NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 30 (2006) 563 – 569

Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus

Riikka Möttönen,[a,b,*] Gemma A. Calvert,[c] Iiro P. Jääskeläinen,[a,b,d] Paul M. Matthews,[e] Thomas Thesen,[c] Jyrki Tuomainen,[a,f] and Mikko Sams[a,b]

[a] *Laboratory of Computational Engineering, Helsinki University of Technology, Finland*
[b] *Advanced Magnetic Imaging Centre, Helsinki University of Technology, Finland*
[c] *Department of Psychology, University of Bath, UK*
[d] *MGH-MIT-HMS A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129, USA*
[e] *Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Oxford University, UK*
[f] *Department of Phonetics, University of Turku, Finland*

Received 13 May 2005; revised 3 October 2005; accepted 4 October 2005
Available online 4 November 2005

The left superior temporal cortex shows greater responsiveness to speech than to non-speech sounds according to previous neuroimaging studies, suggesting that this brain region has a special role in speech processing. However, since speech sounds differ acoustically from the non-speech sounds, it is possible that this region is not involved in speech perception per se, but rather in processing of some complex acoustic features. "Sine wave speech" (SWS) provides a tool to study neural speech specificity using identical acoustic stimuli, which can be perceived either as speech or non-speech, depending on previous experience of the stimuli. We scanned 21 subjects using 3T functional MRI in two sessions, both including SWS and control stimuli. In the pre-training session, all subjects perceived the SWS stimuli as non-speech. In the post-training session, the identical stimuli were perceived as speech by 16 subjects. In these subjects, SWS stimuli elicited significantly stronger activity within the left posterior superior temporal sulcus (STSp) in the post- vs. pre-training session. In contrast, activity in this region was not enhanced after training in 5 subjects who did not perceive SWS stimuli as speech. Moreover, the control stimuli, which were always perceived as non-speech, elicited similar activity in this region in both sessions. Altogether, the present findings suggest that activation of the neural speech representations in the left STSp might be a pre-requisite for hearing sounds as speech.

Keywords: fMRI; Sine wave speech; Speech; Superior temporal sulcus

Introduction

It has not been decisively determined to date whether there are speech-specific neural mechanisms in the human brain or whether speech sounds are processed by the same acoustic signal analysis mechanisms as other equally complex sounds. Theories of speech perception fall roughly into two categories (see, e.g., Diehl et al., 2004): (1) Those that assume that speech sounds are mapped into speech-specific (e.g., articulatory–gestural) representations in the human brain (Liberman and Mattingly, 1985; Liberman et al., 1967), thus making processing of speech sounds different from that of non-speech sounds. (2) Those that posit that general auditory mechanisms are responsible for processing both speech and non-speech sounds, i.e., that speech-specific mechanism do not exist (e.g., Stevens and Klatt, 1974; Miller et al., 1976; Pisoni, 1977).

Previous neuroimaging studies have attempted to address this controversy by comparing responses to speech vs. non-speech sounds (e.g., Demonet et al., 1992; Zatorre et al., 1992; Mummery et al., 1999; Binder et al., 2000; Scott et al., 2000; Vouloumanos et al., 2001; Narain et al., 2003). These studies have consistently demonstrated that hemodynamic responses are greater for speech than for non-speech sounds in the posterior parts of the left superior temporal gyrus/sulcus (STG/STS), suggesting that this region, classically considered as Wernicke's area, includes neuronal systems specialized in the analysis of speech sounds. Furthermore, there is evidence that also the anterior parts of the superior temporal cortex would be involved in the analysis of speech (e.g., Binder et al., 2000; Scott et al., 2000; Narain et al., 2003; Binder et al., 2004; Liebenthal et al., 2005; Obleser et al., in press).

Comparison of brain activations for speech and non-speech sounds that are *acoustically* different is, however, marred by a fundamental problem: it cannot be ruled out that any observed

* Corresponding author. Laboratory of Computational Engineering, Helsinki University of Technology, PO Box 3000, FIN-02015 HUT, Finland. Fax: +358 9 451 6172.
E-mail address: riikka.mottonen@hut.fi (R. Möttönen).

Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.10.002

differences in response are due to differences in the acoustic features of the stimuli. It is possible that the left posterior STG/STS is not involved in speech perception per se but rather in processing of complex acoustic features that are characteristic of speech sounds. Consistent with this interpretation, there is evidence that the left STG/STS is specialized for processing the rapid time-varying acoustic features, such as formant transitions that are typical of consonant sounds (Zatorre et al., 2002; Joanisse and Gati, 2003).

Sine wave speech (SWS) stimuli are modified speech sounds that typically consist of three sine waves tracking the lowest formants of speech sounds (Remez et al., 1981). SWS stimuli are heard as non-speech when perceivers do not know that sounds are modified speech sounds. However, as soon as the perceivers are informed about the origin of the stimuli, they normally start hear the phonetic content of the stimuli. Thus, identical SWS stimuli can be perceived either as non-speech or speech, depending on the perceiver's prior experience of the stimuli (Remez et al., 1981). Consequently, using SWS stimuli, it is possible to study the neural basis of speech perception without the problems associated with differing acoustic stimulus features between experimental conditions.

Here, we used SWS stimuli in a functional magnetic resonance imaging (fMRI) study to find out whether the left superior temporal cortex contains the neuronal substrate for speech perception. We hypothesized that truly *speech-specific* regions should be more active when subjects perceive the SWS stimuli as speech (i.e., as pseudowords) compared with the activity when the *identical* SWS stimuli are perceived as non-speech.

Methods

Subjects

Twenty-one right-handed native English speakers (aged between 18 and 36 years, 9 females) with normal (self-reported) hearing and vision participated in the study after providing written informed consent. The study protocol was approved by the local research ethics committee and adhered to the guidelines of the Helsinki declaration.

Stimuli

The natural speech tokens /omso/ and /onso/ were recorded in a sound-attenuated booth using a condenser microphone. The audio channel was transferred to a computer (digitized 22,050 Hz, 16 bit resolution) and SWS replicas of both /omso/ and /onso/ were created using Praat software (Boersma and Weenink, 1992–2002) with a script provided by Chris Darwin (http://www.biols.susx.ac.uk/home/Chris_Darwin/Praatscripts/SWS). The script created three-tone stimuli by positioning time-varying sine waves at the center frequencies of the three lower formants of the natural speech tokens. The same stimuli have been used previously in the study of Tuomainen et al. (2005).

A noise-vocoded, spectrally rotated sound was also created as a control stimulus using Praat by modulating the /onso/ SWS stimulus (Blesser, 1972; Shannon et al., 1995). Similar stimuli have been previously used by Scott et al. (2000) and Narain et al. (2003) in their neuroimaging studies. This stimulus neither sounded like speech, nor could it be perceived as such even after speech training, in contrast to the SWS stimuli. Consequently, it operated as an ideal control stimulus to examine possible generic session effects on the hemodynamic responses. A control stimulus was crucial because it was necessary to present the experimental sessions in a fixed order, the non-speech SWS session always being the first (see below). Once listeners learn to hear the SWS stimuli as speech, they cannot subsequently perceive the items as non-speech. The fixed session order might yield to non-specific changes in brain activation due to, e.g., increased exposure to stimuli and fatigue, unrelated to the issue of speech specificity. Accordingly, the purpose of the control stimulus was to provide a measure of how brain activation changes over time.

The experiment also included incongruent audiovisual (AV) stimuli which were similar to those used in the study of Tuomainen et al. (2005). Here, we however focus on auditory speech perception and report responses only to acoustic stimuli. The stimuli were presented with Presentation® software.

Procedure

The experiment consisted of following stages:

1. Training to categorize acoustic stimuli outside the scanner. Subjects were taught to categorize the three acoustic stimuli (SWS /omso/, SWS /onso/, control sound) into three non-speech categories ("Sound 1", "Sound 2" and "Sound 3"). Subjects practiced the discrimination task for 5–10 min. Performance was tested when a subject indicated that s/he was able to discriminate the sounds. In a short behavioral test subjects pressed one of three buttons depending on which of the three sounds they heard.
2. Pre-speech training fMRI scanning session. Just prior to the first fMRI scan, subjects were instructed to lie still and classify the sounds into three categories ("Sound 1", "Sound 2" and "Sound 3"). Subjects indicated their decision by pressing one of three buttons. Half of the subjects responded using the right hand and the other half used the left hand. A randomized sequence including SWS, control, AV and baseline (silence) stimuli was presented during a 15-min sparse-sampled scan. One stimulus was presented during each silent period of 11 s between volume acquisitions (for details, see Fig. 1 and Data acquisition).
3. Speech training. After the first scan and before the start of the second, subjects were told that "Sounds 1 and 2" (SWS stimuli) were in fact modulated speech sounds and could be

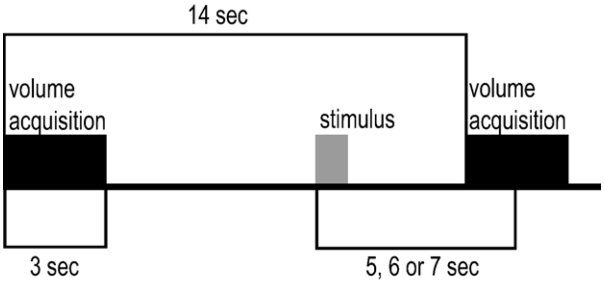

Fig. 1. Schematic representation of the experiment. Each volume acquisition of 3 s was followed by a silent period of 11 s (TR = 14 s). The stimuli were presented 5, 6 or 7 s prior to the mid-point of the volume acquisition.

heard as /omso/ and /onso/. A sound sequence, in which natural tokens of /omso/ and /onso/ stimuli were presented immediately prior to their SWS replicas, was then presented to subjects (15 repetitions of each pair). The subjects were asked to listen carefully to the presented sounds and to try to learn to perceive the "Sounds 1 and 2" (SWS stimuli) as /omso/ and /onso/. Subjects then performed a behavioral test during which they were asked to categorize SWS stimuli into /omso/ and /onso/. Subjects stayed in the scanner during this training period.

4. Post-speech training fMRI scanning session. Just prior to the second scan, subjects were instructed to press button "1" for /omso/, button "2" for /onso/ and "3" when they perceived the "Sound 3". The same randomized stimulus sequence as in the pre-training session was then presented during a 15-min post-training scan. However, a different randomized sequence was presented to different subjects.
5. Post-scan questionnaire. After the scanning, subjects filled in a questionnaire. The subjects were asked to describe how they had perceived "Sounds 1–3" during the pre- and post-training sessions.

Data acquisition

Functional imaging data were acquired on a 3.0-T MRI system with a multislice gradient-echo EPI sequence (TR = 3000 ms; TE = 28 ms, flip angle = 90°, FOV = 256 mm^2, matrix = 64 × 64) at the Oxford FMRIB Center. Twenty-four 5-mm-thick axial slices covering the whole brain were acquired during 3 s every 14 s over the 15-min scans.

The sparse-sampled sequence with silent periods of 11 s was used to minimize contamination caused by hemodynamic responses to scanner noise (Hall et al., 1999; 2000). The 3-s volume acquisition (mid-point) followed the onset of the acoustic stimulus by either 5, 6 or 7 s (see Fig. 1), at which latency the hemodynamic response was assumed to peak based on previous studies (Hickok et al., 1997; Hall et al., 2000). The lengths of the three different delays were balanced across stimulus types.

Each stimulus type was presented 15 times during each 15-min scan. The experiment consisted of two scans (one during pre- and one during post-training session). The decision to run two scans with 15 trials of each stimulus type in each scan was based on the results from a prior pilot study with 5 subjects in which four 15-min scans (two during pre-training and two during post-training) were acquired. These pilot experiments turned out to be too long, and subjects were seemingly tired during the last sessions. Furthermore, hemodynamic responses to all stimulus types decreased over the entire length of the experiment, and yet there was sufficient power to see typical patterns of auditory activation within a single 15-min scan. Therefore, we reduced number of scans to two in the final experiment and increased the number of subjects (>15) to ensure sufficient experimental power for the group analyses (Winer et al., 1991).

After the functional image acquisition a T1-weighted volume was acquired from each subject to aid co-registration (TR = 20 ms, TE = 5 ms, TI = 500 ms, flip angle = 15°, FOV = 256 × 192).

Data analysis

Data analysis was carried out using FEAT (fMRI Expert Analysis Tool) Version 5.1, part of FSL (fMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The following pre-statistics processing was applied: slice-timing correction using Fourier-space time-series phase-shifting, motion correction (Jenkinson et al., 2002), non-brain structure extraction (Smith, 2002), spatial smoothing using a Gaussian kernel of FWHM 5 mm, mean-based intensity normalization and high-pass temporal filtering. The first three volumes from each 15-min scan were omitted. Time-series statistical analyses were performed using a general linear model with local autocorrelation correction (Woolrich et al., 2001). The model used each type of stimulus as an independent explanatory variable. The model was not convolved to a hemodynamic response function, due to the sparseness of the data sampling. Subjects' functional images were registered to their anatomical images and to standard MNI (Montreal Neurological Institute) images (Jenkinson and Smith, 2001; Jenkinson et al., 2002). The MNI coordinates were transformed into Talairach coordinates (Talairach and Tournoux, 1988) by using a matlab script available on: http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml.

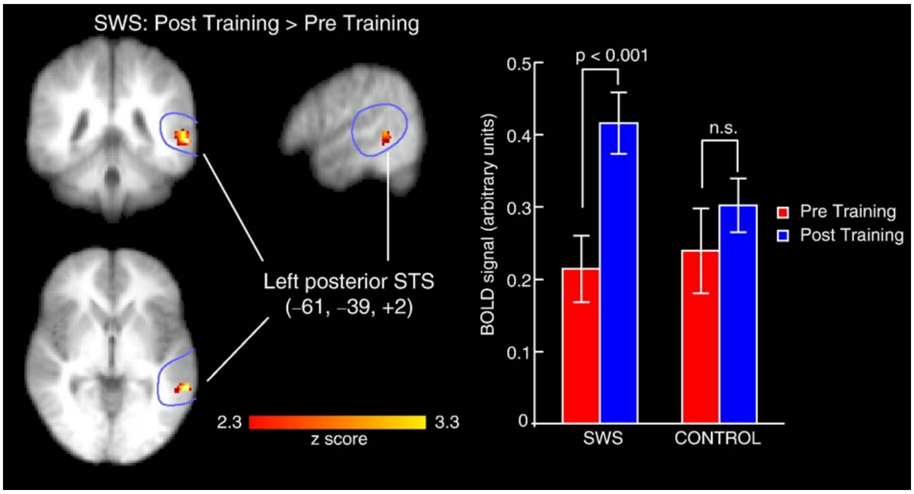

Fig. 2. Speech-specific activation in the left STSp. The left side of the figure shows the region, which was activated more in the post- than in the pre-training session for the SWS stimuli. The analysis was carried out within a left superior temporal ROI (indicated as blue). The right side of the figure depicts the mean (±SEM) BOLD signal changes in the left STSp for SWS and control stimuli in the pre- and post-training sessions (n = 16). The statistical significances are indicated.

To specifically test speech specificity within the left superior temporal cortex, we contrasted pre- and post-training activations for SWS in a mixed-effect analysis within a region of interest (ROI), encompassing areas involved in the classical Wernicke's area. The statistical parametric images were thresholded using clusters determined by $Z > 2.3$ and a cluster significance threshold of $P < 0.05$ (corrected) (Worsley et al., 1992; Friston et al., 1994; Forman et al., 1995). Data from those 16 subjects who reported having perceived SWS stimuli as speech during the post-training session were included in the analysis. Data from five subjects who failed to perceive the SWS stimuli as speech during the post-training session were excluded from all group-level analyses. The ROI for the left superior temporal cortex was obtained from the volumes of interest database (Nielsen and Hansen, 2002). The volumes in this database are established by probability density estimates of locations from the Brain Map database (Fox and Lancaster, 1994). This left superior temporal ROI encompassed the mid- and posterior parts of STS and STG, the medial part of Heschl's gyrus (HG), parieto-occipital junction and planum temporale (see Fig. 2). The ROI extended from the mid STG/STS (Talairach coordinate $y = -18$ mm) to the posterior STG/STS (Talairach coordinate $y = -58$ mm).

Since there is evidence that also more anterior parts of the superior temporal cortex participate in speech processing (see, e.g., Liebenthal et al., 2005; Obleser et al., in press), we contrasted pre- and post-training activations for SWS also within an ROI encompassing the lateral part of the HG and the superior temporal regions anterior to HG. This anterior ROI was of same size and shape as the more posterior ROI, but extended from the mid STS/STG (Talairach coordinate $y = -24$ mm) to the anterior STG/STS (Talairach coordinate $y = 8$ mm). Furthermore, in order to test whether speech-specific processing is lateralized, similar analysis was also carried out within ROIs in the right anterior and posterior superior temporal cortices.

To find out whether a region, which in ROI analyses showed differential activity to SWS stimuli in pre- and post-training sessions, is truly "speech-specific", we inspected individual blood–oxygen-level-dependent (BOLD) signals for each stimulus type within this region. We hypothesized that if the change of activity for SWS stimuli is due to the change in perception (and not due to the fixed order of conditions), activity to control stimuli should be similar in this region during both conditions, as they were always perceived as non-speech. The BOLD signal changes (in relation to rest) were obtained for each stimulus-type from the individual-level analyses for the pre- and post-training sessions. Two-way ANOVA was carried out to test whether training affected differently activity generated by SWS and control stimuli. Furthermore, effects of training on BOLD signals to SWS stimuli were explored in the group of subjects who did not perceive sounds as speech in the post-training session. It was expected that, in these subjects, training does not enhance responses to SWS stimuli in a speech-specific region.

Results

Behavioral results

16 of the 21 subjects reported in the questionnaire that they perceived SWS stimuli as non-speech during the pre-training session and as speech (i.e., /omso/ and /onso/) during the post-training session. Five subjects reported having perceived the SWS stimuli as non-speech during both sessions; their data were excluded from the group analyses. All subjects reported that they had perceived the control stimulus as non-speech in both sessions. They were reported to resemble, e.g., rasping, rattling and shushing sounds.

62 ± 6% of SWS stimuli were correctly categorized during pre-training session and 77 ± 4% during post-training session ($n = 16$). Control stimuli were categorized perfectly by all subjects in both sessions. Two-way ANOVA showed significant main effects of stimulus type (SWS vs. control, $F(1,15) = 66.97$, $P < 0.001$), and session (pre- vs. post-training, $F(1,15) = 5.81$, $P < 0.05$). Interaction between the session and the stimulus type was also significant ($F(1,15) = 5.81$, $P < 0.05$). The proportion of correctly categorized SWS stimuli was significantly greater in the post-training session than in the pre-training session ($t(15) = 2.41$, $P < 0.05$).

fMRI results

In order to test the hypothesis that the left superior temporal areas are speech-specific, we contrasted pre- and post-training activations for SWS within an anatomically defined ROI encompassing the mid- and posterior parts of the left STG/STS (see the left side of Fig. 2). In this analysis, SWS stimuli were found to elicit stronger activity during the post- than pre-training session in the left STSp (Talairach coordinates: $x = -61$ mm, $y = -39$ mm, $z = 2$ mm, cluster size: 117 voxels; see Fig. 2). None of the regions within the ROI showed decreased activity to SWS stimuli in the post-training session contrasted with the pre-training session. Neither were significant differences found between pre- and post-training activations for control stimuli.

No differences were found between pre- and post-training activations in the ROI encompassing the *anterior* parts of the left superior temporal cortex or within the ROIs encompassing the anterior and posterior parts of the *right* superior temporal cortex.

The right side of Fig. 2 shows the BOLD signal intensities in the left STSp to both SWS and control stimuli during the pre- and post-training sessions. Two-way ANOVA showed a significant main effect of session (pre- vs. post-training, $F(1,15) = 18.60$, $P <$

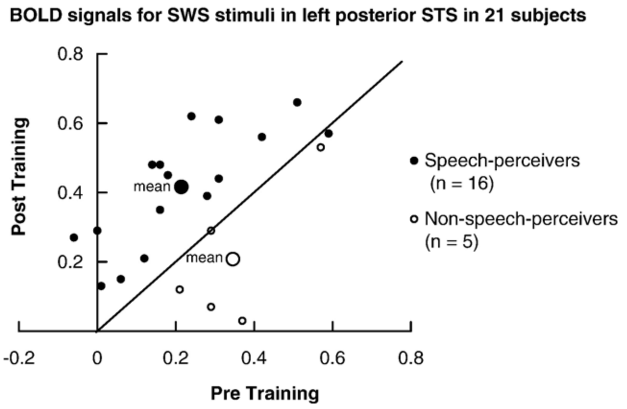

Fig. 3. BOLD signal changes for SWS stimuli in left STSp in all subjects ($n = 21$). X- and Y-axes show the signals during pre- and post-training sessions, respectively. 16 subjects perceived SWS stimuli as speech in the post-training session. 5 subjects perceived them as non-speech in both sessions.

0.001) but no effect of stimulus type (SWS vs. control). The interaction between stimulus type and session was significant ($F(1,15) = 11.96$, $P < 0.01$). Specifically, the BOLD signal within the left STSp increased significantly after speech-training for SWS ($t(15) = 6.92$, $P < 0.001$), but not for the control stimuli.

Fig. 3 depicts BOLD signals in the left STSp for SWS stimuli in post- and pre-training sessions in all 21 subjects. 16 subjects who reported that they had learned to perceive SWS stimuli as speech showed a consistent increase in the BOLD signal after training. No increase was observed in the five subjects who failed to learn to perceive the SWS stimuli as speech.

Discussion

We addressed a fundamental issue in speech perception research: are sounds perceived as speech processed by a specialized neuronal network or are they processed as any other equally complex sounds? We found that activity elicited by SWS was enhanced in the left STSp during the post-training session when subjects perceived the stimuli as speech compared to a pre-training session when the same stimuli were perceived as non-speech. Importantly, activity in the left STSp elicited by the control stimuli always perceived as non-speech was similar in the pre- and post-training sessions. Moreover, activity in the left STSp elicited by SWS was not enhanced after training in subjects who did not learn to perceive SWS as speech. These results provide compelling support for the proposition that specialized neuronal processing within the left posterior superior temporal cortex (i.e., in Wernicke's area) underlies speech perception. Since acoustic stimuli were *identical* in the non-speech and speech perception sessions, modulation of the activity in the left STSp cannot be explained by the acoustical complexity of the stimuli.

The present results confirm and extend the recent results of Dehaene-Lambertz et al. (2005) who also reported an enhancement of activity in the left posterior STG/STS to SWS syllables after speech training in French speaking subjects. In their study, a very similar finding was obtained using different stimuli, subjects with a different language background and a different stimulus presentation paradigm. This emphasizes the robustness of the effect. In addition, the present results show that the effect is found exclusive in the subjects who learned to perceive the stimuli as speech. Moreover, by using the control stimulus always perceived as non-speech, we excluded the possibility that the effect would be due to the fixed order of sessions.

Liebenthal et al. (2003) were the first to study neural basis of SWS processing. They found that activity in the left HG decreased when subjects were informed of the phonetic content of the SWS. No such decrease was observed in the current study, even though the left HG was within our left superior temporal ROIs. Furthermore, Liebenthal et al. (2003) did not observe any enhancements in activity after speech training. This could be due to the fact that only 13 out of 31 subjects perceived the SWS as speech after speech training, yet the data from all subjects were used in group-level statistical analyses.

In the present study, we failed to see enhanced activity in the anterior part of the STG/STS during speech perception, although there is evidence that the anterior "what" stream is involved in processing speech sounds (e.g., Rauschecker and Tian, 2000; Scott and Wise, 2004; Liebenthal et al., 2005; Obleser et al., in press). It is possible that speech processing in the anterior "what" stream is determined by the acoustic features of speech sounds, and therefore, the activity elicited by SWS stimuli did not change between pre- and post-training sessions, although perception changed from non-speech to speech. Moreover, it has been suggested that the anterior "what" stream would by specifically activated by intelligible speech (i.e., sentences, see, e.g., Scott et al., 2000; Narain et al., 2003; Giraud et al., 2004), whereas in the present study, the stimuli were unintelligible pseudowords. In sum, the present data are consistent with the idea that the anterior "what" stream is dedicated to acoustic- and/or meaning-based analysis of speech.

Why was the activity in the left STSp enhanced during speech perception? This region is likely to contain neural representations onto which acoustic input is specifically mapped when we listen to speech. For example, these representations could be articulatory–gestural. In the pre-training session, the SWS sounds were interpreted to be completely artificial, whereas in the post-training session, the same sounds were interpreted to originate from a talker's articulatory gestures, making, for example, vocal imitation of the sounds possible. It is thus plausible to assume that articulatory–gestural representations were activated in the post-training but not in the pre-training session. In line with this interpretation of the present findings, it has been proposed that the left STSp forms part of the posterior "how" stream, which operates in an articulatory–gestural domain and projects to the frontal speech production regions (Scott and Johnsrude, 2003; Scott and Wise, 2004; see also "dorsal stream" in Hickok and Poeppel, 2000; 2004). The left posterior superior temporal cortex (i.e., Wernicke's area) is densely connected with Broca's area, classically considered a motor speech production region, via the articuate fasciculus (e.g., Parker et al., 2005). There is growing evidence that the mirror neuron system, including Broca's area, provides neuronal substrate for embodied simulation of other person's gestures and plays an important role in interpersonal communication (for a review, see, Nishitani et al., 2005). However, there are also alternative views on the role of the left posterior superior temporal cortex in sound processing (see, e.g., Belin and Zatorre, 2000; Rauschecker and Tian, 2000; Romanski et al., 2000), and further experiments are naturally needed in order clarify speech processing in the posterior auditory stream(s).

The main finding of the present study is that the activation of the neural speech representations in the left STSp (i.e., in the putative "how" stream) is not determined by the acoustic features of the sound but is partly dependent on expectancy and experience of the observer. In the present study, observers' knowledge of the phonetic nature of SWS facilitated the activation of the neural speech representations, and accordingly, stimuli were heard as speech by most of the subjects. However, in some subjects, the same knowledge did not lead to this facilitation, and accordingly, the perception did not change to speech. Thus, activation of the neural speech representations in the left STSp could be a pre-requisite for hearing sounds as speech.

Acknowledgments

The study was supported by the Academy of Finland (RM, MS, IPJ), the Wellcome Trust, UK (GAC and TT) and the Medical Research Council, UK (PMM).

References

Belin, P., Zatorre, R.J., 2000. 'What', 'where' and 'how' in auditory cortex. Nat. Neurosci. 3, 965–966.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.

Blesser, B., 1972. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. J. Speech Hear. Res. 15, 5–41.

Boersma, P., Weenink, D., 1992–2002. Praat, a system for doing phonetics by computer v.4.0.13.

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.

Demonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., Frackowiak, R., 1992. The anatomy of phonological and semantic processing in normal subjects. Brain 115, 1753–1768.

Diehl, R.L., Lotto, A.J., Holt, L.L., 2004. Speech perception. Annu. Rev. Psychol. 55, 149–179.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.

Fox, P.T., Lancaster, J.L., 1994. Neuroscience on the Net. Science 266, 994–996.

Friston, K.J., Worsley, K.J., Frakowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1, 214–220.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb. Cortex 14, 247–255.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliot, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Hall, D.A., Summerfield, A.Q., Goncalves, M.S., Foster, J.R., Palmer, A.R., Bowtell, R.W., 2000. Time-course of the auditory BOLD response to scanner noise. Magn. Reson. Med. 43, 601–606.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99.

Hickok, G., Love, T., Swinney, D., Wong, E.C., Buxton, R.B., 1997. Functional MR imaging during auditory word perception: a single-trial presentation paradigm. Brain Lang. 58, 197–201.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. NeuroImage 19, 64–79.

Liberman, A., Mattingly, I.G., 1985. The motor theory of speech perception revised. Cognition 21, 1–36.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74, 431–461.

Liebenthal, E., Binder, J.R., Piorkowski, R.L., Remez, R.E., 2003. Short-term reorganization of auditory analysis induced by phonetic experience. J. Cogn. Neurosci. 15, 549–558.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.

Miller, J.D., Wier, C.C., Pastore, R.E., Kelly, W.J., Dooling, R.J., 1976. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. J. Acoust. Soc. Am. 60, 410–417.

Mummery, C.J., Ashburner, J., Scott, S.K., Wise, R.J., 1999. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J. Acoust. Soc. Am. 106, 449–457.

Narain, C., Scott, S.K., Wise, R.J., Rosen, S., Leff, A., Iversen, S.D., Matthews, P.M., 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb. Cortex 13, 1362–1368.

Nielsen, F.Å., Hansen, L.K., 2002. Automatic anatomical labeling of Talairach coordinates and generation of volumes of interest via the BrainMap database. 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.

Nishitani, N., Schurmann, M., Amunts, K., Hari, R., 2005. Broca's region: from action to language. Physiology (Bethesda) 20, 60–69.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., Rauschecker, J.P., in press. Vowel sound extraction in anterior superior temporal cortex. Hum. Brain Mapp.

Parker, G.J., Luzzi, S., Alexander, D.C., Wheeler-Kingshott, C.A., Ciccarelli, O., Lambon Ralph, M.A., 2005. Lateralization of ventral and dorsal auditory-language pathways in the human brain. NeuroImage 24, 656–666.

Pisoni, D.B., 1977. Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. J. Acoust. Soc. Am. 61, 1352–1361.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806.

Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D., 1981. Speech perception without traditional speech cues. Science 212, 947–949.

Romanski, L.M., Tian, B., Fritz, J.B., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 2000. Reply to "'What', 'where' and 'how' in auditory cortex". Nat. Neurosci. 3, 966.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.

Scott, S.K., Wise, R.J., 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition 92, 13–45.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.

Stevens, K.N., Klatt, D.H., 1974. Role of formant transitions in the voiced–voiceless distinction for stops. J. Acoust. Soc. Am. 55, 653–659.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme, Stuttgart.

Tuomainen, J., Andersen, T., Tiippana, K., Sams, M., 2005. Audio-visual speech perception is special. Cognition 96, B13–B22.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., Liddle, P.F., 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J. Cogn. Neurosci. 13, 994–1005.

Winer, B.J., Brown, D.R., Michels, K.M., 1991. Statistical Principles in Experimental Design. McGraw-Hill, New York.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage 14, 1370–1386.

Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. 12, 900–918.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.