

# Vowel Sound Extraction in Anterior Superior Temporal Cortex

Jonas Obleser,<sup>1</sup> Henning Boecker,<sup>2</sup> Alexander Drzezga,<sup>2</sup>  
Bernhard Haslinger,<sup>3</sup> Andreas Hennenlotter,<sup>3</sup> Michael Roettinger,<sup>3</sup>  
Carsten Eulitz,<sup>1,4</sup> and Josef P. Rauschecker<sup>5\*</sup>

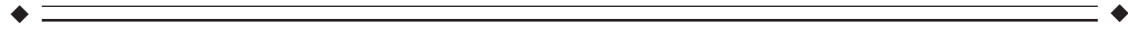
<sup>1</sup>Fachgruppen Psychologie und Linguistik, Universität Konstanz, Konstanz, Germany

<sup>2</sup>Nuklearmedizinische Klinik, Technische Universität München, München, Germany

<sup>3</sup>Neurologische Klinik, Technische Universität München, München, Germany

<sup>4</sup>Klinik für Psychiatrie und Psychotherapie, Universitätsklinikum der RWTH Aachen, Aachen, Germany

<sup>5</sup>Laboratory of Integrative Neuroscience and Cognition, Georgetown University School of Medicine, Washington, DC, USA



**Abstract:** We investigated the functional neuroanatomy of vowel processing. We compared attentive auditory perception of natural German vowels to perception of nonspeech band-passed noise stimuli using functional magnetic resonance imaging (fMRI). More specifically, the mapping in auditory cortex of first and second formants was considered, which spectrally characterize vowels and are linked closely to phonological features. Multiple exemplars of natural German vowels were presented in sequences alternating either mainly along the first formant (e.g., [u]-[o], [i]-[e]) or along the second formant (e.g., [u]-[i], [o]-[e]). In fixed-effects and random-effects analyses, vowel sequences elicited more activation than did nonspeech noise in the anterior superior temporal cortex (aST) bilaterally. Partial segregation of different vowel categories was observed within the activated regions, suggestive of a speech sound mapping across the cortical surface. Our results add to the growing evidence that speech sounds, as one of the behaviorally most relevant classes of auditory objects, are analyzed and categorized in aST. These findings also support the notion of an auditory “what” stream, with highly object-specialized areas anterior to primary auditory cortex. *Hum Brain Mapp* 27:562–571, 2006. © 2005 Wiley-Liss, Inc.

**Key words:** fMRI; speech; vowel; auditory; temporal lobe



## INTRODUCTION

The great ease with which the human brain decodes and categorizes vastly different sounds of speech is one of the key enigmas of the faculty of language. Which, if any, structures of the auditory pathway are specialized in categorizing different classes of speech sounds, and how they are organized functionally is largely unknown. Traditional views of speech perception, based mainly on the study of stroke patients, have implicated posterior regions of the superior temporal gyrus (pSTG), including the planum temporale, in these functions [Geschwind and Levitsky, 1968; Wernicke, 1874]. However, there is growing evidence that areas in the anterior parts of the superior temporal cortex are of even greater relevance to this extraordinary capability of extracting and classifying species-specific vocalizations, including speech sounds in humans.

Contract grant sponsor: Deutsche Forschungsgemeinschaft; Contract grant number: FOR 348; Contract grant sponsor: Center for Junior Research Fellows, University of Konstanz, Germany; Contract grant sponsor: Landesstiftung Baden-Wuerttemberg; Contract grant sponsor: National Science Foundation; Contract grant number: 0350041; Contract grant sponsor: Tinnitus Research Consortium; Contract grant sponsor: Humboldt Foundation.

\*Correspondence to: Dr. Josef P. Rauschecker, Department of Physiology and Biophysics, Georgetown University Medical Center, The Research Building WP15, 3970 Reservoir Road NW, Washington, DC 20057.

Received 25 March 2005; Accepted 12 July 2005

This assumption draws upon recent evidence from both human brain imaging and animal neurophysiology for at least two segregated processing streams within the auditory cortical system. The evidence suggests the existence of an auditory “what” system specialized in identifying and categorizing auditory objects, among which speech sounds form a behaviorally relevant class [Rauschecker and Tian, 2000]. This processing stream is believed to comprise areas of anterior superior temporal gyrus (aSTG) [Tian et al., 2001; Zatorre et al., 2004; Zielinski and Rauschecker, 2000] and superior temporal sulcus (STS) [Binder et al., 2000, 2004; Scott et al., 2000, Scott and Johnsrude, 2003], which project on to inferior frontal areas [Romanski et al., 1999; Hackett et al., 1999]. A recent meta-analysis of 27 human imaging studies that involved nonspatial auditory object processing is consistent with these ideas [Arnott et al., 2004]: Areas in anterior STG/STS and inferior frontal cortex generally show greater activation when tested with object-related auditory stimuli and tasks than when confronted with spatially differentiated stimuli or spatial localization tasks.

Beginning with the seminal work of Petersen et al. [1988], neuroimaging studies have made a substantial contribution to understanding the functional neuroanatomy of speech perception [e.g., Fiez et al., 1995; Scott et al., 2000]. The most important thread of research for the present study comprises work that has contrasted speech sounds or running speech (i.e., auditory objects) with unintelligible speech, backward speech or nonspeech sounds. What has been found consistently in such studies is a greater responsiveness of the superior temporal cortex to speech compared to nonspeech sounds [Binder et al., 1997, 2000; Jäncke et al., 2002; Scott et al., 2000]. In general, nonprimary auditory cortex seems to be increasingly activated by increasingly complex natural stimuli, which suggests a hierarchy of processing levels within object-related auditory cortex from core and belt areas [Kaas and Hackett, 2000; Rauschecker and Tian, 2000] to more anterior areas in STG and STS [Binder et al., 2000; Hart et al., 2003; Wessinger et al., 2001].

There are few studies that have utilized neuroimaging to scrutinize the mapping of the native speech sound inventory in primary and nonprimary auditory cortex. Using magnetic source imaging (MSI), it was demonstrated that shifts of the N100m evoked-field activity in the superior temporal gyrus were dependent on the nature of the vowel or consonant-vowel syllable, both on a single-subject basis [Eulitz et al., 2004; Obleser et al., 2003a] and on a group basis when using more extensive, i.e., more natural sets of stimuli [Obleser et al., 2003b, 2004]. Two other recent investigations also found N100m source displacements in STG for different classes of vowels [Makela et al., 2003; Shestakova et al., 2004].

Despite these promising results from MSI, cortical maps representing features of auditory objects and speech sounds are most likely to be resolved using functional magnetic

resonance imaging (fMRI) techniques with their superior spatial resolution. The mapping of stop consonant syllables was investigated [Rauschecker and Tian, 2000; Zielinski and Rauschecker, 2000] and differential activation patterns for various syllable categories, which activated contiguous but overlapping areas in the anterior STG, were identified. Jäncke et al. [2002] compared vowels and consonant-vowel (CV) syllables to pure tones and white noise to identify brain regions specifically involved in phonetic perception, and they also found evidence for neuronal assemblies especially sensitive to speech in the STS. A very recent study by Callan et al. [2004] confirmed activation of anterior STG/STS when native and nonnative speakers listened to consonant-vowel syllables and carried out a discrimination task, and found that this region shows more activation in first-language than in second-language speakers.

Completely unresolved are the organizing principles upon which any functional maps of speech sounds in the anterior STG/STS might be structured. Physically, different vowel categories are distinguished by their spectral properties, as characterized mainly by the first and second formants, i.e., restricted bands of high spectral energy created by the resonance properties of the vocal tract and its articulators [Peterson and Barney, 1952]. These spectral characteristics of a vowel can be considered as features of an auditory object, and we would expect them to be mapped within areas of the auditory cortex that are specialized for object identification and categorization, as described above.

Here, we compared vowels that varied along a given feature dimension while keeping other feature characteristics constant. Our study posed two specific questions. First, where exactly in the human brain is the perceptual segregation of different vowel categories accomplished? Second, are these stimulus dimensions implemented topographically in an orthogonal manner, as they vary orthogonally in natural speech (Fig. 1)? Most specifically, one would expect a topographic segregation between two vowel categories (Front vowels, such as [i] and [e], and Back vowels, such as [o] and [u]) that are characterized by widely distinct values of their (perceptually relevant) second formant.

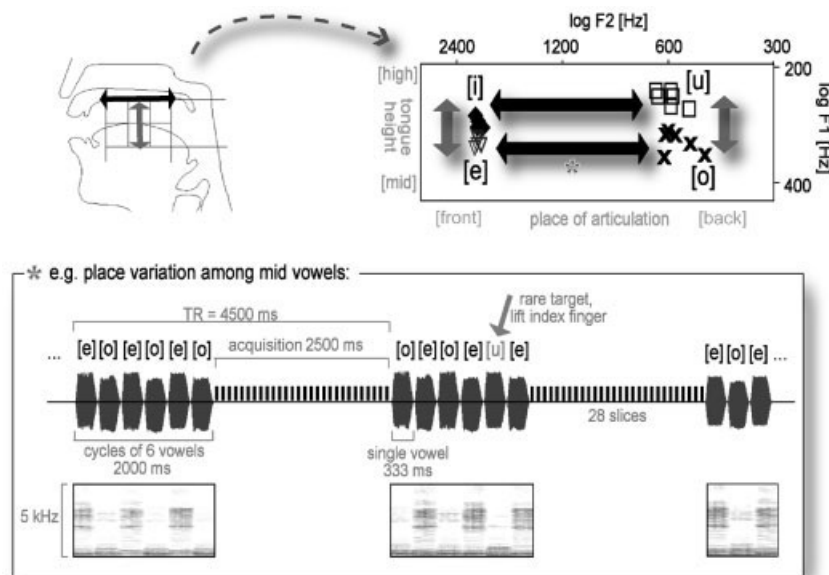
## SUBJECTS AND METHODS

### Subjects

Thirteen healthy right-handed participants (5 women; mean age  $\pm$  standard deviation [SD],  $31 \pm 5$  years) underwent fMRI scanning. They signed an informed consent form and received 20 Euros. The local ethics committee of the Technical University Munich approved all experimental procedures.

### Experimental Design

The study was set up to investigate the processing of vowels and their main feature dimensions: place of articulation (Place, mainly a variation of the second formant,  $F_2$ ) and tongue height (Height, a variation of the first formant,



**Figure 1.**

Upper panel, right:  $F_1, F_2$ -vowel space of the vowel stimuli used. Note the spectral properties of the vowel categories and their correspondence to the articulatory and phonological features “place of articulation” and “tongue height” (left). Lower panel: Illustration of the experimental setup with corresponding oscillograms and spectrograms, exemplified on the [e]-[o] condition.

$F_1$ ). To study the brain structures involved in extracting these feature dimensions, we generated alternating sequences of natural German vowels that oscillated either within the  $F_2$  or the  $F_1$  dimension (Fig. 1). All vowels (eight different tokens per vowel category) were edited from spoken words of a single male speaker (duration 333 ms; 50-ms Gaussian on- and offset ramp) and were free from any formant transitions. Pitch varied within the male voice’s range but did not differ systematically between vowel categories (Table I; for a more extensive description of the raw vowel material see Obleser et al. [2004]).

Two vowel conditions were conceived to track  $F_2$  or Place distinctions. Condition [u]-[o] (hereafter called Back condition) included vowels with an almost constant  $F_2$  in a range that is typical for male German Back vowels (450–650 Hz) and varied only within the  $F_1$  dimension. Correspondingly, condition [i]-[e] (Front) also varied mainly along  $F_1$ , but at much higher  $F_2$  values. Comparing brain activation in these two conditions would allow us to track voxels that are involved more strongly in coding one or the other  $F_2$  or Place condition. Two other conditions were to reflect mainly differences in  $F_1$  or Height: Condition [u]-[i] (High) included vowels with an almost constant and low  $F_1$  that varied within the  $F_2$  dimension. Correspondingly, condition [o]-[e]

(Mid) also varied mainly within  $F_2$ , but at higher  $F_1$  values (Fig. 1; Table I).

Subjects were instructed to listen carefully to the vowel alternations (alternation rate 3 Hz) and to detect rare target vowels that violated the vowel sequence pattern. For example, in the [u]-[o] sequence, this was either [i] or [e]. Subjects were expected to lift their right (or their left, counterbalanced across subjects) index finger whenever they detected such a violation to the vowel sequence. Targets appeared in 8.6% of vowel sequences (i.e., less than 2% of all single vowels were violating targets).

In total, each vowel sequence condition lasted for about 8 min (i.e., 104 vowel sequences). An additional functional reference condition with duration- and energy-matched band-passed noise (BPN) stimuli of various center frequencies (randomized presentation of BPNs with one-third octave bandwidth and center frequencies of 0.5, 1, 2, 4, and 8 kHz; stimulus length 2 s) was presented. Noise bursts were also ramped with 50-ms Gaussian on- and offsets. The BPN condition required subjects to listen to the noise bursts passively and was identical to vowel conditions in all acquisition parameters.

**TABLE I. Range of pitch (F0) and formant frequencies, and the assignment of phonological features in the vowel categories used**

Vowel	Place of articulation	Tongue height	Formant frequencies min-max (Hz)			
			F0	F1	F2	F3
[i]	Front	High	127–132	267–287	2,048–2120	2,838–3,028
[e]	Front	Mid	109–125	302–322	2,055–2143	2,711–2,890
[u]	Back	High	112–118	231–256	522–645	2,117–2,292
[o]	Back	Mid	109–125	293–346	471–609	2,481–2,688

## Functional MRI Scanning

Echo planar imaging (EPI) was carried out on a Siemens Magnetom Symphony scanner with 1.5-T field strength, using the standard head coil for radio-frequency transmission and signal reception. Stimulus sequences were triggered by the MR system and were presented using *Presentation v0.6* software (Neurobehavioral Systems) and an air conduction sound delivery system with sound-attenuated headphones at approximately 60 dB SPL.

One hundred and four whole-brain volumes of a  $T_2^*$ -weighted EPI scan with transverse orientation were acquired (repetition time [TR] = 4,500 ms; acquisition time [TA] = 2,505 ms, echo time [TE] = 50 ms, flip angle 90 degrees, 28 slices with 5-mm thickness and 0.5-mm gap, field of view  $64 \times 64$ , 192 mm, voxel size  $3 \times 3 \times 5$  mm<sup>3</sup>, orientation parallel to anterior commissure–posterior commissure [AC–PC], slice order interleaved) per vowel sequence conditions.

Using a TR of 4.5 s, our design strikes a compromise between continuous scanning (TR ~2 s) and extremely sparse sampling designs with a TR > 12 s [Gaab et al., 2003; Hall et al., 1999]. Like the longer TRs, such an intermediate TR avoids auditory masking of stimuli through scanner noise but minimizes total acquisition time. Although there remains potential overlap between the hemodynamic responses to scanner noise and stimuli, this does not seem to lead to a significant reduction in signal, as previous auditory fMRI studies of speech using comparable designs and TRs have demonstrated [e.g., Jäncke et al., 2002; Van Atteveldt et al., 2004].

All four vowel sequence conditions were presented in a counterbalanced order. After accomplishing the four vowel sequences, a high-resolution anatomical  $T_1$ -weighted scan (TR = 11.08 ms, TE = 4.3 ms, flip angle 15 degrees, voxel size  $1 \times 1 \times 1$  mm<sup>3</sup>, sagittal orientation) plus 104 volumes of the BPN reference condition were acquired. The first and last four images of each run were discarded, leaving 96 volumes of each condition and subject for statistical inference using *SPM2b* (Wellcome Department of Imaging Neuroscience; <http://www.fil.ion.ac.uk>). Before statistical inference, all images were slice-time corrected, realigned to the first image acquired, co-registered with subjects' individual high-resolution  $T_1$ -weighted anatomical scans, and normalized using a Talairach-transformed template image. All images were smoothed using a non-isotropic  $8 \times 8 \times 12$  mm<sup>3</sup> Gaussian kernel.

Data from 13 participants were subjected to statistical analysis. At the single-subject level, a boxcar design model with five conditions (Back, Front, High, Mid, nonspeech BPN) plus the individual realignment parameters as additional regressors was estimated (no high-pass filter was applied due to the low-frequency cycling of conditions; time series were low-pass filtered with a Gaussian full-width half-maximum [FWHM] of 4 s, and autocorrelation of time series was taken into account).

A fixed-effects model with 13 subjects, five conditions, and the individual realignment parameters as additional

regressors was calculated. Specific contrasts of conditions were masked inclusively by an “all-vowel conditions > BPN” contrast ( $P < 10^{-2}$ ) and thresholded at a  $P < 10^{-6}$  level (cluster extent  $k \geq 30$ ). We also calculated a group-level analysis (random-effects model). Herein, main-effect contrast images from the single-subject analyses were submitted to one-way repeated-measures analysis of variance (ANOVA; *SPM2b*). Vowel conditions were compared to the reference nonspeech BPN condition in contrasts at a  $P < 10^{-3}$  (uncorrected, cluster extent  $k \geq 3$ ) threshold. (Pertaining to the problem of multiple comparisons and Type I error inflation, estimating just one repeated-measures ANOVA for all conditions is supposed to be less affected by such bias than is estimating  $n$  separate random-effects models for  $n$  contrasts of interest.)

## RESULTS

### Fixed-Effects Analysis

The principal contrast of interest was a comparison of all vowel-related activation to activation in the BPN condition: Focused activations in the anterior superior gyri of both hemispheres were found, plus activations of the middle temporal and inferior frontal gyrus (IFG; Fig. 2). The center of activation in the aSTG was also evident in all single vowel conditions, most strongly in the right hemisphere. The IFG also seemed to be involved in Front and Back vowel processing; however, no suprathreshold activation was evident there in High and Mid vowel conditions. No primary auditory cortex activation was observed.

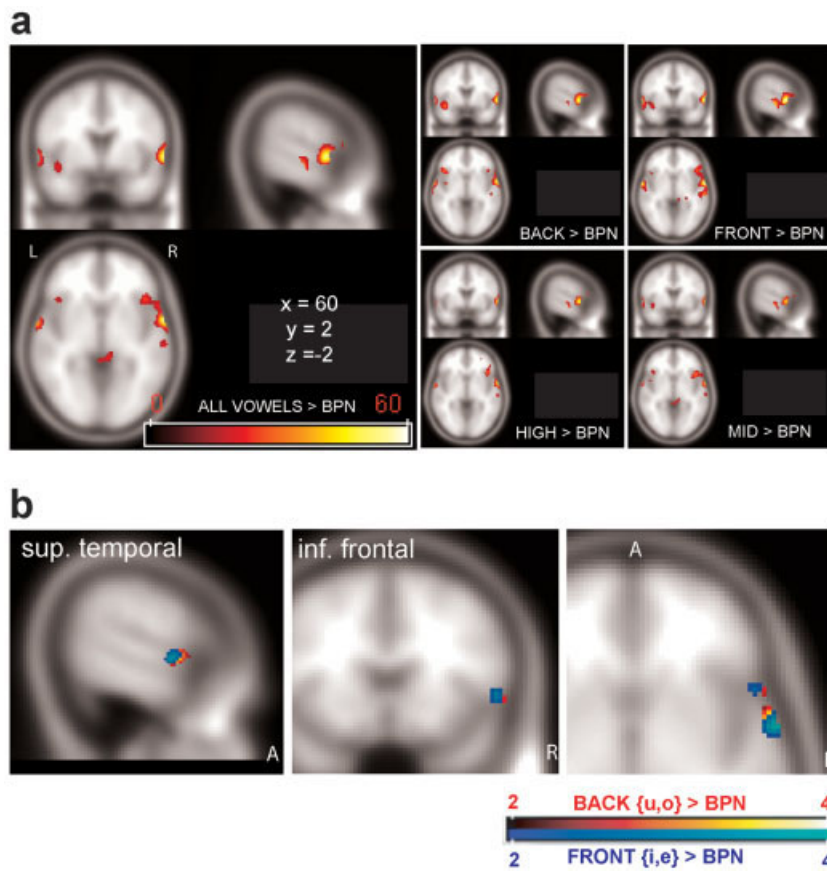
### Random-Effects Analysis

Most notably, suprathreshold activations in the random-effects model matched the areas activated in the fixed-effects analysis very closely (Table II; cf. Fig. 2a,b fixed-effects and random-effects, respectively). Activation was also confined to anterior temporal and inferior frontal areas. The right anterior spot in Brodmann area (BA) 22 also appeared as the most reliable activation across all conditions.

### Specific Comparison of Front and Back Vowels

As illustrated in Figures 2b, 3, and 4, partially distinct centers of activation of different classes of vowels could be observed, especially for the widely distinct classes of Front ([i], [e]), and Back ([u], [o]) vowels. The bar plots in Figure 3 show percent of signal change in all five conditions of the random-effects model. Interestingly, more posterior voxels exhibited a preference for the vowel categories [i] and [e] (Front, characterized by high  $F_2$  values) and showed no signal change that was significantly different from zero for Back vowels [u] and [o]. In contrast, the anterior-most voxels showed the reverse pattern (strongest signal change for Back vowels, characterized by low  $F_2$  values). The two other vowel conditions, [u]-[i] (High) and [o]-[e] (mid), which switched back and forth along  $F_2$  between these widely different categories, exhibited no significant signal change compared to the global mean of all conditions.





**Figure 2.**

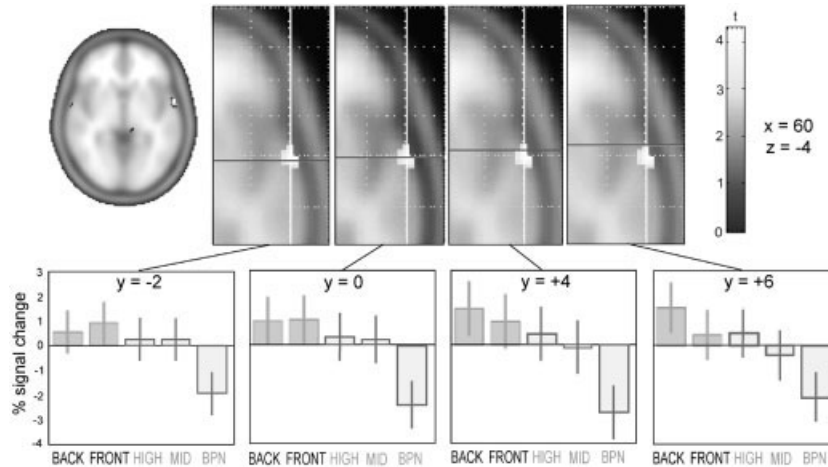
**a:** Overview of overactivations compared to band-passed noise (BPN;  $t$  values) seen in the fixed-effects analysis ( $P < 10^{-6}$ , cluster extent  $k > 50$ ). Left: Activations of all vowel conditions greater than BPN. Smaller panels on the right: activations of single vowel conditions greater than BPN. **b:** Comparison of Front  $>$  BPN (blue colors) and Back  $>$  BPN (red colors) seen in the random-effects analysis ( $P < 10^{-3}$ ,  $k > 3$ ). Although activation overlaps, a change in center of gravity is seen in both anterior superior temporal gyrus (aSTG) and inferior frontal gyrus (IFG).

**TABLE II. Overview of supra-threshold activation in the random-effects analysis after small volume correction**

	Talairach coordinates	Z	P	Approximate anatomical structures*
All vowels $>$ BPN	61, 0, -3	3.93	0.000	Right STG/STS, BA 22
	-63, -2, -2	3.03	0.001	Left STG, BA 22
	53, 15, -4	2.59	0.005	Right IFG
	61, -18, -8	2.36	0.009	Right STS, BA 21
Back $>$ BPN	59, 4, -2	3.69	0.000	Right STG
	-63, -2, -2	2.80	0.003	Left STG/STS, BA22/21
	54, 15, -4	2.59	0.005	Right IFG
	-65, -10, 4	2.58	0.005	Left STG, BA 22
	-63, -10, 7	2.46	0.007	Left MTG, BA 21
Front $>$ BPN	59, 0, -3	3.47	0.000	Right STG/STS, BA 22/21
	51, 16, -1	3.03	0.001	Right IFG
	-63, -2, -2	2.88	0.002	Left STG, BA 22
High $>$ BPN	61, 0, -3	2.97	0.001	Right STG/STS, BA22
Mid $>$ BPN	61, 0, -3	2.95	0.002	Right STG/STS, BA22
	63, -18, -8	2.57	0.007	Right MTG, BA 21

Small volume correction: 20-mm spheres centered at right- and left-hemispheric peak voxels.

\* BA, Brodmann area; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; STG, superior temporal gyrus; STS, superior temporal sulcus.



**Figure 3.**

Upper panel: Axial slices of the standard statistical parametric mapping (SPM) brain template (Montreal Neurological Institute coordinates,  $x = 60, z = -4$ ) with activation in the random-effects analysis overlaid ( $P < 10^{-3}$ ). Lower panels: For the corresponding locations along the supratemporal plane (anterior–posterior dimension; bold line indicating different  $y$ -coordinates), bar plots indicate percent signal change in all five conditions of the random-effects model. The left-most graph displays a preference of the

most posterior voxels for the Front vowel categories [i] and [e] and shows no signal change that is significantly different from zero for Back vowels [u] and [o]. The reverse is seen in the two right-most graphs (most anterior voxels). The two intermediate vowel conditions High ([u]-[i]) and Mid ([o]-[e]) show no significant signal change compared to the global mean; the reference condition of band-passed noise (BPN) bursts displays a relative suppression.

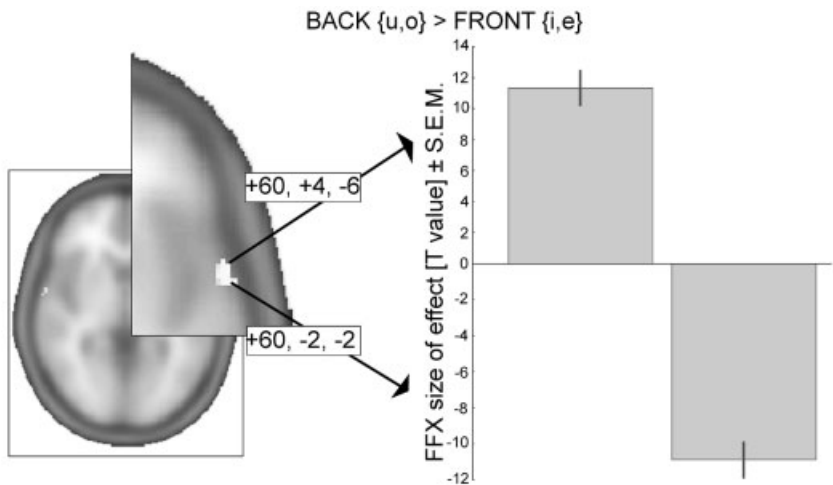
A more stringent demonstration of spatial segregation between the vowel categories is illustrated in Figure 4: The direct statistical comparison of Back and Front conditions attained significance in the more sensitive fixed-effects model. One cluster of activation was significant in the Front  $>$  Back comparison ( $60, -2, -2$ ; cluster size  $k = 19$ , corrected cluster  $P = 0.002, t = 10.77, P < 10^{-11}$ ), and another cluster was found for the reverse case, Back  $>$  Front ( $60, 4, -6$ ; cluster size  $k = 6$ , corrected cluster  $P = 0.003, t = 9.77, P < 10^{-11}$ ; search volume was confined to those voxels identified previously to be involved in speech sound processing, that is  $P$  values were small volume-corrected by the

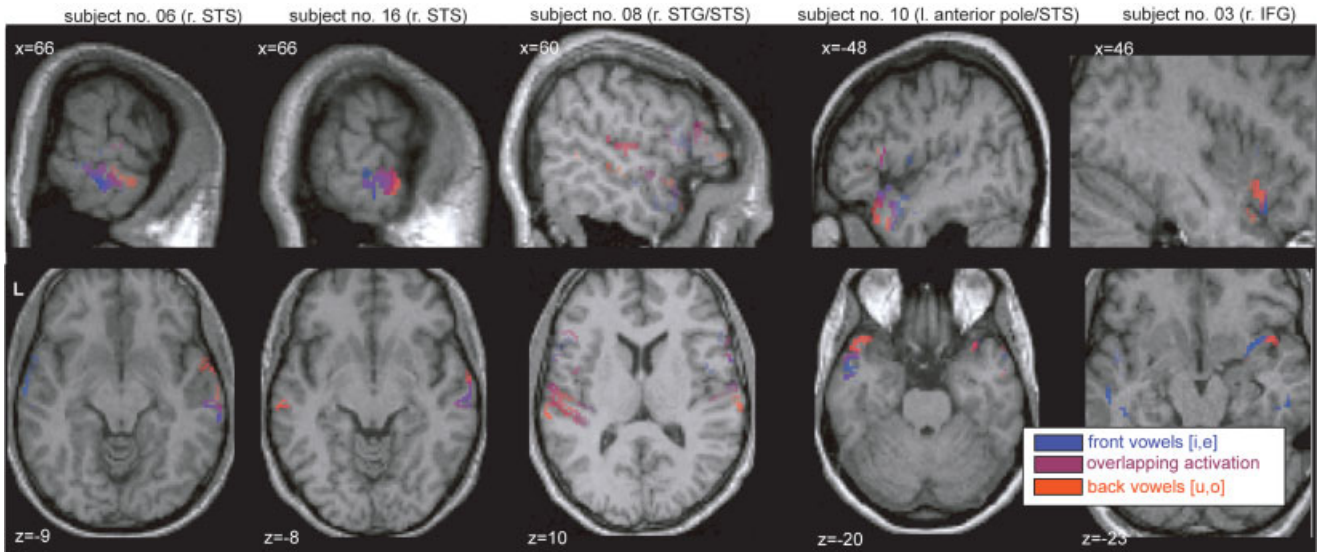
“all-vowel conditions  $>$  band-passed noise” contrast image). The clusters had different centers of gravity and were shifted mainly along the anterior–posterior axis. Adding further evidence to this, a small topographic shift in center of activation was also observed in inferior frontal cortex (in the right hemisphere, cf. Fig. 2b).

Figure 5 illustrates this effect in sagittal and transverse projections of results from five representative single subjects. Within the activation spot in anterior superior temporal cortex, three of five subjects show a clear gradual shift from stronger [i,e]-related activation (shown in blue) to stronger [u,o]-related activation. Among the other subjects,

**Figure 4.**

Direct statistical comparison of Back vowel to Front vowel activation. Given are  $t$ -values ( $P < 10^{-11}$ ) for the clusters in the fixed-effects (FFX) model that survived small-volume correction for the “all-vowels  $>$  band-passed noise” comparison derived from the random-effects model (shown in the right-hemispheric brain activation; right panel). The two clusters differ largely along the posterior–anterior axis (Montreal Neurological Institute coordinates).





**Figure 5.**

Single-subject activations to Back vowels [u,o] (red), Front vowels [i,e] (blue), as well as areas of overlap (shown in purple), plotted onto individual T<sub>1</sub> structural images (neurological convention, left being left). All activations are confined to regions of interest in temporal (Brodmann area [BA] 21, 22, 38, 41, and 42) and frontal

lobe (BA 44 and 45), compared against band-passed noise (BPN) and corrected to have an overall false discovery rate  $P < 0.05$  (cluster extent  $k > 15$ ). Note the gradual shift due to vowel features (from blue to red) in anterior superior temporal cortex (aST).

Subject 8 differs by showing much more prominent activation in the left transverse gyri in this contrast, although some degree of separation is seen also in right aST. Subject 3 shows suprathreshold activation accompanied by a gradual shift of feature-driven responses in right IFG.

## DISCUSSION

In this study, we report a consistent activation of the anterior STG and STS cortical regions when subjects listened to natural German vowels. In comparison to nonspeech BPN stimuli, which mainly activate human analogues of core and belt areas [Wessinger et al., 2001; cf. Kaas and Hackett, 2000; Rauschecker and Tian, 2000], it is these areas anterior to Heschl’s gyrus that seem to contain mainly neuronal assemblies specialized in the processing of simple human communication sounds. This result closely matches Zielinski and Rauschecker’s [2000] demonstration that fMRI mapping of different consonant-vowel syllables reveals activation predominantly anterior of primary auditory cortex, in cortical patches highly similar to the locus of ST activation described here.

What is the functional significance of this activity in the anterior parts of the ST? To track differences between rapidly changing vowels, subjects had to detect spectral changes of either the first or second spectral peak (first or second formant). Evidence is growing that such spectral properties are characteristic features of auditory objects, which are processed predominantly in an anterior “what” stream [Rauschecker and Tian, 2000; Romanski et al., 1999].

A recent study by Zatorre and Belin [2001] contrasted primarily temporal processing with primarily spectral processing using positron emission tomography (PET). This study revealed a bilateral spot in the aSTG that showed strong involvement in spectral processing, although the right aSTG apparently exhibited a preference for primarily spectral differences, as they were also evident between vowels. The same areas have also been shown to be especially responsive to object-like, i.e. discriminable, sounds, which include but are not restricted to speech sounds [Zatorre et al., 2004], although a functional distinction between STG and STS as activated in the latter study remains to be disentangled. Interestingly, a recent study by Warren et al. [2005] showed the right superior temporal cortex (with peak voxels located in both STG and STS) to respond preferably to changes in spectral envelope, the extraction and abstraction of which is most likely a crucial step in vowel identification.

Within the activated area in aST, we obtained clear hints toward a topographic shift of vowel categories. Previous MSI studies reported a statistical difference in source location for Back and Front vowels, presumably also anterior to primary auditory cortex [Obleser et al., 2003a, 2004]. The finding reported here, namely two different centers of gravity for the Back > Front and the Front > Back comparison, reflect a topographic segregation of cortical representations tuned more finely to one or the other class of German vowels. It is these two classes of vowels that are most widely distinct in terms of their acoustic and phonological features (cf. Fig. 1, Table I). In acoustic terms, they differ predomi-

nantly in second formant frequency (by  $\sim 1,500$  Hz in our stimulus material). In phonological terms, vowels such as [u] and [i] have minimal articulatory overlap and will never co-occur at the same time in the speech signal a listener perceives. In other words, they are mutually exclusive and therefore their topographic separation in aST is most likely. A topographic mapping of speech sound features has been proposed on theoretical grounds [Kohonen and Hari, 1999], and the present results are among the first to study and imply such a mapping using fMRI.

It has been argued that speech sound-specific maps may well be subject to high interindividual variance due to individual experience-dependent connectivity in the maturing cortex [Buonomano and Merzenich, 1998; Rauschecker, 1999]. This diminishes the statistical power of a random-effects analysis, which is in principle the method of choice when inference about a population (i.e., native speakers of German) is intended, because the subjects are treated as random samples from a population [Friston et al., 1999]. It also can result in a difficulty to unravel higher-order functional cortical maps using noninvasive methods on a group-statistic level [Diesch et al., 1996; Obleser et al., 2003a]. Even tonotopic maps that are well established from findings in animal research [Kaas et al., 1999; Merzenich and Brugge, 1973; Rauschecker, 1997; Read et al., 2002] are sometimes elusive in human noninvasive magnetoencephalography (MEG) and fMRI research [Formisano et al., 2003; Pantev et al., 1995; Talavage et al., 2000; but see also Lütkenhöner et al., 2003]. In the present study, fixed-effects and random-effects analyses show highly congruent results, although the fixed-effects analysis delivered the necessary sensitivity to unravel the topographical separation between vowel categories (cf. Fig. 4).

Effectively, no decision can be made whether the activation seen here reflects purely acoustic properties of the stimuli (i.e., different ranges of formant frequencies) or whether these cortical representations preferably process human speech sounds. However, one should keep in mind that our reference stimuli were third-octave BPN bursts of different center frequencies, which resemble the acoustic properties of single formants quite closely [Rauschecker and Tian, 2000]. If there were a considerable overlap in the neuronal networks preferring BPN and those preferring human vowels in the aST, we would not have expected as strong an effect in the general “all-vowel conditions  $>$  band-passed noise” comparisons (cf. Fig. 2a).

Pitch variation has also been considered as an influence on aST activation [Patterson et al., 2002]. This corroborates the important role of aST in auditory object recognition. However, the categorical distinction of vowels as reported here can hardly be explained by pitch: the variation in pitch is not systematic between vowel categories, and the maximal pitch difference between vowels amounts to less than 25 Hz. In contrast, formant frequency differences spread over hundreds of Hz (cf. Table I).

## Involvement of IFG

Another open question arises from our results: what functional significance in speech sound decoding does the activation of the IFG bear? The random-effects and the fixed-effects models are inconsistent here, as no IFG voxels were activated according to the random-effects analyses (Table II) in the High and mid vowel conditions. In the fixed-effects model, the right IFG is activated consistently in all vowel conditions (cf. Fig. 2a). The most feasible explanation for this incompatibility arises from a difference between models: in the more sensitive fixed-effects model, strong activation in a few subjects suffices for a significant group effect [Friston et al., 1999]. This may indeed be the case with IFG activation; subjects may have differed in their use of frontal areas to accomplish the vowel discrimination task (cf. Fig. 5). In the random-effects model, the variance across subjects in this brain area may have yielded the inconclusive results across conditions.

Previous imaging studies of speech are inconsistent as to whether IFG involvement is mandatory in receptive language tasks [Chein et al., 2002; Davis and Johnsruide, 2003; Fiez et al., 1995; Fitch et al., 1997]. For example, Chein et al. [2002] pointed out that the IFG may play a vital role in verbal working memory, which in turn could indicate that the variability of IFG activation may be due to effects of the task we employed (subjects had to track ongoing vowel sequences for rare oddball vowels). However, in a recent study where task demands were abolished to mimic natural, effortless speech comprehension (mere listening to simple narrative speech was contrasted with backward versions of the same speech stimuli), Crinion et al. [2003] reported the only extratemporal activation in a confined area of ventrolateral prefrontal cortex (VLPFC). VLPFC also has a role to play in auditory object working memory [Romanski and Goldman-Rakic, 2002] and is anatomically connected directly with anterior lateral belt areas in the superior temporal cortex as part of the auditory “what” stream [Romanski et al., 1999]. In our study, a comparison of active versus passive vowel processing would have been helpful to answer the question of IFG involvement but was omitted from the protocol for reasons of experiment duration.

## CONCLUSIONS

The attentive processing of vowels, which changed along confined feature dimensions, activated mainly regions in the anterior parts of the superior temporal cortical regions (BA 22, 21), with additional, less consistent activation of inferior frontal cortex. This activity pattern matches the notion of an auditory “what” system specialized in object analysis that has been substantiated in nonhuman primates [Tian et al., 2001] as well as humans [Alain et al., 2001; Rauschecker and Tian, 2000]. It is these specialized areas in the anterior parts of the superior temporal cortex [Binder et al., 2004; Liebenthal et al., 2005; Zatorre et al., 2004] that seem to



accomplish the tracking and extraction of vowel features in human speech processing.

## ACKNOWLEDGMENTS

We thank C. Dresel, J. Kissler, M. Junghöfer, and C. Wienbruch for technical support and helpful comments. This work was supported by the Deutsche Forschungsgemeinschaft (FOR 348 to C.E.), the Center for Junior Research Fellows at the University of Konstanz, Germany (to J.O.), a postdoctoral grant by the Landesstiftung Baden-Wuerttemberg (to J.O.), the Cognitive Neuroscience Initiative of the National Science Foundation (0350041 to J.P.R.), the Tinnitus Research Consortium (to J.P.R.), and a Humboldt Research Award (to J.P.R.).

## REFERENCES

- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001): "What" and "where" in the human auditory system. *Proc Natl Acad Sci USA* 98:12301–12306.
- Arnott SR, Binns MA, Grady CL, Alain C (2004): Assessing the auditory dual-pathway model in humans. *Neuroimage* 22:401–408.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000): Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997): Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004): Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295–301.
- Buonomano DV, Merzenich MM (1998): Cortical plasticity: from synapses to maps. *Annu Rev Neurosci* 21:149–186.
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004): Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22:1182–1194.
- Chen JM, Fissell K, Jacobs S, Fiez JA (2002): Functional heterogeneity within Broca's area during verbal working memory. *Physiol Behav* 77:635–639.
- Crinion JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ (2003): Temporal lobe regions engaged during normal speech comprehension. *Brain* 126:1193–1201.
- Davis MH, Johnsrude IS (2003): Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Diesch E, Eulitz C, Hampson S, Ross B (1996): The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain Lang* 53:143–168.
- Eulitz C, Obleser J, Lahiri A (2004): Intra-subject replication of brain magnetic activity during the processing of speech sounds. *Brain Res Cogn Brain Res* 19:82–91.
- Fiez JA, Raichle ME, Miezin FM, Petersen SE, Tallal P, Katz WF (1995): Activation of a left frontal area near Broca's area during auditory detection and phonological access tasks. *J Cogn Neurosci* 7:357–375.
- Fitch RH, Miller S, Tallal P (1997): Neurobiology of speech perception. *Annu Rev Neurosci* 20:331–353.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003): Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859–869.
- Friston KJ, Holmes AP, Worsley KJ (1999): How many subjects constitute a study? *Neuroimage* 10:1–5.
- Hart HC, Palmer AR, Hall DA (2003): Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex* 13:773–781.
- Jäncke L, Wustenberg T, Scheich H, Heinze HJ (2002): Phonetic perception and the temporal cortex. *Neuroimage* 15:733–746.
- Kaas JH, Hackett TA (2000): Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97:11793–11799.
- Kaas JH, Hackett TA, Tramo MJ (1999): Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9:164–170.
- Kohonen T, Hari R (1999): Where the abstract feature maps of the brain might come from. *Trends Neurosci* 22:135–139.
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005): Neural substrates of phonemic perception. *Cereb Cortex* (in press).
- Lütkenhöner B, Krumbholz K, Seither-Preisler A (2003): Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *Neuroimage* 19:935–949.
- Makela AM, Alku P, Tiitinen H (2003): The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neurosci Lett* 353:111–114.
- Merzenich MM, Brugge JF (1973): Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res* 50:275–296.
- Obleser J, Elbert T, Lahiri A, Eulitz C (2003a): Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Res Cogn Brain Res* 15:207–213.
- Obleser J, Lahiri A, Eulitz C (2003b): 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. *J Cogn Neurosci* 16:31–39.
- 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. *Electroencephalogr Clin Neurophysiol* 94:26–40.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002): The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Peterson G, Barney H (1952): Control methods used in a study of the vowels. *J Acoust Soc Am* 24:175–184.
- Rauschecker JP (1997): Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl* 532:34–38.
- Rauschecker JP (1999): Auditory cortical plasticity: a comparison with other sensory systems. *Trends Neurosci* 22:74–80.
- Rauschecker JP, Tian B (2000): Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci USA* 97:11800–11806.
- Read HL, Winer JA, Schreiner CE (2002): Functional architecture of auditory cortex. *Curr Opin Neurobiol* 12:433–440.
- Romanski LM, Goldman-Rakic PS (2002): An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15–16.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999): Dual streams of auditory afferents target

- multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M (2004): Orderly cortical representation of vowel categories presented by multiple exemplars. *Brain Res Cogn Brain Res* 21:342–350.
- Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR (2000): Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res* 150:225–244.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001): Functional specialization in rhesus monkey auditory cortex. *Science* 292:290–293.
- Van Atteveldt N, Formisano E, Goebel R, Blomert L (2004): Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- Warren JD, Jennings AR, Griffiths TD (2005): Analysis of the spectral envelope of sounds by the human brain. *Neuroimage* 24:1052–1057.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001): Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.
- Zatorre RJ, Belin P (2001): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Bouffard M, Belin P (2004): Sensitivity to auditory object features in human temporal neocortex. *J Neurosci* 24:3637–3642.
- Zielinski BA, Rauschecker JP (2000): Phoneme-specific functional maps in the human superior temporal cortex. *Soc Neurosci Abstr* 26:1969.