

# Detection of Differential Speech-Specific Processes in the Temporal Lobe Using fMRI and a Dynamic “Sound Morphing” Technique

Karsten Specht,<sup>1,2\*</sup> Berge Osnes,<sup>1</sup> and Kenneth Hugdahl<sup>1,3</sup>

<sup>1</sup>Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

<sup>2</sup>Clinical Engineering Department, Haukeland University Hospital, Bergen, Norway

<sup>3</sup>Division of Psychiatry and Bergen Mental Health Center, Haukeland University Hospital, Bergen, Norway

---

**Abstract:** Contrary to the classical view, recent neuroimaging studies claim that phonological processing, as part of auditory speech perception, is subserved by both the left and right temporal lobes and not the left temporal lobe alone. This study seeks to explore whether there are variations in the lateralization of response to verbal and nonverbal sounds by varying spectral complexity of those sounds. White noise was gradually transformed into either speech or music sounds using a “sound morphing” procedure. The stimuli were presented in an event-related design and the evoked brain responses were measured using fMRI. The results demonstrated that the left temporal lobe was predominantly sensitive to gradual manipulation of the speech sounds while the right temporal lobe responded to all sounds and manipulations. This effect was especially pronounced within the middle region of the left superior temporal sulcus (mid-STS). This area could be further subdivided into a more posterior area, which showed a linear response to the manipulation of speech sounds, and an anteriorly adjacent area which showed the strongest interaction between the speech and music sound manipulations. Such a differential and selective response was not seen in other brain areas and not when the sound “morphed” into a music stimulus. This gives further experimental evidence for the assumption of a posterior-anterior processing stream in the left temporal lobe. In addition, the present findings support the notion that the left mid STS area is more sensitive to speech signals compared to the right homologue. *Hum Brain Mapp* 30:3436–3444, 2009. © 2009 Wiley-Liss, Inc.

**Key words:** brain; auditory perception; temporal lobe; speech; music; fMRI; laterality

---

## INTRODUCTION

There is broad consensus in the field of neuroscience that speech processes are predominantly lateralized to the left hemisphere, dating back to the ground breaking work

of Broca and Wernicke [see, e.g., Hickok and Poeppel, 2007, for a review]. However, these early accounts of functional asymmetry, which were based on clinical observations, have been challenged. Recent reviews [see e.g., Hickok and Poeppel, 2007] and meta-analyses [see e.g., Indefrey and Cutler, 2004] also take into account results from functional imaging studies and point towards a more bilateral processing of acoustic speech signals. Therefore, current research on the functional asymmetry of speech and music perception focuses, in particular, on temporal lobe structures such as Heschl’s Gyrus, planum temporale (PT), superior temporal gyrus (STG), and superior temporal sulcus (STS). It is now widely accepted that there exists

---

\*Correspondence to: Karsten Specht, Department of Biological and Medical Psychology, University of Bergen, Jonas Lies vei 91, Bergen 5009, Norway. E-mail: karsten.specht@psybp.uib.no

Received for publication 7 July 2008; Accepted 3 February 2009

DOI: 10.1002/hbm.20768

Published online 3 April 2009 in Wiley InterScience (www.interscience.wiley.com).

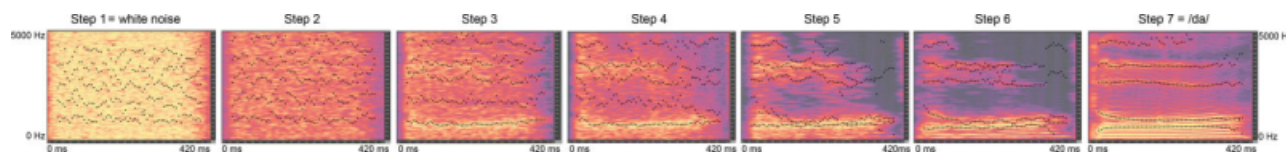


Figure 1.

Example of a “morphed” sound, changing from white noise (left) to a “da” syllable (right). Black dots are displaying the results from a formant extraction, using Praat. Note, that there are now detectable formants in the beginning, but becoming more and more detectable through the morphing procedure.

both a functional, as well as structural, asymmetry within the primary and secondary auditory system. This functional asymmetry is reflected by a higher temporal resolution in the left auditory cortex and higher spectral resolution in the right homologue [Zatorre and Belin, 2001]. Alternatively, one could think of different integration windows of the left and right auditory system, which then may result in an asymmetric sampling and processing of acoustic signals [Boemio et al., 2005]. Anatomical studies have also shown that there are macroscopic, as well as microscopic, structural differences in the primary and secondary auditory cortex. Cytoarchitectonic maps, as well as results obtained with voxel-based morphometry (VBM), have consistently shown that cell densities in a subregion (Te1.1) of the primary auditory cortex is higher [Morosan et al., 2001] and that the PT is larger in the left compared to the right hemisphere [Beaton, 1997; Binder et al., 1996; Dos Santos et al., 2006; Good et al., 2001; Hugdahl et al., 1998; Jancke et al., 1994; Jancke and Steinmetz, 1993]. This was originally seen as an indication that the PT is involved in the auditory processing of speech. Recent studies have modified this view by claiming a more general function in analyzing complex sound structures, like rapidly changing cues, which are essential for differentiating stop consonants or place of articulation. On the other hand, PT seems not to be exclusively sensitive to those phonetic signals [Griffiths and Warren, 2002; Jancke et al., 2002; Krumbholz et al., 2005]. It therefore remains unclear whether these functional and structural asymmetries in the posterior third of the temporal lobes are necessary for speech perception or for merely amplifying and/or facilitating the processing of speech in the left temporal lobe.

By contrast, the left STS, especially the middle region of the STS (mid-STS) has repeatedly been reported in studies focusing on phonological and prelexical processing [Jancke et al., 2002; Price et al., 2005; Rimol et al., 2006b; Scott and Wise, 2004; Scott et al., 2000; Specht and Reul, 2003; Specht et al., 2005]. Conversely, the processing of music as well as prosodic information is assumed to be predominantly mediated through right temporal lobe structures [Grimshaw et al., 2003; Meyer et al., 2002; Samson, 2003; Tervaniemi, 2001; Tervaniemi and Hugdahl, 2003; Zatorre, 2001; Zatorre et al., 2002].

However, research on the functional asymmetry in auditory perception is often methodologically limited. Two

main problems should be mentioned in this respect. First, studies often employ fixed stimulus categories such as pure tones, synthetic sounds, sounds from musical instruments, or speech sounds. The contrasts between these stimulus categories may include several aspects occurring together, such as acoustic complexity along with differences between verbal and nonverbal processing. A second problem in studying functional asymmetry is the instruction of the participants or, respectively, the task they have to perform. A recent study by Dufor et al. [2007] demonstrated that the activation pattern changed when participants were instructed to focus on phonological aspects in sine-wave speech stimuli. Prior to this instruction, participants demonstrated an almost bilateral activation pattern which subsequently shifted to an activation pattern more common for speech perception tasks, when given the instruction. In other words, the expectancy of hearing speech-like sounds modulates the activation pattern. Studies which attempt to overcome the first problem often use speech comprehension tasks and may therefore be biased by expectancy and attentional mechanisms.

Therefore, this study used a stimulus paradigm that attempted to overcome these limitations to study the automatic, as opposed to attention-biased processing of speech, speech-like, and nonspeech stimuli. This was achieved by introducing stimuli which gradually changed from one category into another. This was paired with an instruction which focused the attention of the participants onto an irrelevant aspect. Thereby, this study was a continuation of the earlier introduced “sound morphing” technique [Specht et al., 2005] by generating stimuli which changed from white noise to either a speech or a music sound (see Fig. 1). Such a stepwise approach is advantageous in that the changes from one step to the next are minimal, while the whole set of stimuli covers a much broader range. Using white noise as the same starting point the analysis can be based on the differential evolution of brain responses while the sound “morphs” into either a speech or music sound. On the basis of the abovementioned differential functional asymmetry in speech and music processing, we expected to observe differentially evolving hemispheric laterality between those stimuli which evolved into a speech sound as compared to those which became a musical instrument sound. It was hypothesized that speech sound would predominantly increase

activation in the left temporal lobe, particularly in the posterior and middle part of STG and STS [Hickok and Poeppel, 2007; Scott and Wise, 2004], as the speech specific information in the sounds increased, triggering phonological and prelexical processing. That is, we expected to see a more leftward activation pattern when the sound could be processed more as speech rather than nonspeech sound. By contrast, it was expected that the musical instrument stimuli would lead to more bilateral or even right lateralized activation in the right temporal lobe [Hugdahl et al., 1999; Specht and Reul, 2003; Tervaniemi and Hugdahl, 2003] as the musical instrument specific information increased.

## METHODS

### Participants

Participants were 15 male, right-handed, healthy adults (25–35 years of age) recruited from the student population at the University of Bergen. All participants were screened with an audiometer (250, 500, 1,000, 2,000, and 3,000 Hz). Exclusion criteria were a hearing threshold greater than 20 dB or interaural difference greater 10 dB on either frequency. Handedness was determined according to the Edinburgh Inventory [Annett, 1970]. The highest possible score in this rating for right handedness is 15 and the exclusion criterion was set to 13. All participants gave written informed consent in accordance with the Declaration of Helsinki and institutional guidelines.

### Stimuli

For this study, we selected two speech and two music sounds. Because variations in the voice onset time (VOT) of different speech sounds can give differential lateralization effects [see e.g. Rimol et al., 2006a; Sandmann et al., 2007], we chose the consonant-vowels (CV) /da/ with a short VOT and /ta/ with a long VOT to control for this. The CVs were read by a male voice and lasted 420 ms. For a nonspeech control condition that could easily be perceived within the same time-window and have a richer spectro-temporal characteristic than a pure tone, we used two different musical instrument sounds; a guitar sound ( $A_3$ ) and a piano chord (C major triad on a  $C_3$  root). After recording and digitizing, all natural stimuli were edited so that they matched in duration and mean intensity (Gold-wave Software). Additionally, white noise was generated and matched in duration and mean intensity to the natural stimuli.

### Parametric Manipulation

To have a gradual transition from a nonspeech sound into a speech sound, we constructed seven different steps where we parametrically varied the level of white noise

perturbation in the sounds. This was accomplished using a morphing procedure whereby the sounds (speech and music sounds) were mixed with white noise, using increasingly larger interpolation factors (SoundHack; www.soundhack.com). This resulted in a continuum from white noise towards a more speech or musical instrument like sound, from which we selected in seven total distinct steps. Thereby, the manipulation procedure gradually revealed the specific spectral and temporal characteristic of the speech/music instrument sounds in a stepwise manner.

The speech analysis program Praat (www.praat.org) was subsequently used to obtain an account of the phonetic structures present at the different manipulation steps for the speech sounds. This analysis revealed that the different phonetic information in the sounds increased over the seven manipulation steps. Analyzing formant structures in the different sound stimuli showed an energy peak just below 1,000 Hz that is in the middle of the F1 and F2 of the original/da/sound at step 3 (see Fig. 1). A clearer differentiation between F1 and F2, although still somewhat perturbed, was present in step 5 followed by an even clearer differentiation between F3 and F4 in step 6.

### Scanning Procedures

The fMRI study was performed on a 3T GE Signa Exite scanner. The experiment was performed as a single session event-related stochastic design [Friston et al., 1999] which included 224 regular events, 90 null-events (e.g., trials with no stimuli), and 28 target trials. One event lasted 2 s and consisted of four repetitions of the same sound. The order of the events was pseudorandomized across manipulation steps and categories (speech/music) so that the morphing sequence was never presented in a consecutive order. There were 16 repetitions for each manipulation step and each category. The design differentiated only speech and music but not between the underlying stimuli themselves (e.g., between /da/ and /ta/ or between the piano and guitar sound). To exclude activation related to top-down processes [see Dehaene-Lambertz et al., 2005; Dufor et al., 2007; Sabri et al., 2008] and to keep attention relatively constant during fMRI data acquisition [see Jancke et al., 1999], participants were given an arbitrary task that was not related to the specific quality of different sound categories. They were asked to report only when they heard a stimulus in one ear. These target trials were randomly distributed and there were an equal number of trials for the left and right ear, respectively.

To present the stimuli without scanner noise in the background, a sparse sampling technique was used with 1.5 s of image acquisition and an additional silent gap of 2.3 s. Axial slices for the functional imaging were positioned parallel to the AC-PC line with reference to a high resolution anatomical image of the entire brain volume and obtained using a  $T_1$ -weighted gradient echo pulse

sequence (MPRAGE). The functional images were acquired using an EPI sequence with 370 EPI volumes, each containing 25 axial slices ( $64 \times 64$  matrix,  $3 \text{ mm} \times 3 \text{ mm} \times 5.5 \text{ mm}$  voxel size, TE 30 ms) that covered the cerebrum and most of cerebellum. The stimuli were presented through MR compatible headphones with insulating materials that also compensated for the ambient scanner noise by 24 dB (NordicNeuroLab, [www.nordicneurolab.no](http://www.nordicneurolab.no)). Presentation of the stimuli and recording of the behavioral responses was controlled by the E-prime software (Psychology Software Tools Inc.) running on a PC placed outside of the MR chamber. The intensity of the stimuli was constant across manipulation steps and set to 87 dB (LAeq), as measured by a Brüel and Kjær Measuring Amplifier (Type 2250 connected to a Brüel and Kjær head and torso-simulator 4128C).

### Data Analysis

The BOLD-fMRI data were preprocessed and statistically analyzed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm>). The EPI images were first realigned to adjust for head movements during the image acquisition and the images were corrected for movement-induced distortions (“unwarping”). Data were subsequently inspected for residual movement artefacts. The realigned image series were then normalized to the stereotaxic Montreal Neurological Institute (MNI) reference space provided by the SPM5 software package, and resampled with a voxel-size of  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ . The images were finally smoothed by using a Gaussian kernel of 8-mm. Single subject statistical analysis was performed using a fixed-effects statistical model with the hemodynamic response function (HRF) and its time derivative as basic functions. A design matrix was specified using “the general linear model” for single subject analysis that specified individual vector of onsets for every level of manipulation in both categories of stimuli and left/right targets, in all 16 conditions. A single contrast was specified for each condition. The group data were analyzed with a  $2 \times 7$  ANOVA model (2 stimulus categories, 7 manipulations). Main effects and interactions were explored with *F*-contrasts by applying an FWE-corrected significance threshold of  $P < 0.05$ , and a cluster-extension threshold of at least five voxels.

To explore the parametric variation in more detail, we also specified linear contrasts, highlighting those areas with increased activations through the seven manipulation steps. Therefore, we specified a linear contrast on the single-subject level and grouped these individual results in one-sample *t*-tests (one for music and one for speech). These analyses were also analyzed with an FWE corrected threshold of  $P < 0.05$  and at least five voxels per cluster.

For anatomical localization, a nonlinear transformation from MNI to Talairach coordinates [Talairach and Tournoux, 1988] was performed ([www.mrc-cbu.cam.ac.uk](http://www.mrc-cbu.cam.ac.uk)) and anatomically characterized by using the Talairach Daemon

software (<http://ric.uthscsa.edu/resources>) and the Talairach atlas [Talairach and Tournoux, 1988]. Cross validation was performed using overlays on anatomical reference images from the Brodmann and AAL (automatic anatomic labeling) maps as part of the MriCro software ([www.mricro.com](http://www.mricro.com)).

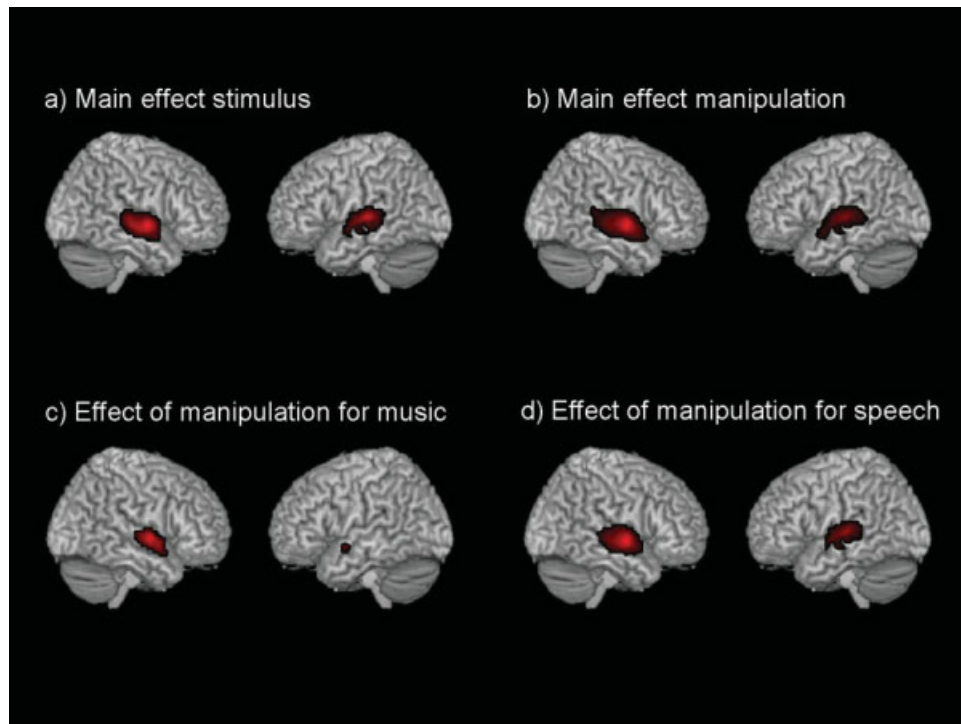
As a post-hoc analysis, a region of interest (ROI) analyses was performed focusing on the middle part of the superior temporal sulcus (mid-STS) from which the BOLD signal was extracted. The specification of the ROIs followed a comparable procedure as described by Liebenthal et al. [2005]. First, the ROIs were anatomically specified using the MARINA software [Walter et al., 2003] for deriving the STG and MTG (corresponding to BA 21 and 22). They were then postprocessed using the MriCro software in order to restrict the ROIs to mid-STS and to exclude primary auditory areas (BA 41/42). The ROIs were separately specified for the left and right hemisphere to account for the different localization on the two hemispheres. Instead of a weighted mean, as used by Liebenthal et al. [2005], we used the overall averaged BOLD signal from each ROI and for each manipulation, stimulus, and participant. These data were then subjected to a  $2 \times 2 \times 7$  ANOVA, with hemisphere (left/right), stimulus (music/speech), and manipulation (1st–7th manipulation) as factors. Standard *F*-tests for exploring main effects and interactions were applied and effects were considered to be significant when a sphericity corrected significance threshold of  $P < 0.05$  was reached (Greenhouse Geisser). The obtained results were followed up with a post-hoc  $2 \times 7$  ANOVA on the laterality index ( $LI = (\text{left} - \text{right})/(\text{left} + \text{right})$ ). To explore different levels of lateralization in more detail, the data were subjected to pair-wise post-hoc comparisons between left and right, that is 14 comparisons (7 for the music sounds and 7 for the speech). We then applied a Bonferroni correction by accepting only those comparisons as significant, which fulfilled  $P < 0.0036$  ( $= 0.05/14$ ).

### RESULTS

Significant main effects of the factor stimulus (that is the two different stimulus categories—speech and music sounds) as well as manipulation (that is the seven manipulations within each category) were found bilaterally in the posterior part of the temporal lobe, comprising the STG and STS. Exploring the effects of manipulation for the speech and music stimuli separately, the analyses revealed that the left STG and STS responded almost solely to the manipulation of the speech sounds, while the right STG and STS area responded to the manipulation of both speech and music sounds (see Fig. 2 and Table I).

In addition, a significant interaction of stimulus  $\times$  manipulation was found in the middle region of the left STS, showing greater activation by speech compared to music (see Fig. 3, red area, and Table I). This finding was further supported by the parametric analysis showing





**Figure 2.**

Analysis of the  $2 \times 7$  ANOVA: Displayed are the (a) main effects of stimulus, (b) Main effects of manipulation, (c) main effects of the manipulation only for music, (d) main effects of the manipulation only for speech. All results are  $F$ -contrasts, and a FWE-corrected threshold of  $P < 0.05$  with at least five voxel per cluster was applied.

significantly increased activation throughout the seven manipulation steps. For the music sounds, only the right STS showed a parametric modulation whereas the speech sounds generated a modulation in the STS of both hemi-

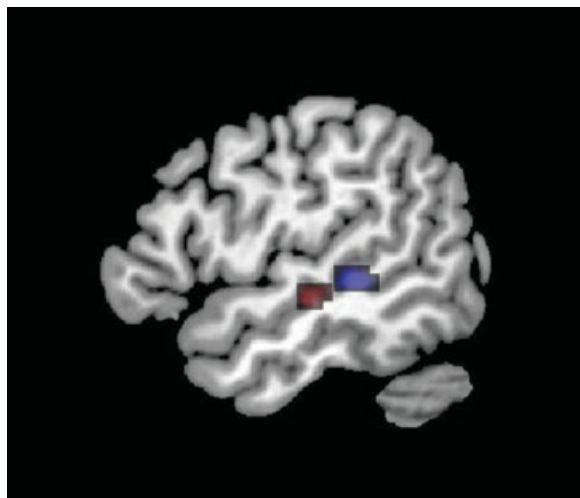
spheres (see Fig. 3, blue area, and Table II). This differential behavior became insignificant when directly comparing between the two stimulus types using a corrected threshold. However, based on our a priori

**TABLE I. Anatomical location of the main effects and interaction, given with coordinates,  $F$ -value and cluster size (voxel size  $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ )**

Cluster size	$P$ (FWE)	$F$ -value	$x$	$y$	$z$	Localization
Main effect of stimulus						
371	0.000	102.03	-57	-18	-3	Left STG/STS
331	0.000	99.55	60	-6	-3	Right STG/STS
Main effect of manipulation						
572	0.000	39.65	60	-6	-3	Right STG/STS
398	0.000	23.84	-57	-18	-3	Left STG/STS
Interaction: stimulus $\times$ manipulation						
9	0.001	8.5	-54	-18	-6	Left STS
Effect of manipulation for music						
167	0.000	16.79	57	-6	0	Right STG/STS
11	0.003	8.05	-51	-3	-9	Left STG/STS
Effect of manipulation for speech						
418	0.000	33.22	60	-6	-3	Right STG/STS
334	0.000	27.56	-57	-18	-3	Left STG/STS

STG, superior temporal gyrus; STS, superior temporal sulcus.

All results are  $F$ -contrasts, and a FWE-corrected threshold of  $P < 0.05$  with at least five voxel per cluster was applied.



**Figure 3.**

Red area: speech  $\times$  manipulation interaction (F-contrast,  $2 \times 7$  ANOVA); blue area: linear response to the manipulation of the speech sounds (linear  $t$ -contrast). All results are explored with an FWE-corrected threshold of  $P < 0.05$  with at least five voxel per cluster.

hypotheses of differential activation in the STG/STS area, we applied an exploratory post-hoc analysis with an uncorrected voxel-threshold of  $P < 0.001$  and a corrected extend threshold of 68 voxels ( $P < 0.05$ , corrected threshold for the extension of a cluster). The obtained results confirmed our hypothesis by demonstrating a significant difference in the parametric modulation in the left posterior and middle STS. In addition, we applied a hypothesis driven small-volume correction based upon the Gaussian random field theory [Kiebel et al., 1999]. The effect became significant, on a FWE-corrected voxel-threshold level, using the entire left STG/STS/MTG as the volume for the correction.

These fMRI results were further evaluated by an ROI analysis, where the BOLD signal was extracted from mid-STS. The results demonstrated a diverging time course between the left and right STS, as the sound became more like speech. The results from the  $2 \times 2 \times 7$  ANOVA, with hemisphere as an additional factor, showed significant main effects and interactions of all factors ( $P < 0.05$ ). More importantly, the threefold interaction of stimulus by manipulation by hemisphere was significant. As a post-hoc test, we performed a  $2 \times 7$  ANOVA on the laterality indices, where the main effects and interaction became significant. Accordingly, the paired  $t$ -tests performed by the post-hoc analysis revealed significant differences between the left and right mid-STS ( $P < 0.0036$ ,  $df = 14$ ) for the 5th ( $t = 4.968$ ), 6th ( $t = 4.086$ ), and 7th step ( $t = 6.878$ ) for the speech stimuli. This divergent time course was not present for the musical stimuli. Here, the signal developed synchronously in the left and right STS. Interestingly, the results for the speech stimuli demonstrated a ceiling effect, with no further increase in activation after the fifth step in the left STS and a slight decrease in the right STS (Fig. 4).

## DISCUSSION

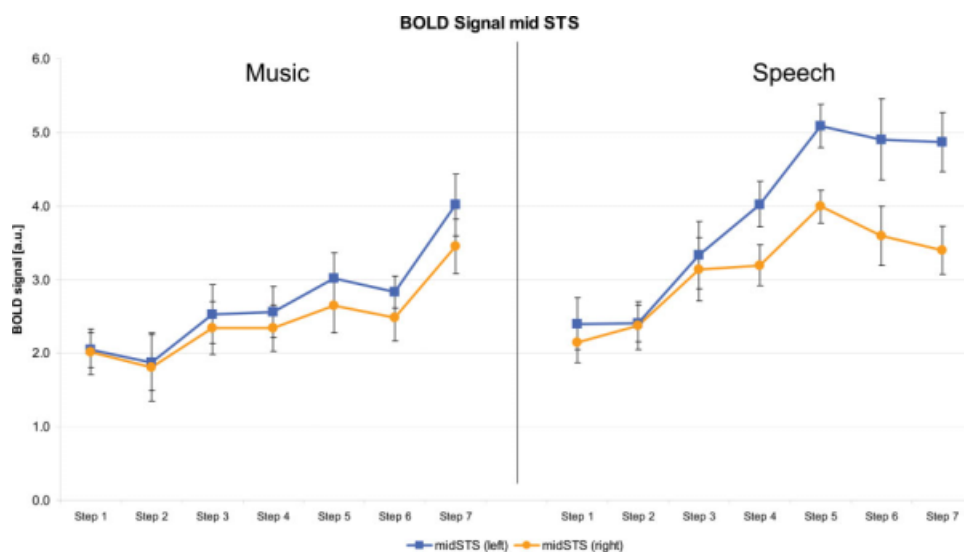
By “morphing” the sound spectrum from white noise into either a speech sound or a musical instrument sound, we have shown that neuronal responses in the temporal lobes differ between the left and right hemisphere as the acoustic spectrum becomes more similar to the speech or music-instrument spectrum. The main activations were found in the posterior and middle part of the STG and STS, with a significantly more lateralized response to the speech compared to music sound manipulations. These findings were revealed by the imaging data as well as the ROI analysis. Moreover, modulated activity was seen in the right STG and STS regardless of whether it was speech, speech-like, musical-instrument, or musical-instrument-like sound. By contrast, the same areas of the left temporal lobe were more sensitive to the manipulation of

**TABLE II. Anatomical location of the parametric responses, given with coordinates,  $t$ -value and cluster size (voxel size 3 mm  $\times$  3 mm  $\times$  3 mm)**

Cluster size	$P$ (FWE)	$F$ -value	$x$	$y$	$z$	Localization
Parametric response to music						
23	0.002	10.07	60	-6	0	Right STG
Parametric response to speech						
33	0.002	10.43	-57	-36	0	Left STS
10	0.005	9.53	60	-24	-9	Right STS
18	0.007	9.17	60	-6	0	Right STG
Differential parametric response (speech > music)						
80	0.222	6.52	-60	-36	12	Left STS
[0.011 (SVC: entire STG/STS/MTG)]						

STG, superior temporal gyrus; STS, superior temporal sulcus; MTG, middle temporal gyrus.

All results are  $t$ -contrasts, and a FWE-corrected threshold of  $P < 0.05$  with at least five voxel per cluster was applied.



**Figure 4.**

BOLD signal (in arbitrary units [a.u.]), extracted from the left and right mid-STS; error bars are denoting the standard error. Note that the left-right differences for the speech sounds are significant ( $P < 0.0036$ ,  $df = 14$ ) for the 5th, 6th, and 7th step.

speech rather than music sounds, which was in accordance with our hypothesis. In particular, we observed a significant stimulus by manipulation interaction only in the middle region of the left superior temporal sulcus (mid-STS), reflecting higher sensitivity to speech compared to musical stimuli. This observation is in accordance with earlier reports [Indefrey and Cutler, 2004; Scott et al., 2000; Specht et al., 2005; Specht and Reul, 2003]. This was further supported by the fact that only the parametric modulation of the speech sounds became significant on the left side, in the region of the left mid-STS. Interestingly, the area within the mid-STS was found to be posterior and adjacent to the area that was significant in the stimulus  $\times$  manipulation interaction. Finally, after adding hemisphere as a factor in a  $2 \times 2 \times 7$  ANOVA, the three-way interaction became significant, reflecting significantly different processing of the speech- and music-sound manipulations in the two hemispheres.

There are several important findings in this study. The most prominent being the differential response of the STG and STS to the acoustical parametric manipulation. It is unlikely that such a response would be observed by merely using “static” stimulus categories, which is a common procedure in most neuroimaging studies on auditory perception. Another finding is the difference in response sensitivity to the speech stimuli of the left compared to the right temporal lobe, as indicated by the significant left-right differences (see Fig. 4). Such a differentiation was not observed in the whole brain analysis or in the ROI analysis during the manipulation of the music sounds. There is an ongoing discussion with respect to the differential specialization of the left and right audi-

tory cortex. There is reasonable evidence that the left auditory cortex is more specialized in processing temporal information while the right auditory cortex is more specialized in processing spectral information [Zatorre and Belin, 2001]. There is also a discussion concerning a possible asymmetry in temporal sensitivity between the auditory cortices (Asymmetric sampling theory, AST). Although the left side is assumed to analyze auditory signals on a timescale of 25–50 ms, the right side is assumed to have a longer integration window of 200–300 ms [Boemio et al., 2005]. Such a differentiation at a lower level of auditory cortical processing might also influence further processing along the proposed processing streams in the temporal lobes. One important aspect of our findings in this circumstance is that the detected areas of the right temporal lobe demonstrated responses to all auditory stimuli, irrespective of the acoustic properties, that is they were sensitive to the manipulation but not specific to the type of stimulation. By contrast, the response of the left temporal lobe, and here in particular the mid-STS, was predominantly modulated by the manipulation of speech and not by musical sounds. These results support the notion of differing temporal resolutions and spectro-temporal sensitivities between the left and right auditory systems. The changes in the sound spectra were very subtle between two adjacent manipulation steps, leaving the overall magnitude and intensity almost unchanged. Therefore, the differences between two adjacent manipulation steps were outside of the assumed integration window of the right auditory cortex of 200–300 ms [Boemio et al., 2005]. This is also reflected by the flatter slope from the right mid-STS area (see Fig. 4), and could explain why

only the main effect became significant on the right side, and not the interaction (see Table I). However, because speech sounds by nature have a specific spectro-temporal characteristic, the respective areas of the left temporal lobe are probably more “tuned” to detect those phonetic structures also within a disturbed signal, such as incomplete or altered spectrum. Especially the middle part of the STS demonstrated the strongest response to phonetic structures, as indicated by the significant interaction (see Table I, Fig. 3). As the formant structures became more and more detectable, this area seemed also to be more sensitive to the subtle changes from one step to another. One could further speculate that the often observed right lateralization during the perception of nonverbal stimuli, like music, is perhaps caused by a relatively reduced response of the left temporal lobe rather than a higher sensitivity of the right temporal lobe to nonverbal material, as visible in the lower row of Figure 2. One might argue that the detected differences were only caused by the different complexity of the spectro-temporal characteristic of the sounds and not by the categorical differentiation, which is speech versus music. However, this would primarily have caused a differential response within the primary and secondary auditory cortex although not in mid-STS. In addition, the analysis mainly focused on the evolution of the brain response throughout the morphing sequence. Therefore, differential responses resulting exclusively from physical differences would have occurred only at the very end of the morphing sequence, because the starting point was the same for both stimulus categories. In this case, the interaction should also have been significant for the primary auditory cortex and not only in mid-STS. Furthermore, one aim of the study was to use natural stimuli, which also could be identified by the subjects. The usage of artificial sounds, which match the complexity but do not sound natural, may have resulted in additional activations. This, however, is a basic drawback in all studies using a natural stimulation approach.

Besides this more general effect between speech and music sounds, the second relevant aspect of the results is the differentiation between a linear response to the manipulation and a more step-wise processing, when focusing only on the sequence of speech sounds. There is a qualitative difference in the time course of the mid-STS area before and after the fifth manipulation step. Although the signal increased between the second and fifth step, it remained constant (left mid-STS) or even slightly decreased (right mid-STS) after this step, resulting in the strongest and most significant left-right difference in the last step (see Fig. 4). Interestingly, the area showing the most linear response with respect to the manipulation is more posterior to the areas showing the most significant interaction effect (see Fig. 3). One could interpret this anatomical differentiation through the model of a posterior-anterior processing stream [Scott et al., 2000; Specht and Reul, 2003], where the sensitivity to speech specific signals increases from posterior to anterior, parallel with more

phonological and pre-lexical processing of the signal. Furthermore, the mid-STS area demonstrated a constant activation after the 5th manipulation step (see Fig. 4), which underlines further the importance of the first two formants for speech perception, as they were present from step 5 onwards (see Fig. 1). In this circumstance, one could further infer that the mid-STS act as a filter or gate that analyses every incoming signal for speech elements. In other words, as soon as a sound is identified as speech (or speech-like), the STS probably gates the signal to other language areas in the left temporal lobe for further analysis, analogous to prelexical processing. One could further speculate that the decrease of activation in the right mid-STS reflects an inhibitory process, facilitating speech processing on the left side by reducing sensitivity on the right side.

Finally, it is important to note that the discussed areas show only speech-specific responses to the selected manipulation of the stimuli. This does not mean that these areas are speech-specific areas, e.g., that they respond exclusively to either speech or musical sounds [Price et al., 2005]. As could be seen in Figure 2, several areas of both temporal lobes were involved in processing the sounds, but only some areas, discussed earlier, showed additional speech-specific responses.

In summary, this study demonstrated differential processing of spectral dimensional properties between speech and music-instrument sounds. Using this new approach of a gradual manipulation of the acoustic properties of speech and music sounds, we were able to detect differential and category-specific responses in the temporal lobe, especially in the mid-STS. Furthermore, these data demonstrated that the right mid-STS area responded to any type of manipulation while the response of left mid-STS was predominantly modulated by the manipulation of the speech sounds. One could therefore conclude that the STG and STS areas of the right temporal lobe are involved in speech perception although, on the other hand, this involvement was on a more general level, because the measured responses were not different between the two selected categories (speech/music). By contrast, the same substructures of the left temporal lobe demonstrated a high level of sensitivity to phonological signals. The present findings provide further evidence that, although speech perception is predominantly a bilateral process, regions of the left temporal lobe, in particular the mid-STS area, are more responsive in the presence of speech signals.

## ACKNOWLEDGMENTS

We are grateful to Morten Eide Pedersen and Ruben Sverre Gjertsen for technical advice and help when constructing the new stimulus paradigm. We would further thank the radiographers at the Haukeland University Hospital for their assistance and technical expertise.



## REFERENCES

- Annett M (1970): A classification of hand preference by association analysis. *Br J Psychol* 61:303–321.
- Beaton AA (1997): The relation of planum temporale asymmetry and morphology of the corpus callosum to handedness, gender, and dyslexia: A review of the evidence. *Brain Lang* 60:255–322.
- Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW (1996): Function of the left planum temporale in auditory and linguistic processing. *Brain* 119 (Part 4):1239–1247.
- Boemio A, Fromm S, Braun A, Poeppel D (2005): Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S (2005): Neural correlates of switching from auditory to speech perception. *Neuroimage* 24:21–33.
- Dos Santos SS, Woerner W, Walter C, Kreuder F, Lueken U, Westerahausen R, Wittling RA, Schweiger E, Wittling W (2006): Handedness, dichotic-listening ear advantage, and gender effects on planum temporale asymmetry—A volumetric investigation using structural magnetic resonance imaging. *Neuropsychologia* 44:622–636.
- Dufor O, Serniclaes W, Sprenger-Charolles L, Demonet JF (2007): Top-down processes during auditory phoneme categorization in dyslexia: A PET study. *Neuroimage* 34:1692–1707.
- Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM (1999): Stochastic designs in event-related fMRI. *Neuroimage* 10:607–619.
- Good CD, Johnsrude I, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001): Cerebral asymmetry and the effects of sex and handedness on brain structure: A voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* 14:685–700.
- Griffiths TD, Warren JD (2002): The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Grimshaw GM, Kwasny KM, Covell E, Johnson RA (2003): The dynamic nature of language lateralization: Effects of lexical and prosodic factors. *Neuropsychologia* 41:1008–1019.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hugdahl K, Heiervang E, Nordby H, Smievoll AI, Steinmetz H, Stevenson J, Lund A (1998): Central auditory processing, MRI morphometry and brain laterality: Applications to dyslexia. *Scand Audiol Suppl* 49:26–34.
- Hugdahl K, Bronnick K, Kyllingsbaek S, Law I, Gade A, Paulson OB (1999): Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: A 15O-PET study. *Neuropsychologia* 37:431–440.
- Indefrey P, Cutler A (2004): Prelexical and lexical processing in listening. In: Gazzaniga MS, editor. *The Cognitive Neuroscience*, 3rd ed. Cambridge, MA: MIT Press. pp 759–774.
- Jancke L, Steinmetz H (1993): Auditory lateralization and planum temporale asymmetry. *Neuroreport* 5:169–172.
- Jancke L, Schlaug G, Huang Y, Steinmetz H (1994): Asymmetry of the planum parietale. *Neuroreport* 5:1161–1163.
- Jancke L, Mirzazade S, Shah NJ (1999): Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neurosci Lett* 266:125–128.
- Jancke L, Wustenberg T, Scheich H, Heinze HJ (2002): Phonetic perception and the temporal cortex. *Neuroimage* 15:733–746.
- Kiebel SJ, Poline JB, Friston KJ, Holmes AP, Worsley KJ (1999): Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *Neuroimage* 10:756–766.
- Krumbholz K, Schonwiesner M, Rubsamen R, Zilles K, Fink GR, von Cramon DY (2005): Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur J Neurosci* 21:230–238.
- Liebenthal E, Binder JR, Spitzer SM, Posing ET, Medler DA (2005): Neural substrates of phonemic perception. *Cereb Cortex* 15:1621–1631.
- Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002): fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17:73–88.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001): Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13:684–701.
- Price C, Thierry G, Griffiths T (2005): Speech-specific auditory processing: where is it? *Trends Cognit Sci* 9:271–276.
- Rimol LM, Eichele T, Hugdahl K (2006a): The effect of voice-onset-time on dichotic listening with consonant-vowel syllables. *Neuropsychologia* 44:191–196.
- Rimol LM, Specht K, Hugdahl K (2006b): Controlling for individual differences in fMRI brain activation to tones, syllables, and words. *Neuroimage* 30:554–562.
- Sabri M, Binder JR, Desai R, Medler DA, Leitl MD, Liebenthal E (2008): Attentional and linguistic interactions in speech perception. *Neuroimage* 39:1444–1456.
- Samson S (2003): Neuropsychological studies of musical timbre. *Ann N Y Acad Sci* 999:144–151.
- Sandmann P, Eichele T, Specht K, Jancke L, Rimol LM, Nordby H, Hugdahl K (2007): Hemispheric asymmetries in the processing of temporal acoustic cues in consonant-vowel syllables. *Restor Neurol Neurosci* 25:227–240.
- Scott SK, Wise RJ (2004): The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92:13–45.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Part 12):2400–2406.
- Specht K, Reul J (2003): Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: An auditory rapid event-related fMRI-task. *Neuroimage* 20:1944–1954.
- Specht K, Rimol LM, Reul J, Hugdahl K (2005): “Soundmorphing”: A new approach to studying speech perception in humans. *Neurosci Lett* 384:60–65.
- Talairach J, Tournoux P (1988): *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart, New York: Thieme-Verlag.
- Tervaniemi M (2001): Musical sound processing in the human brain. Evidence from electric and magnetic recordings. *Ann N Y Acad Sci* 930:259–272.
- Tervaniemi M, Hugdahl K (2003): Lateralization of auditory-cortex functions. *Brain Res Brain Res Rev* 43:231–246.
- Walter B, Blecker C, Kirsch P, Sammer G, Schienle A, Stark R, Vaitl D (2003): MARINA: An easy to use tool for the creation of MAsks for Region of Interest Analyses [abstract]. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, New York, NY. Available on CD-Rom in *NeuroImage*, Vol. 19.
- Zatorre RJ (2001): Neural specializations for tonal processing. *Ann N Y Acad Sci* 930:193–210.
- Zatorre RJ, Belin P (2001): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. *Trends Cognit Sci* 6:37–46.