parameter mapped onto the cortical sheet is  $F_1$ – $F_2$  distance, and that there seem to be no isolated cortical patches that are specifically activated by a particular vowel.

The present observation of a significant vowel distance difference in the sagittal plane is also in line with Ohl and Scheich's as well as Diesch and Luce's notion of a  $F_1$ – $F_2$  difference mapping perpendicular to a tonotopic gradient [5,18], which in humans is typically greatest along the medial–lateral axis. If vowels with varying  $F_1$ – $F_2$  differences activate cortical patches of varying size along the isofrequency stripes, this could result in different dipole locations (which reflect the center of gravity of a given

activation) that should vary to a great extent perpendicularly to the major tonotopic axis.

Our finding of high interindividual variance in absolute ECD locations matches the results of recent studies in language acquisition [2,14,23] that underline the experience-dependent shaping of phoneme representation in early childhood. When representations of stimuli are formed in sensory systems, gross anatomical substrates are probably predetermined by genetic programs and the brain's architecture (e.g. the preference of belt and parabelt areas in auditory cortices for spectrally complex sounds and the distinction of a 'what' and a 'where' system, cf. [10,33]). In contrast, highly experience-dependent representations like the distinction of different phonemes develop individually and self-organizing in the maturing brain (for a review see [25]).

While our results indicate reliable spatial arrangements of cortical vowel representations, we cannot address the question whether the spectral, the phonological or both characteristics are the major determinants of this phonemotopic map: formant frequencies are strongly correlated with phonological features such as place of articulation (corresponds to  $F_2$ ) and tongue height ( $F_1$ ). Our finding of a greater [a]–[i] distance for N100m source topography is compatible with a purely spectral extraction algorithm as well as with a more language-influenced phonological feature extraction algorithm, as the German vowels [a], [e] and [i] cluster in a similar manner when applying a spectral or a phonological categorization. The question arises whether non-native phonemes would fit into a N100m map as we found it by solely mirroring their spectral properties or whether auditory processing as early as the N100m is shaped by learned phonological feature categories (as implied by [2,17]). If the latter holds true, one would expect that the extraction of a feature from the auditory signal predicts the place a phoneme is assigned to in a cortical map [15].

In sum, N100m peak latency, amplitude and generator location demonstrate that different vowels elicit differential brain responses. The representational centers of the auditory N100m component – determined via an ECD model – revealed relative distances that resemble a  $F_1$ – $F_2$ -vowel space and indicate a phonemotopic organization in the supratemporal plane. Our findings suggest a cortical map of vowels that mirrors spectral dissimilarities and corresponds to abstract phonological features. Over repeated measures, the relative distances in this cortical vowel topography are preserved. Further studies of additional vowels may examine whether the N100m location differences reflect primarily spectral or phonological representations or a mixture of both.

## Acknowledgements

This research was supported by the Deutsche For-

<sup>2</sup>As the remaining distance [a] to [e] is fully determined when knowing both other distances in this vowel triangle, the statistical tests were confined to comparing the hypothetically largest with the hypothetically smallest distance.

schungsgemeinschaft and the Volkswagen-Stiftung. The authors wish to thank Michaela Schlichtling, Ursula Lommen and Isabella Paul for their help in data acquisition. We especially thank Eugen Diesch for supplying the stimulus material.

## References

- [1] D.V. Buonomano, M.M. Merzenich, Cortical plasticity: from synapses to maps, *Annu. Rev. Neurosci.* 21 (1998) 149–186.
- [2] M. Cheour, R. Ceponiene, A. Lehtokoski, A. Luuk, J. Allik, K. Alho, R. Naatanen, Development of language-specific phoneme representations in the infant brain, *Nat. Neurosci.* 1 (1998) 351–353.
- [3] E. Diesch, C. Eulitz, S. Hampson, B. Ross, The neurotopography of vowels as mirrored by evoked magnetic field measurements, *Brain Lang.* 53 (1996) 143–168.
- [4] E. Diesch, T. Luce, Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation, *Psychophysiology* 34 (1997) 501–510.
- [5] E. Diesch, T. Luce, Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex, *J. Cogn. Neurosci.* 12 (2000) 878–893.
- [6] J.J. Eggermont, Between sound and perception: reviewing the search for a neural code, *Hear. Res.* 157 (2001) 1–42.
- [7] T. Elbert, H. Flor, N. Birbaumer, S. Knecht, S. Hampson, W. Larbig, E. Taub, Extensive reorganization of the somatosensory cortex in adult humans after nervous system injury, *NeuroReport* 5 (1994) 2593–2597.
- [8] C. Eulitz, E. Diesch, C. Pantev, S. Hampson, T. Elbert, Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli, *J. Neurosci.* 15 (1995) 2748–2755.
- [9] B. Høse, G. Langner, H. Scheich, Linear phoneme boundaries for German synthetic two-formant vowels, *Hear. Res.* 9 (1983) 13–25.
- [10] J.H. Kaas, T.A. Hackett, M.J. Tramo, Auditory processing in primate cerebral cortex, *Curr. Opin. Neurobiol.* 9 (1999) 164–170.
- [11] D.H. Klatt, Software for a cascade/parallel formant synthesizer, *J. Acoust. Soc. Am.* 67 (1980) 971–995.
- [12] T. Kohonen, R. Hari, Where the abstract feature maps of the brain might come from, *Trends Neurosci.* 22 (1999) 135–139.
- [13] N. Kraus, M. Cheour, Speech sound representation in the brain, *Audiol. Neurotol.* 5 (2000) 140–150.
- [14] P.K. Kuhl, A new view of language acquisition, *Proc. Natl. Acad. Sci. USA* 97 (2000) 11850–11857.
- [15] A. Lahiri, H. Reetz, Underspecified recognition, in: C. Gussenhoven, N. Warner, T. Rietveld (Eds.), *Labphon VII*, Mouton, Berlin, 2002, pp. 637–675.
- [16] G. Langner, M. Sams, P. Heil, H. Schulze, Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography, *J. Comp. Physiol. A* 181 (1997) 665–676.
- [17] R. Naatanen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huottilainen, A. Iivonen, M. Vainio, P. Alku, R.J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, K. Alho, Language-specific phoneme representations revealed by electric and magnetic brain responses, *Nature* 385 (1997) 432–434.
- [18] F.W. Ohl, H. Scheich, Orderly cortical representation of vowels based on formant interaction, *Proc. Natl. Acad. Sci. USA* 94 (1997) 9440–9444.
- [19] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, *Neuropsychologia* 9 (1971) 97–113.
- [20] C. Pantev, O. Bertrand, C. Eulitz, C. Verkindt, S. Hampson, G. Schuierer, T. Elbert, Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings, *Electroencephalogr. Clin. Neurophysiol.* 94 (1995) 26–40.
- [21] G. Peterson, H. Barney, Control methods used in a study of the vowels, *J. Acoust. Soc. Am.* 24 (1952) 175–184.
- [22] D. Poeppel, C. Phillips, E. Yellin, H.A. Rowley, T.P. Roberts, A. Marantz, Processing of vowels in supratemporal auditory cortex, *Neurosci. Lett.* 221 (1997) 145–148.
- [23] L. Polka, J.F. Werker, Developmental changes in perception of nonnative vowel contrasts, *J. Exp. Psychol. Hum. Percept. Perform.* 20 (1994) 421–435.
- [24] J.P. Rauschecker, Parallel processing in the auditory cortex of primates, *Audiol. Neurotol.* 3 (1998) 86–103.
- [25] J.P. Rauschecker, Auditory cortical plasticity: a comparison with other sensory systems, *Trends Neurosci.* 22 (1999) 74–80.
- [26] T.P. Roberts, P. Ferrari, S.M. Stufflebeam, D. Poeppel, Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception, *J. Clin. Neurophysiol.* 17 (2000) 114–129.
- [27] G.L. Romani, S.J. Williamson, L. Kaufman, Tonotopic organization of the human auditory cortex, *Science* 216 (1982) 1339–1340.
- [28] C.E. Schreiner, Spatial distribution of responses to simple and complex sounds in the primary auditory cortex, *Audiol. Neurotol.* 3 (1998) 104–122.
- [29] M. Steinschneider, D. Reser, C.E. Schroeder, J.C. Arezzo, Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey, *Brain Res.* 674 (1995) 147–152.
- [30] H. Sussman, Phonemic representation: a twenty-first century challenge, *Brain Lang.* 71 (2000) 237–240.
- [31] N. Titova, R. Naatanen, Preattentive voice discrimination by the human brain as indexed by the mismatch negativity, *Neurosci. Lett.* 308 (2001) 63–65.
- [32] X. Wang, M.M. Merzenich, R. Beitel, C.E. Schreiner, Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics, *J. Neurophysiol.* 74 (1995) 2685–2706.
- [33] C.M. Wessinger, J. VanMeter, B. Tian, J. Van Lare, J. Pekar, J.P. Rauschecker, Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging, *J. Cogn. Neurosci.* 13 (2001) 1–7.