Task-Dependent Activations of Human Auditory Cortex to Prototypical and Nonprototypical Vowels

Kirsi Harinen,^{1*} Olli Aaltonen,¹ Emma Salo,¹ Oili Salonen,²
and Teemu Rinne^{1,3}

¹*Institute of Behavioural Sciences, University of Helsinki, Finland*

²*Helsinki Medical Imaging Center, Helsinki University Central Hospital, Finland*

³*Advanced Magnetic Imaging Centre, Aalto University School of Science, Finland*

Abstract: Research in auditory neuroscience has largely neglected the possible effects of different listening tasks on activations of auditory cortex (AC). In the present study, we used high-resolution fMRI to compare human AC activations with sounds presented during three auditory and one visual task. In all tasks, subjects were presented with pairs of Finnish vowels, noise bursts with pitch and Gabor patches. In the vowel pairs, one vowel was always either a prototypical /i/ or /ae/ (separately defined for each subject) or a nonprototype. In different task blocks, subjects were either required to discriminate (same/different) vowel pairs, to rate vowel “goodness” (first/second sound was a better exemplar of the vowel class), to discriminate pitch changes in the noise bursts, or to discriminate Gabor orientation changes. We obtained distinctly different AC activation patterns to identical sounds presented during the four task conditions. In particular, direct comparisons between the vowel tasks revealed stronger activations during vowel discrimination in the anterior and posterior superior temporal gyrus (STG), while the vowel rating task was associated with increased activations in the inferior parietal lobule (IPL). We also found that AC areas in or near Heschl’s gyrus (HG) were sensitive to the speech-specific difference between a vowel prototype and nonprototype during active listening tasks. These results show that AC activations to speech sounds are strongly dependent on the listening tasks. *Hum Brain Mapp* 34:1272–1281, 2013. © 2012 Wiley Periodicals, Inc.

Key words: attention; auditory cortex; functional magnetic resonance imaging; humans; speech

INTRODUCTION

Although the neural bases of speech processing are not well understood, it is generally accepted that human auditory cortex (AC) has evolved to process speech-related information. In the current theoretical models of AC operations, auditory information is processed in parallel pathways (e.g., separate “what” and “where” pathways for identification of objects and localization of sounds in space, respectively) linking different auditory cortical fields and other cortical areas [Hickok and Poeppel, 2007; Kaas and Hackett, 2000; McLachlan and Wilson, 2010; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000; Recanzone and Cohen, 2010; Romanski et al., 1999; Scott

Contract grant sponsor: Academy of Finland; Contract grant numbers: 1135900, 1141006, 1141563 (to T.R.); Contract grant sponsor: Research Funds of the University of Helsinki (to T.R.).

*Correspondence to: Kirsi Harinen, Institute of Behavioural Sciences, P. O. Box 9, FI-00014 University of Helsinki, Finland. E-mail: kirsi.harinen@helsinki.fi

Received for publication 19 July 2011; Revised 6 October 2011; Accepted 10 October 2011

DOI: 10.1002/hbm.21506

Published online 30 January 2012 in Wiley Online Library (wileyonlinelibrary.com).

and Johnsrude, 2003]. The models also typically maintain a hierarchical organization in which areas of primary AC in or near Heschl's gyrus (HG) are involved in spectrotemporal analysis, whereas other regions in the supratemporal gyrus (STG), supratemporal sulcus (STS), and adjacent inferior parietal lobule (IPL) are more involved in processing of higher-order auditory information (such as speech) and are less sensitive to spectrotemporal variation. It is not known whether some AC regions are dedicated to speech processing or whether all AC regions are involved in processing both speech and nonspeech information [Griffiths and Warren, 2002]. Animal and human studies have demonstrated that active listening tasks strongly modulate information processing in AC [Fritz et al., 2005; Hall et al., 2000; Lomber and Malhotra, 2008; Petkov et al., 2004; Rinne et al., 2009]. However, previous studies on speech processing have not systematically investigated the effect of different listening tasks on AC activations to speech sounds. In the present study, we examined how AC activations to Finnish prototype and nonprototype vowels differ during vowel discrimination and vowel rating tasks.

Earlier studies suggest that prototypical and nonprototypical speech sounds are processed differently in the human brain [Aaltonen et al., 1997; Iverson and Kuhl, 1991; Kuhl, 1991]. This is because the prototypes (i.e., excellent category exemplars of one's native language) act as "perceptual magnets" so that speech sounds near a prototype are perceived more similar to each other than speech sounds near a nonprototype, even though the physical difference between the sounds is kept constant [Kuhl, 1991]. Therefore, a prototype vowel is more difficult to discriminate from its neighbors than a less prototypical one. At least one previous study suggests that the difference between processing of prototypical and nonprototypical vowels also modulates AC activations observed with fMRI [Guenther et al., 2004].

Previous fMRI studies have shown differences in the distribution of AC activations during spatial and nonspatial auditory tasks. In general, the results of these studies are consistent with the assumption that areas of the posterior STG and IPL are more involved in processing of spatial information ("where" stream), while areas of the anterior STG process nonspatial information ("what" stream) [Alain et al., 2009; Arnott et al., 2004; Barrett and Hall, 2006; Degerman et al., 2006; Warren and Griffiths, 2003]. However, our recent study [Rinne et al., 2009] showed that activations in the anterior and posterior AC are differently modulated also during two nonspatial tasks. In that study, the activations of the anterior AC were increased during pitch discrimination but not during n-back pitch memory tasks, while activations in the posterior AC were enhanced during n-back pitch memory tasks but not during pitch-discrimination tasks. These results suggest that task-dependent modulations of AC activations observed during active auditory tasks are more complicated than commonly assumed. It is likely that AC activations to speech are also strongly modulated by different tasks [Hickok and Poeppel, 2007].

In the present study, we used high-resolution fMRI to investigate AC activations during two different auditory tasks performed on speech sounds. Subjects were presented with stimulus pairs from two vowel (Finnish /i/ and /ae/) continua in all conditions. During one task, subjects were required to discriminate (same/different) vowel pairs from either of the continua. During the other task, subjects listened to the same sounds and rated the "goodness" of the vowels (first/second sound was a better exemplar of the vowel category). The vowel pairs were arranged so that one sound was always either a prototypical /i/ or /ae/ (separately defined for each subject) or a nonprototype and the other sound was either the same or a different sound from that vowel continuum. Concurrent with the vowel stream, subjects were also presented with independent streams of noise bursts [iterated rippled noise (IRN)] and visual stimuli (Gabor gratings). This paradigm allowed us to compare task-dependent AC activations to sounds during two tasks performed on either vowel pairs organized around a prototype or nonprototype, during IRN pitch discrimination, and, in addition, stimulus-dependent AC activations (i.e., in the absence of an auditory task) during a visual task (Gabor orientation discrimination). We hypothesized that this paradigm would reveal task-dependent activation differences due to processing of speech-specific properties of the sounds.

MATERIALS AND METHODS

Subjects

Subjects ($N = 20$, 12 women) were right handed, 19 to 41 years of age (mean 24 years) with normal hearing (self reported), normal or corrected-to-normal vision, and no history of psychiatric or neurological illnesses. An informed written consent was obtained from each subject before the experiment. The study protocol was approved by the Ethical Committee of the Hospital District of Helsinki and Uusimaa, Finland. All subjects were native Finnish speakers. One subject was excluded from the analysis due to excessive motion during fMRI.

Stimuli

Two vowel continua were synthesized using the Praat software package (version 5.1.12; available at: www.praat.org). Each continuum consisted of 19 Finnish vowels. One continuum ranged from /a/ (as in "father") to /ae/ (as in "cat") and the other from /y/ (high rounded vowel) to /i/ (as in "beet") (Fig. 1a,b). The duration of each vowel was 200 ms including a linear 5 ms onset and offset ramp. In the /a/-/ae/ continuum, the frequency of the second formant (F2) varied from 1,046 to 2,119 Hz and in the /y/-/i/ continuum from 1,518 to 2,882 Hz in steps of 30 mels. The first formant (F1) was fixed at 720 Hz in the /a/-/ae/ continuum and at 250 Hz in the /y/-/i/ continuum. The other

Figure 1.

Spectrograms of five vowel stimuli from (a) /a/–/ae/ continuum and (b) /y/–/i/ continuum. Both continua consisted of 19 vowels separated by 30 mel steps in F2. (c) Oscillograms of IRN (top) and vowel (bottom) streams and (d) structure of the visual stream. Note that the iterated rippled noise (IRN) bursts (dura-

tion 100 ms), vowel pairs (two 200 ms vowels separated by a 100 ms gap), and Gabor gratings (duration 100 ms) were presented in concurrent, asynchronous streams at a relative fast rate in all task blocks. For details, see Methods.

formants (F3–F5) were in both continua as follows: F3 = 3,010 Hz, F4 = 3,300 Hz, and F5 = 3,850 Hz. A linear falling contour from 150 to 100 Hz was used for F0.

Prototype and nonprototype vowels were defined separately for each subject. First, subjects were presented with self-paced vowels from the two vowel continua. Each vowel variant was presented 20 times in random order. Subjects were required to indicate by pressing one of the two buttons whether they heard /a/ or /ae/ or, in a separate run, whether they heard /y/ or /i/. Next, the vowels (variants of /ae/ and /i/) that were categorized consistently (more than 75% of the cases) as /ae/ or /i/ were used to define prototype (P) and nonprototype (NP) vowels for each subject. Subjects were required to rate these sounds (1 = poor category exemplar, 4 = good category exemplar) relative to a good Finnish pronunciation of

/ae:/ as in “sääri” /sae:ri/ (“leg”) and of /i:/ as in “tiili” /ti:li/ (“brick”). For each subject, the /ae/ and /i/ vowels with the highest rating score were used as prototypes and the ones with the lowest rating as nonprototypes.

Fourteen different bursts of iterated rippled noise (IRN) were generated by iteratively adding delayed noise (three iterations, pitch range corresponding to 97–120 Hz, duration 100 ms). Visual stimuli consisted of 14 Gabor gratings (duration 100 ms, orientation range –60 to 60 degrees) presented at fixation.

The auditory stimuli were delivered binaurally with an UNIDES ADU2a audio system (Unides Design, Helsinki, Finland) via plastic tubes through a porous EAR-tip (ER3, Etymotic Research) acting as an earphone. The noise of the scanner was attenuated by the earplugs, circumaural ear protectors (Bilsom Mach 1, SNR 23 dB), and viscoelastic

Figure 2.

Structure of the stimulus streams (time scale on the horizontal axis is schematic). (a) In 30 s blocks (alternating with 10 s rest with no stimuli), subjects were presented with concurrent and asynchronous streams of iterated rippled noise (IRN) bursts with varying pitch, (b) Gabor gratings with varying orientation, and (c, d) vowel pairs. In pitch and Gabor discrimination tasks (a, b), subjects were required to detect pitch or orientation changes, respectively, and to indicate the direction of the change. During the vowel (same/different) discrimination task and vowel rating task (first/second part was better exemplar of the phoneme category), there were both (c) prototype (50%) or (d) nonprototype vowel pair blocks. Prototypes and nonprototypes were defined individually before fMRI scanning. The difference between the first and second part of the vowel pairs was 30, 60, 90, or 120 mels. For details, see Methods.

mattress inside the body coil and around the head coil and under the subject. The visual stimuli were presented in the middle of a screen viewed through a mirror fixed to the head coil.

Stimulus Streams and Behavioral Tasks

During fMRI, subjects were presented with concurrent streams of IRN bursts, Gabor gratings, and vowel pairs. IRN bursts were presented with a 110 to 190 ms onset-to-onset interval so that one pitch was repeated 8 to 10 times after which the pitch slightly increased or decreased (Fig. 1c, top; Fig. 2a). Gabor gratings were presented with a 140 to 200 ms onset-to-onset interval so that one orientation was repeated seven to nine times after which the orientation changed slightly (Fig. 1d; Fig. 2b). The vowel stream consisted of pairs of 200 ms vowels separated by a 100 ms gap (Fig. 1c, bottom; Fig. 2c,d). In prototype blocks, the vowel pairs were arranged so that one sound was always a prototypical /i/ or /ae/ and the other sound was either the same (50%) or a different sound from that

vowel continuum (F2 30, 60, 90, or 120 mel lower than F2 of the prototype). In nonprototype blocks, the vowel pairs were constructed in the same way around a nonprototype (16 subjects: F2 30–120 mel higher than F2 of the nonprototype in both vowel continua; three subjects: F2 30–120 mel lower than F2 of the nonprototype in one or the other vowel continuum). Vowel pairs were presented with a 1,200 to 1,600 ms onset-to-onset interval. Prototype and nonprototype pairs were presented in separate blocks. Note that the IRN bursts, Gabor patches, and vowel pairs were presented at a relative fast rate. Thus, a stimulus could be overlapped by 0 to 2 concurrent stimuli in the other streams. However, the streams were presented asynchronously with independent onset-to-onset intervals (Fig. 1c,d).

There were four different tasks. In vowel discrimination task, subjects were required to indicate whether the vowel pair consisted of same or different vowels. In vowel rating task, subjects indicated whether the first or second vowel of the pair was a better exemplar of the phoneme category. In both tasks, subjects responded by pressing one of two buttons (left button for “same” and “first” responses) with their right hand. In the IRN pitch discrimination task, subjects were required to detect pitch changes in the IRN stream and indicate the direction of the pitch change by pressing two buttons (left button for “down”). Analogously, in the visual task, subjects were required to detect orientation changes in the Gabor grating stream and to indicate the direction of the change (left button for “counterclockwise”).

The stimuli were presented in 30 s blocks alternating with 10 s silent breaks. During the breaks, subjects focused on a fixation mark (x) presented in the middle of a screen and waited for the next task. A task instruction symbol (“=” vowel discrimination task; “+” vowel rating task; “^” IRN pitch discrimination task; “○” visual task), replaced the fixation mark 3 s before the onset of the next task block and remained on the screen until the end of that block. The order of the task conditions was randomized.

For the IRN pitch discrimination, visual, vowel discrimination, and vowel rating task there were 11, 11, 22, and 22 blocks, respectively, so that 66 blocks were presented in total. Half of all task blocks were with prototype vowel pairs. Each block contained 20 to 24 targets (i.e., pitch changes, orientation changes, or vowel pairs). Total duration of the experiment was 45 min. The experiment was controlled using Presentation software (Neurobehavioral Systems, Albany, CA).

Analysis of the Behavioral Data

Mean hit rates (HRs) and reaction times (RTs) were calculated separately for each task. In vowel tasks, responses (with both buttons, i.e., same/different and first/second responses were not analyzed separately) occurring between 500 and 1,500 ms from the onset of the vowel pair were accepted as hits. In the IRN pitch discrimination and visual tasks responses (with both buttons) occurring between 200

and 1,200 ms from change onset were accepted as hits. Other responses (i.e., extra responses after a hit or responses outside the response window) were considered as false alarms. HR was defined as the number of hits divided by the number of targets (i.e., vowel pairs or pitch/orientation changes). False alarm rate (FaR) was defined as the number of false alarms divided by the number of all responses.

fMRI Data Acquisition and Analysis

Functional images were acquired using a T2*-weighted gradient-echo echo-planar-imaging (GE-EPI) sequence (TR 2,048 ms, TE 32 ms, flip angle 90°, voxel matrix 96 × 96, FOV 20 cm, slice thickness 2.1 mm with no gap, in-plane resolution 2.1 mm × 2.1 mm, number of slices 24; 3.0 T GE Signa system retrofitted with an Advanced NMR operating console and 8-channel head coil). The middle EPI slices were aligned along Sylvian fissures to cover the superior temporal lobe, insula, and most of the inferior parietal lobe in both hemispheres.

The functional scanning was performed in one 45 min run (approximately 1,322 volumes). After the functional scans, a fluid-attenuated inversion recovery image using the same imaging slices but with denser in-plane resolution was acquired (FLAIR; TR 10,000 ms, TE 120 ms, voxel matrix 320 × 192, FOV 20 cm, slice thickness 2.1 mm). Finally, at the end of the session, a high-resolution anatomical image was acquired (voxel matrix 256 × 256, FOV 25.6 cm, resolution 1.0 mm × 1.0 mm × 1.0 mm).

Global voxel-wise analysis was performed using the tools developed by the Analysis Group at the Oxford Centre for Functional MRI of the Brain (FMRIB) and implemented within FMRIBs software library (FSL, release 4.1; available at: www.fmrib.ox.ac.uk/fsl). The data was motion-corrected, high-pass filtered (cutoff 100 s), and spatially smoothed (Gaussian kernel of 5 mm full-width half-maximum). First-level statistical analysis of fMRI responses to different task blocks was carried out using FMRIBs improved linear model. Based on the timing information recorded during the experiment, each functional image was labeled as vowel rating (prototype or nonprototype), vowel discrimination (prototype or nonprototype), IRN pitch discrimination (with prototype or nonprototype vowels), visual task (with prototype or nonprototype vowels), or baseline (10 s breaks with no stimuli). The hemodynamic response function was modeled with a gamma function (mean lag 6 s, SD 3 s) and its temporal derivative. Contrasts were specified to create Z-statistic images testing for task and prototype/nonprototype effects.

For analysis across participants, the data were anatomically normalized in the following steps: first, cortical surfaces were extracted from high-resolution anatomical images, transformed to spherical standard space, and anatomically normalized on the basis of the cortical gyral and sulcal patterns using FreeSurfer (v5.0.0; available at: <http://surfer.nmr.mgh.harvard.edu>). Next, the three-dimensional (3D) spherical cortical surfaces were rotated

and projected to a two-dimensional (2D) space separately for each hemisphere using equal area Mollweide projection (Python libraries matplotlib and basemap; available at: <http://matplotlib.sourceforge.net>). This procedure produced 3D-to-2D anatomical transformation matrices that were then applied separately for each subject to transform the results of the 3D second-level statistical analysis to 2D.

Finally, the group analysis (FMRIBs local analysis of mixed effects, $N = 19$) was run on these flattened data. Z-statistic images were thresholded using clusters determined by $Z > 2.1$ (ANOVA) or $Z > 2.3$ (separate contrasts) and a (corrected) cluster significance threshold of $P < 0.05$ (using Gaussian random field theory).

RESULTS

Behavior

On average, subjects ($N = 19$) responded to 95% of the vowel pairs (i.e., no response was given to 5% of the vowel pairs), 80% of the pitch targets, and 86% of the visual orientation changes. The proportion of accepted responses was slightly lower in the vowel rating task (94%) than in the vowel discrimination task [97%; repeated measures ANOVA with factors Task (vowel discrimination, vowel rating) and Stimulus (prototype, nonprototype), main effect of Task, $F_{(1,18)} = 5.4$, $P < 0.05$] and slightly lower for non-prototype (95%) than prototype (96%) vowel pairs [main effect of Stimulus, $F_{(1,18)} = 5.5$, $P < 0.05$] especially in the vowel rating task [93% vs. 96%; interaction of Task and Stimulus, $F_{(1,18)} = 6.6$, $P < 0.05$]. Responses in the vowel rating task (923 ms) were also slower than in the vowel discrimination task [878 ms; main effect of Task, $F_{(1,18)} = 21$, $P < 0.001$]. Mean RTs for visual discrimination task and IRN pitch discrimination task were 551 ms and 686 ms, respectively. Mean FaRs in vowel rating, vowel discrimination, visual discrimination, and IRN pitch discrimination task were 0.02, 0.01, 0.05, and 0.09, respectively.

fMRI

A second-level (across subjects) repeated measures ANOVA with factors Task (vowel discrimination, vowel rating) and Stimulus (prototype, nonprototype) revealed significant ($N = 19$, threshold $Z > 2.1$, cluster-corrected $P < 0.05$) main effects. The main effect of Task (Fig. 3a, blue) was detected bilaterally in wide areas of the anterior insula, STG, and IPL. The main effect of Stimulus (red), in turn, was significant in areas in or near HG. Some of these areas also showed a significant interaction of Task and Stimulus factors (yellow).

Separate contrasts were then used to analyze AC activations in more detail. Activations to sounds in the absence of auditory task were isolated by contrasting activations during the visual task with sounds with activations during the 10 s breaks with no sounds. As expected, task-

irrelevant sounds were associated with bilateral activations ($N = 19$, threshold $Z > 2.3$, cluster-corrected $P < 0.05$) in the posterior STG including areas in or near HG (Fig. 3b, blue). A comparison of activations during vowel tasks (discrimination or rating) and the visual task with sounds revealed task-specific activation enhancement in the anterior insula, posterolateral STG, and areas of the IPL (c, blue and yellow). During the IRN pitch discrimination task (vs. visual task with sounds) activations were modulated in the anterior insula and wide areas of the anterior and posterior STG (c, red and yellow) including most of the areas that were activated by presentation of task-irrelevant sounds during the visual task (b, blue). Direct comparison of activations during vowel rating and vowel discrimination tasks showed stronger activations during vowel rating in areas of the IPL bilaterally (d, blue). During vowel discrimination, in turn, activations were enhanced in the anterior insula, anterior STG, in areas in or near HG, and posterior STG (d, red). A contrast between vowel-task blocks with prototype and nonprototype vowel pairs showed stronger activations during tasks with prototype than nonprototype vowel pairs bilaterally in areas in or near HG (e, red). Further, separate contrasts between blocks with prototype and nonprototype vowel pairs during vowel discrimination and vowel rating tasks revealed that this effect was mainly due to stronger activations in the prototype than nonprototype blocks during vowel discrimination (f, red). In the vowel rating task, no significant differences were detected between prototype and nonprototype blocks. There were no significant differences between prototype and nonprototype blocks presented during the IRN pitch discrimination or during the visual task.

DISCUSSION

The present study was designed to test whether AC activations to speech sounds are modulated by active listening tasks and whether these modulations are different during tasks with prototype and nonprototype vowels. Our subjects were presented with concurrent streams of vowel pairs (organized around a prototype or nonprototype), IRN noise bursts with pitch changes and Gabor gratings with orientation changes in all task conditions, the only difference being in the behavioral instruction. We found that activations in wide AC areas were differently modulated by vowel discrimination and vowel rating tasks (Fig. 3a, main effect of Task, blue). Direct comparison of vowel discrimination and vowel rating tasks revealed stronger activations associated with discrimination in the anterior and posterior STG including areas in or near HG (d, red), while areas in the IPL were more activated during vowel rating (b, blue). These activation differences between vowel discrimination and vowel rating task are quite remarkable as, in both tasks, subjects were to give a discriminative response and, thus, it would seem that in both tasks processing requirements were quite similar. Further, such

Figure 3.

(a) Results of a second-level repeated measures ANOVA with factors Task (vowel discrimination, vowel rating) and Stimulus (prototype, nonprototype). Activations ($N = 19$, threshold $Z > 2.1$, cluster-corrected $P < 0.05$) are shown on a flattened mean 2D cortical surface. Blue, main effect of Task; Red, main effect of Stimulus; Yellow, interaction of Task and Stimulus. (b–f) Results of separate contrasts. (b) Activations to task-irrelevant sounds during visual task (blue). (c) Comparisons of activations during vowel tasks versus visual task (blue and yellow) and IRN pitch discrimination versus visual task (red and yellow). Areas showing significant activations in both contrasts are shown in yellow. (d) Areas where activations were stronger during vowel rating than vowel discrimination tasks (blue) and areas where activations were stronger during vowel discrimination than vowel rating (red). (e) Comparison of activations during nonprototype (NP) and prototype (P) blocks collapsed across the vowel tasks. (f) Areas where activations were stronger during prototype than nonprototype blocks during vowel rating (blue) and vowel discrimination (red). (g) Anatomical labels. STG superior temporal gyrus, HG Heschl's gyrus, IPL inferior parietal lobule (angular gyrus and supramarginal gyrus). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

activation differences between two quite similar nonspatial tasks performed on nearly identical vowel stimuli are not easily explained by the current auditory models.

Originally, it was assumed that anterior and posterior AC belong to functionally separate auditory processing streams so that areas in the anterior STG are more involved in nonspatial processing (e.g., pitch; “what” stream) while areas immediately posterior to HG (planum temporale) and areas of IPL are associated with processing of spatial information (“where” stream). More recently, it has been proposed that the areas within the posterior stream are probably not associated with spatial processing as such but rather have a more general functional role in supporting integration of auditory and motor information [Arnott and Alain, 2011; Hickok et al., 2009, 2011; Rauschecker, 2011; Smith et al., 2010]. In general, the present results are in line with the idea that areas in the anterior and posterior AC have different functional roles in processing of auditory information. However, our results are not consistent with the assumption that areas immediately posterior to HG (planum temporale) and areas of IPL have a similar functional role in auditory processing. In the present study, all tasks required similar discriminative (motor) responses. It could be argued, though, that the vowel rating task required more speech-specific processing (e.g., auditory-motor integration) than the vowel discrimination task. This could have caused the stronger IPL activations during vowel rating than during discrimination tasks (Fig. 3d, blue). However, areas immediately posterior to HG were more associated with the vowel discrimination task (d, red). Thus, the present results suggest that areas immediately posterior to HG and IPL are functionally distinct.

In the vowel rating task, subjects were required to identify whether the vowel pair was from the /ae/ or /i/ category and then to compare both vowels with a mental representation of this vowel category. It is possible that such categorical processing in the vowel rating task explains the observed activation differences between vowel rating and vowel discrimination tasks. This account is consistent with the results of our previous study showing enhanced IPL activations associated with a categorical pitch memory task as compared with a pitch discrimination task [Rinne et al., 2009]. Areas of the posterior STG and supramarginal gyrus (part of IPL) have been implicated in categorical processing also in other previous studies [Chang et al., 2010; Raizada and Poldrack, 2007].

Importantly, we found that activations in or near HG were different during processing of vowels in prototype and nonprototype pairs (Fig. 3a, main effect of Stimulus, red; note that the present study was not designed to localize primary AC or other auditory cortical fields, see Woods et al., 2010). Some AC areas also showed an interaction of task (discrimination vs. rating) and stimulus (prototype vs. nonprototype) effects (yellow). Direct comparison of activations during prototype and nonprototype vowel-task blocks showed enhanced activations during prototype blocks in areas in or near HG bilaterally (e, red).

This activation increase was mainly due to stronger AC activations during vowel discrimination with prototype pairs than with nonprototype pairs (f, red). In the vowel rating task, we detected no significant differences between prototype and nonprototype blocks. Note that the stimuli in the prototype and nonprototype blocks were spectrotemporally highly similar as all vowels were taken from the same two vowel continua and some vowel pairs could even appear in both prototype and nonprototype blocks. Consistently, a comparison of activations during visual task with prototype and nonprototype vowel pairs showed no significant stimulus-dependent activation differences (in the anterior or posterior STG) in the absence of an auditory task. There also were no significant activation differences between IRN discrimination task blocks with prototype and nonprototype vowel pairs. Thus, these results demonstrate that AC areas in the vicinity of HG are sensitive to the speech-specific difference between a vowel prototype and nonprototype in a task-dependent manner. In most previous imaging studies, speech-specific activations have been mainly observed in the anterior or posterior STG [Binder et al., 2004; Chang et al., 2010; Desai et al., 2008; Jäncke et al., 2002; Obleser et al., 2006] or in the superior temporal sulcus [STS; Binder et al., 2000; Desai et al., 2008; Hickok and Poeppel, 2007; Liebenthal et al., 2005; Scott and Johnsrude, 2003]. However, recent results of studies using multiparametric pattern-recognition analysis techniques have also suggested speech effects in AC areas in or near HG. Kilian-Hütten et al. [2011] showed that perceptual interpretation of physically identical phonemes can be decoded from activations of areas near HG and Staeren et al. [2009] argued that these areas may participate in categorical auditory representations. In line with these results, a recent study by Woods et al. [2011] used population-based cortical-surface analysis to characterize the processing of consonant–vowel–consonant syllables (CVCs) and spectrally matched amplitude-modulated noise bursts (AMNBs) in AC and reported that lateral areas of primary AC (A1) showed a CVS preference. Together these results suggest that areas in or near HG, possibly including areas of primary AC, are specialized to process speech sounds.

The present enhanced activations during the discrimination-task prototype blocks were probably because the vowels in the prototype pairs were perceptually more similar with each other than the vowels in the nonprototype pairs (the magnet effect) [Kuhl, 1991]. However, it is possible that, due to the perceptual similarity, the discrimination task with prototype pairs required more elaborate spectrotemporal analysis of the vowels. Thus, although the activation differences between prototype and nonprototype blocks are definitely a consequence of speech-specific features of the stimuli, it could be argued that the enhanced activations observed in or near HG are not necessarily related to speech-specific processing as such. This account is in line with the reverse hierarchy theory which maintains that perception is mostly based on high-level

representations and only when a finer discrimination is needed access to the lower level information representing the detailed information may be gained [Kilian-Hütten et al., 2011; Nahum et al., 2008]. Thus, in the present vowel discrimination task, the enhanced activations to prototype pairs might be due to task-dependent analysis of lower-level representations in areas in or near HG.

In the vowel rating tasks, no statistically significant differences were detected between prototype and nonprototype blocks. This seems rather surprising as one would expect that also in this task the vowels had to be analyzed in detail and that, similar to the vowel discrimination task, this analysis would require different resources in the prototype and nonprototype blocks. However, in the vowel rating task, there was no objectively correct answer as no reference vowel was given. Thus, subjects may have based their decisions on more subjective criteria and on less throughout analysis of the vowels. Further, in the vowel rating task subjects were first required to identify the category of the vowel pair and then to compare both vowels with a mental representation of the vowel category (/i/ or /ae/). It is possible that such activation of mental representations of prototype vowels during both prototype and nonprototype blocks in the vowel rating task concealed any activation differences between these conditions. It is also possible that due to these and other features of the vowel rating task, it was associated with more varying activations making it more difficult to detect systematic differences between prototype and nonprototype blocks.

In the present study, we found stronger activations to vowel pairs organized around a prototype than to vowel pairs organized around a nonprototype. In contrast, a previous study reported *weaker* AC activations to prototype than nonprototype vowels [Guenther et al., 2004]. In that study, subjects were presented with one repetitive prototype vowel /i/ (i.e., “judged to be a good example of /i/ by most listeners”) and one repetitive nonprototype vowel (i.e., “judged by most to be a bad example of /i/, chosen from near the boundary with the vowel /ɪ/”) in separate blocks. Subjects were to “attend to the stimuli by listening for differences from sound to sound, although all stimuli within a block were the same.” The authors found significantly less activation in areas of posterior AC to the prototypical vowel blocks than to the nonprototypical vowel blocks. They proposed that this was because the prototype vowel was associated with a smaller neural representation than the nonprototype vowel (due to category learning). This account rests on the assumption that the authors were able to measure pure representations (or stimulus-dependent activations) associated with processing of prototypical and nonprototypical vowels. However, this assumption is not necessarily true as it is possible that there were also (uncontrolled) task-level differences between the prototypical and nonprototypical vowel blocks. The prototype blocks consisted of one repetitive vowel (/i/). The nonprototype vowel, in turn, was selected near a category boundary, so that it, by definition, was perceived as either /i/ or /ɪ/.

During the prototypical vowel blocks, it was probably quite evident that a repetitive vowel /i/ was presented and that in a block beginning with that vowel all the stimuli would be identical. During nonprototype blocks, however, the repetitive vowel was perceptually ambiguous so that the task (i.e., detect differences from one sound to the next) probably was more interesting and required more effort throughout the block. As a result, subjects were more engaged in the task during nonprototype blocks than during the prototype blocks. Such task-level difference between the blocks, rather than differences between processing of prototypical and nonprototypical vowels as such, could have resulted in activation differences in areas of posterior AC.

Previous studies have shown that AC activations observed with fMRI are strongly modulated by attention-engaging listening tasks, especially in areas posterior and lateral to HG [Grady et al., 1997; Petkov et al., 2004; Rinne et al., 2008, 2009; Woodruff et al., 1996; Woods et al., 2009]. Consistently, in the present study, activations were stronger during the vowel tasks than during the visual tasks in the posterior STG (Fig. 3c, blue and yellow). However, during the IRN discrimination task, activations were enhanced in wide AC areas extending from the anterior to posterior STG (c, red and yellow). This difference between activations during the IRN discrimination and the vowel tasks performed on similar stimuli could be due to several reasons. First, it is possible that the vowels were more salient than the auditory background consisting of the high-rate IRN stream and scanner noise and were thus difficult to ignore. Because of this, it could be that the vowels were associated with relatively strong activations also during the visual task. This could have decreased the activation differences between the vowel and visual tasks. Further, our previous study showed that the magnitude of attention-related modulation of AC activations increases with increasing presentation rate [Rinne et al., 2005]. The fact that the IRN bursts were presented with a higher rate than the vowels could have caused the stronger and wider task-dependent activations observed during the IRN discrimination than during the vowel tasks. Thus, the difference between activations during the IRN pitch discrimination and vowel tasks could be due to interactions of task and stimulus effects and not due to differences between pitch and speech processing as such. Nevertheless, these results further strengthen the view that activations in wide AC areas are susceptible to task-dependent modulations [Rinne, 2010; Rinne et al., 2009].

CONCLUSIONS

Current theoretical models on speech processing in human AC are to a large degree based on the results of imaging studies comparing activations to speech and non-speech stimuli. Typically these studies have ignored the possible effects of different tasks on AC activations. In the present study, we obtained distinct activation patterns to

identical sounds (vowels and IRN bursts) during visual, vowel discrimination, vowel rating, and IRN pitch discrimination tasks. We found that areas of the anterior and posterior AC were differently activated depending on the characteristics of the vowel task. Such activation patterns are not easily predicted by the current auditory models. This suggests that subsequent studies should systematically investigate task-dependent modulations of AC activations in order to understand the functional organization of AC operations in processing of speech and nonspeech auditory information [Hickok and Poeppel, 2007].

ACKNOWLEDGMENTS

The authors thank Dr. David L. Woods, Ms. Susan McLaughlin, and two anonymous reviewers for insightful comments.

REFERENCES

- Aaltonen O, Eerola O, Hellström A, Uusipaikka E, Lang AH (1997): Perceptual magnet effect in the light of behavioral and psychophysiological data. *J Acoust Soc Am* 101:1090–1105.
- Alain C, McDonald KL, Kovacevic N, McIntosh AR (2009): Spatio-temporal analysis of auditory “what” and “where” working memory. *Cereb Cortex* 19:305–314.
- Arnott SR, Alain C (2011): The auditory dorsal pathway: Orienting vision. *Neurosci Biobehav Rev* 35:2162–2173.
- Arnott SR, Binns MA, Grady CL, Alain C (2004): Assessing the auditory dual-pathway model in humans. *NeuroImage* 22:401–408.
- Barrett DJ, Hall DA (2006): Response preferences for “what” and “where” in human non-primary auditory cortex. *Neuroimage* 32:968–977.
- 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, 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.
- Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010): Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428–1432.
- Degerman A, Rinne T, Salmi J, Salonen O, Alho K (2006): Selective attention to sound location or pitch studied with fMRI. *Brain Res* 1077:123–134.
- Desai R, Liebenthal E, Waldron E, Binder JR (2008): Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci* 20:1174–1188.
- Fritz JB, Elhilali M, Shamma SA (2005): Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci* 25:7623–7635.
- Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997): Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8:2511–2516.
- Griffiths TD, Warren JD (2002): The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA (2004): Representation of sound categories in auditory cortical maps. *J Speech Lang Hear Res* 47:46–57.
- Hall DA, Haggard MP, Akeroyd MA, Summerfield AQ, Palmer AR, Elliott MR, Bowtell RW (2000): Modulation and task effects in auditory processing measured using fMRI. *Hum Brain Mapp* 10:107–119.
- Hickok G, Houde J, Rong F (2011): Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69:407–422.
- Hickok G, Okada K, Serences JT (2009): Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101:2725–2732.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Iverson P, Kuhl PK (1996): Influences of phonetic identification and category goodness on American listeners’ perception of /r/ and /l/. *J Acoust Soc Am* 99:1130–1140.
- Jäncke L, Wüstenberg 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.
- Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011): Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31:1715–1720.
- Kuhl PK (1991): Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Percept Psychophys* 50:93–107.
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005): Neural substrates of phonemic perception. *Cereb Cortex* 15:1621–1631.
- Lomber SG, Malhotra S (2008): Double dissociation of ‘what’ and ‘where’ processing in auditory cortex. *Nat Neurosci* 11:609–616.
- McLachlan N, Wilson S (2010): The central role of recognition in auditory perception: A neurobiological model. *Psychol Rev* 117:175–196.
- Nahum M, Nelken I, Ahissar M (2008): Low-level information and high-level perception: The case of speech in noise. *PLoS Biol* 6:e126.
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roetinger M, Eulitz C, Rauschecker JP (2006): Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp* 27:562–571.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004): Attentional modulation of human auditory cortex. *Nat Neurosci* 7:658–663.
- Raizada RD, Poldrack RA (2007): Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740.
- Rauschecker JP (2011): An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res* 271:16–25.
- Rauschecker JP, Scott SK (2009): Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724.
- 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.
- Recanzone GH, Cohen YE (2010): Serial and parallel processing in the primate auditory cortex revisited. *Behav Brain Res* 206:1–7.
- Rinne T (2010): Activations of human auditory cortex during visual and auditory selective attention tasks with varying difficulty. *Open Neuroimag J* 7:187–193.

-
- Rinne T, Balk MH, Koistinen S, Autti T, Alho K, Sams M (2008): Auditory selective attention modulates activation of human inferior colliculus. *J Neurophysiol* 100:3323–3327.
- Rinne T, Koistinen S, Salonen O, Alho K (2009): Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *J Neurosci* 29:13338–13343.
- Rinne T, Pekkola J, Degerman A, Autti T, Jääskeläinen IP, Sams M, Alho K (2005): Modulation of auditory cortex activation by sound presentation rate and attention. *Hum Brain Mapp* 26:94–99.
- 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, Johnsrude IS (2003): The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Smith KR, Hsieh IH, Saberi K, Hickok G (2010): Auditory spatial and object processing in the human planum temporale: No evidence for selectivity. *J Cogn Neurosci* 22:632–639.
- Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009): Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol* 19:498–502.
- Warren JD, Griffiths TD (2003): Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J Neurosci* 23:5799–5804.
- Woodruff PW, Benson RR, Bandettini PA, Kwong KK, Howard RJ, Talavage T, Belliveau J, Rosen BR (1996): Modulation of auditory and visual cortex by selective attention is modality-dependent. *Neuroreport* 7:1909–1913.
- Woods DL, Herron TJ, Cate AD, Kang X, Yund EW (2011): Phonological processing in human auditory cortical fields. *Front Hum Neurosci* 5:42.
- Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010): Functional properties of human auditory cortical fields. *Front Syst Neurosci* 4:155.
- Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X (2009): Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS ONE* 4:e5183.
-