◆ Human Brain Mapping 24:11–20(2005) ◆

Voice Perception: Sex, Pitch, and the Right Hemisphere

Sonja Lattner,[1] Martin E. Meyer,[1,2]* and Angela D. Friederici[1]

[1]*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2]*Department of Neuropsychology, University of Zurich, Switzerland*

Abstract: The present functional magnetic resonance imaging (fMRI) study examined the neurophysiological processing of voice information. The impact of the major acoustic parameters as well as the role of the listener's and the speaker's gender were investigated. Male and female, natural, and manipulated voices were presented to 16 young adults who were asked to judge the naturalness of each voice. The hemodynamic responses were acquired by a 3T Bruker scanner utilizing an event-related design. The activation was generally stronger in response to female voices as well as to manipulated voice signals, and there was no interaction with the listener's gender. Most importantly, the results suggest a functional segregation of the right superior temporal cortex for the processing of different voice parameters, whereby (1) voice pitch is processed in regions close and anterior to Heschl's Gyrus, (2) voice spectral information is processed in posterior parts of the superior temporal gyrus (STG) and areas surrounding the planum parietale (PP) bilaterally, and (3) information about prototypicality is predominately processed in anterior parts of the right STG. Generally, by identifying distinct functional regions in the right STG, our study supports the notion of a fundamental role of the right hemisphere in spoken language comprehension. *Hum Brain Mapp 24:11–20, 2005.* © **2004 Wiley-Liss, Inc.**

Key words: speech; voice; pitch; formant structure; functional magnetic resonance imaging (fMRI); superior temporal region; right hemisphere

INTRODUCTION

In speech communication, the listener does not only decode the linguistic message from the speech signal, but at the same time he or she also infers paralinguistic information such as the age, gender, and other properties of the speaker. We will term this type of information voice information. In terms of acoustic properties, voice information has been described by several acoustic parameters. Two of these parameters are of major perceptual relevance: (1) the fundamental frequency and (2) the spectral formant frequencies [Fant, 1960; Lavner et al., 2000; Van Dommelen, 1990]. The fundamental frequency (F0) of the vocal fold vibration determines the perceived pitch of a voice; most often it is higher in females than in males. The spectral formants (F1, F2. . .) determine the characteristic timbre of each speaker's voice, so that the listener can recognize familiar voices and discriminate or categorize unfamiliar voices. A formant is a prominence in the frequency spectrum that depends on the size, length, and shape of the vocal tract. It has been reported that particularly the third and fourth formant correspond to the perceived speaker's gender, because they depend on the shape of the pharyngeal cavity, which is disproportionably larger in males [Lavner et al., 2000]. Despite the importance of voice information for the interpretation of an utterance,

Contract grant sponsor: Schweizer National Fonds (Swiss National Foundation); Contract grant number: SNF 46234101.

*Correspondence to: Martin Meyer, Department of Neuropsychology, University of Zurich, Treichlerstrasse 10, CH-8032 Zurich, Switzerland. E-mail: mmeyer@access.unizh.ch

Received for publication 22 November 2003; Accepted 17 May 2004

DOI: 10.1002/hbm.20065
Published online in Wiley InterScience (www.interscience.wiley.com).

hitherto neurocognitive research has largely ignored the domain of voice information. For decades, the study of neurologically impaired patients was the only source of information about the neural processing of voices. The clinical observations suggest that lesions of the temporal lobe of either hemisphere may lead to deficits in the discrimination of unfamiliar voices [Van Lancker and Kreiman, 1987; Van Lancker et al., 1988, 1989] whereas a lesion in the right hemisphere only leads to a deficit in the recognition of familiar voices [Van Lancker and Kreiman, 1987]. Recently, the neural processing of voice information has also been investigated by functional imaging experiments corroborating and complementing the clinical data. In a recent functional magnetic resonance imaging (fMRI) study [von Kriegstein et al., 2003], it was shown that a task targeting on the speaker's voice (in comparison to a task focussing on verbal content) leads to a response in the right anterior temporal sulcus of the listener. In another series of studies [Belin et al., 2000, 2002], it was shown that temporal lobe areas in both hemispheres responded more strongly to human voices than to other sounds (e.g., bells, dog barks, machine sounds) but that, again, it is the right anterior STS that responded significantly stronger to nonspeech vocalizations than to scrambled versions of the same stimuli. However, while these studies employed different categories of sounds as control stimuli for voice-selective effects, the present experiment sought to investigate the processing of differential acoustic parameters within the voice domain by identifying areas that are sensitive to these parameters.

In an earlier study using magnetoencepalography (MEG) [Lattner et al., 2003], we showed that a violation of the listeners' expectations by a non-prototypical voice (e.g., an extraordinarily low female voice) leads to a voice-specific brain response. This response occurs in primary auditory cortices as early as 200 ms after the stimulus onset. The data suggest that listeners have a certain expectation (i.e., a memory trace or template) about the combination or configuration of pitch and spectral properties of male and female voices that influences the brain response even at a preattentive processing level. The separate manipulation of either parameter, therefore, offers a valuable method to gain further information about the way voice information is processed. However, while MEG allows an excellent temporal resolution in the range of milliseconds, the spatial resolution achieved by this method is coarse. In the present fMRI experiment, we investigated the functionally relevant neuroanatomical correlates of specific information processing within the voice domain, focussing on the supratemporal cortex and in particular on the role of the anterior part of the right hemisphere.

A further aspect that to our knowledge has not been systematically investigated at all is the difference in the neural processing of voice gender in male and female listeners. Behavioural data suggest that the male-female distinction in voice perception is of major importance; e.g., infants are able to categorize voices at the age of 8 months [Patterson and Werker, 2002], and adults judge the similarity of voices according to a male-female categorization [for a discussion, see Mullenix et al., 1995]. In order to explore the neurophysiological correlates of voice gender perception, we systematically varied the voices we presented. In addition to varying the factor gender in the percept, we also varied this factor in the perceiver in order to see whether expectations about gender-specific properties of a voice are different between male and female persons listening to presented speech.

MATERIALS AND METHODS

Subjects

Sixteen subjects (8 male, mean age 26 $\pm$ 2 years; 8 female, mean age 24 $\pm$ 4 years) participated in the experiment. They were right-handed native speakers of German, reported no audiological or neurological disorders, and had normal structural MRI scans. Subjects were paid for participation and gave written informed consent in accordance with the guidelines approved by the Ethics Committee of the Leipzig University Medical Faculty.

Stimuli and Design

The stimuli consisted of 144 two-word sentences (e.g., *Albert lacht*/Albert laughs). The sentences were uttered both by a male and a female speaker. The average fundamental frequency for each speaker was determined (male, 124 Hz; female, 205 Hz) and the stimuli were shifted in fundamental frequency by the amount of the F0-difference, i.e., 81 Hz. The recordings were then manipulated using the PSOLA resynthesis function of the PRAAT speech editing software. Intensities were normalized using the Cool edit speech editing software; durations were matched across conditions (female speaker 1.509 ms vs. male speaker 1.318 ms).

Altogether, four types of experimental stimuli were employed and each type constituted an experimental condition. Condition M consisted of the sentences read by the male speaker, condition F consisted of the sentences read by the female speaker, condition F$-$ was comprised of the female-voice utterances that were lowered in pitch, and condition M+ contained the male utterances but with a high (female voice) pitch (Fig. 1).[1] Note that all conditions consisted of intelligible speech.

Procedure and Task

Four lists of 36 items per condition were made up of the stimulus material. Each subject was presented with one list. By this procedure, it was ensured that no subject heard a sentence more than once, but that across subjects each sen-

[1]Example sound files are available at "http://www.psychologie.unizh.ch/neuropsy/home_mmeyer/HBM-03-0175".

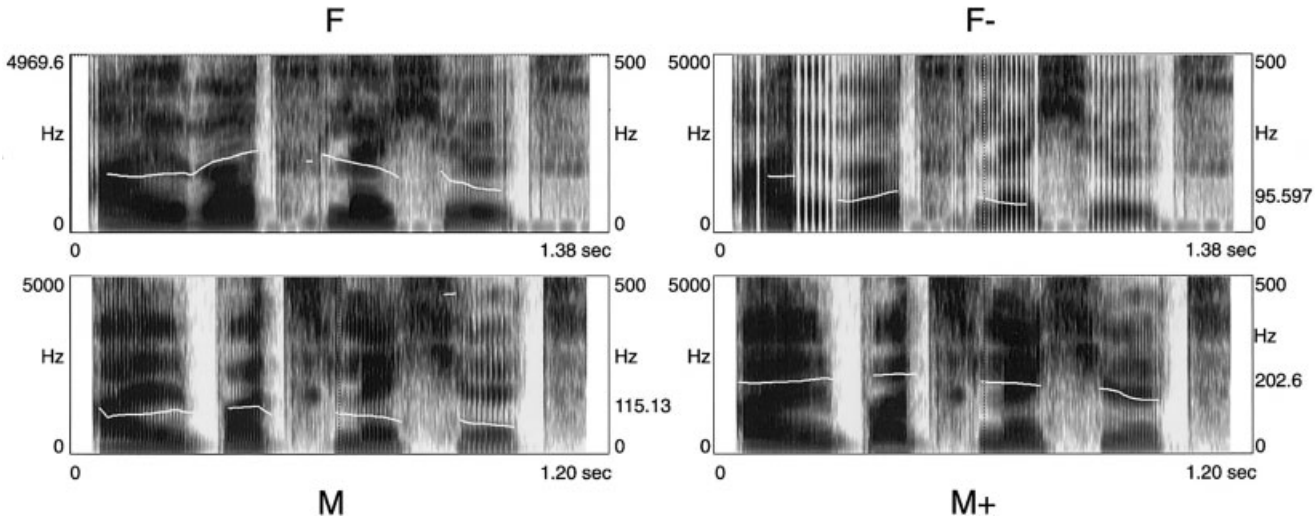

Figure 1.

Spectrograms and pitch contour derived from the four different voices speaking the same sentence. Each spectrogram corresponds to one experimental condition: F = natural female voice; F− = pitch-shifted female voice; M = natural male voice; M+ = pitch-shifted male voice. (Example sound files are available online at http://www.psychologie.unizh.ch/neuropsy/home_mmeyer/HBM-03-0175.)

tence was presented in each voice condition. This constraint also holds for the two listener–gender groups separately. The relation between image acquisition and trial presentation was systematically varied so that numerous time points along the response are sampled [Burock et al., 1998], with an average onset-to-onset interval of 6 s, allowing the fMRI signal to decrease adequately between trials. The subjects were instructed to indicate by a button press after presentation of each sentence whether the sentence was presented in a natural or unnatural voice (finger and hand for the answering reaction were balanced).

MR Imaging

MRI data were collected at 3T using a Bruker 30/100 Medspec system (Bruker Medizintechnik GmbH, Ettlingen, Germany). The standard bird cage head coil was used. Before MRI data acquisition, field homogeneity was adjusted by means of "global shimming" for each subject. Then, scout spin echo sagittal scans were collected to define the anterior and posterior commissures on a midline sagittal section. For each subject, structural and functional (echoplanar) images were obtained from 8 axial slices (5 mm thickness, 2 mm spacing, 64 × 64 with a FOV of 19.2 mm) parallel to the plane intersecting the anterior and posterior commissures (AC–PC plane). After defining the slices' position a set of two-dimensional (2-D) T1-weighted anatomical images (MDEFT sequence: TE 20 ms, TR 3,750 ms, in-plane resolution 0.325 mm^2) were collected in plane with the echoplanar images, to align the functional images to the 3-D images. A gradient-echo EPI sequence was used with a TE 30 ms, flip angle 90°, TR 2000 ms. In a separate session, high-resolution whole-head 3-D MDEFT brain scans (128 sagittal slices, 1.5-mm thickness, FOV 25.0 × 25.0 × 19.2 cm, data matrix of 256 × 256 voxels) were acquired additionally for reasons of improved localization [Norris, 2000; Ugurbil et al., 1993].

Data Analysis

For data analysis we used the LIPSIA software package [Lohmann et al., 2001]. Data preparation proceeded as follows: slice-wise motion correction (time step 50 as reference), sinc interpolation in time (to correct for fMRI slice acquisition sequence); baseline correction (cut-off frequency of 1/36 Hz); spatially smoothing using a Gaussian kernel with FWHM 6 mm. To align the functional data slices onto a 3-D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (3 rotational, 3 translational) was performed. The rotational and translational parameters were acquired on the basis of the 2-D MDEFT volume to achieve an optimal match between these slices and the individual 3-D reference dataset. Geometrical distortions of the EPI-T1 images were corrected by using additional EPI-T1 refinement on the transformation matrices. The resulting parameters, scaled to standard size, were then used to transform the functional slices using trilinear interpolation so that the resulting functional slices were aligned with the stereotactic coordinate system of [Talairach and Tournoux 1988].

Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations [Friston, 1994; Zarahn et al., 1997]. First,

TABLE I. Natural male vs. natural female voices*

Location	BA	Left hemisphere				Right hemisphere			
		Z score	*x*	*y*	*z*	Z score	*x*	*y*	*z*
Natural female > natural male									
SCG/STP	6/43/4	−3.60	−49	−15	21	−4.85	52	−13	15
	42/22					−3.53	62	−12	8
pSTG	22					−3.60	50	−34	15
IPL	40	−3.55	−56	−25	27	−3.53	56	−34	24
INS		−3.55	−29	−10	18	−3.80	32	−19	21
HeG	41	−4.59	−37	−34	12				

* This table and Tables II–IV list results of averaged contrast images based on individual contrasts between conditions. To assess the significance of an activation locus, averaged contrast images were thresholded with ($P < 10^{-3}$ one-tailed, uncorrected for multiple comparisons). Localization is based on stereotactic coordinates [Talairach and Tournoux, 1988]. These coordinates refer to the location of maximal activation indicated by the Z score in a particular anatomical structure. Distances are relative to the intercommissural (AC–PC) line in the horizontal (*x*), anterior-posterior (*y*), and vertical (*z*) directions. The table only lists activation clusters exceeding a minimal size of 50 voxels.
SCG, subcentral gyrus; STP, supratemporal plane; pSTG, posterior superior temporal gyrus; IPL, inferior parietal lobe; INS, insula; HeG, Heschl's gyrus.

statistical parametric maps (SPM) were generated for each subject. The design matrix was generated with the standard hemodynamic response function considering a response delay of 6 s and its first and second derivative. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of FWHM 4 s. The model includes an estimate of temporal autocorrelation. The effective degrees of freedom were estimated as described by Worsley and Friston [1995]. Thereafter, contrast maps (i.e., estimates of the raw-score differences of the beta coefficients between specified conditions), were generated for each subject. As the individual functional datasets were all aligned to the same stereotactic reference space, a group analysis was subsequently performed. A one-sample *t*-test of contrast maps across subjects was computed to indicate whether observed differences between conditions were significantly different from zero as suggested by Holmes and Friston [1998]. Obtained *t*-values were subsequently transformed into Z-values giving an SPM Z for each subject and condition. Voxels exceeding the threshold $|Z = 3.09|$, corresponding to $P < 10^{-3}$, were reported as significant results.

RESULTS

Behavioural Results

In 92% of the cases, the subjects judged the naturalness of the voices as expected (correct response: "natural" for condition M and F; "unnatural" for conditions M+ and F−). The individual analysis showed that 12 of the 16 subjects had above 90% correct responses. One subject had 88% correct responses because he tended to judge the manipulated male voice (M+) on some occasions as "natural". The other three subjects had above 70% correct. The lower rate of these three subjects is due to a consistent behaviour: While they made no deviant judgements about the other voices, two of them rated the high male voice as "natural"; the third subject judged the low female voice as "natural".

fMRI Results

Four contrasts were performed, looking at the processing of different voice parameters: In the first contrast, the processing of natural male and female voices (M vs. F) was evaluated; the second contrast focused on the role of the fundamental frequency by comparing high- and low-pitched voice conditions (M/F− vs. F/M+); in the third contrast, pitch was balanced across conditions, but the formant related activation was investigated (M/M+ vs. F/F−); and in the fourth contrast, the role of voice prototypicality or naturalness was explored (M/F vs. M+/F−).

The role of speaker's gender

Table I shows that the contrast of the natural voice conditions M vs. F revealed stronger responses to the female voice than to the male voice in the right hemisphere and was comprised of three main centres of gravity: (1) in the supratemporal plane (STP) anteriorly to Heschl's gyrus extending to the foot of the central sulcus and adjacent ventral pre- and postcentral gyri; (2) in the posterior part of the superior temporal gyrus (STG) and adjacent parietal operculum extending into the inferior parietal lobe (IPL); and (3) in the superior part of the first long insular gyrus in the right hemisphere.

In the left hemisphere, activation was observed in the posterior central gyrus, the inferior parietal lobe, and the insula.

A second level analysis revealed no significant differences between the activation patterns of female and male listeners;

TABLE II. Pitch*

Location	BA	Left hemisphere Z score	*x*	*y*	*z*	Right hemisphere Z score	*x*	*y*	*z*
Low pitch > high pitch									
aCG/Gs	25	4.18	−2	2	−3				
High pitch > low pitch									
STP	42/22					−3.82	54	−7	10
INS						−3.93	38	−7	15

* Functional activation indicated separately for contrasts between conditions. For further explanation, see Table I.
aCG, anterior cingulate gyrus; Gs, subcallosal gyrus; STP, supratemporal plane; INS, insula.

both groups showed the right hemisphere dominance in response to the female voices.

The role of the fundamental frequency (pitch)

A contrast of the brain responses to high-pitched voices (F/M+) and low-pitched voices (M/F−) was performed. In this contrast the formant spectra are balanced across the conditions (Table II, Fig. 2A).

Stronger activity in response to the high-pitched voices was observed in the right-hemispheric temporal lobe portion anterior to Heschl's gyrus and in the insula. In the left hemisphere, stronger activation in response to the low-pitched voices was observed in the subcallosal area of the anterior cingulate gyrus.

The role of the formant structure (vocal tract)

A contrast of the brain activation in response to all voices with a male formant spectrum (M, M+), and the activation patterns in response to all voices with a female formant spectrum (F, F−) was performed, whereby the fundamental frequency level was balanced (Fig. 2B, Table III).

Stronger activation in response to the female vocal tract voices was found in the posterior part of the left and right superior temporal region (planum temporale, planum parietale). Subcortical activation was also observed in left hemisphere brain areas, namely in the white matter and thalamus. Stronger activation in response to the male

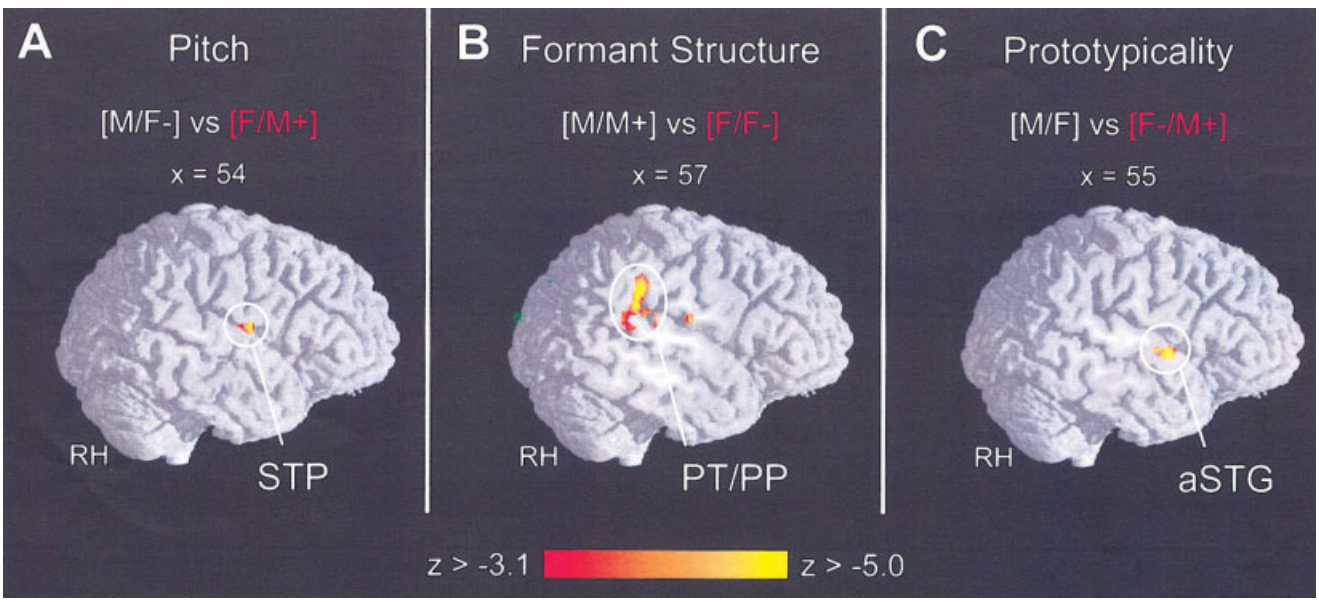

Figure 2.
The brain scans show right hemisphere parasagittal view of direct comparison along three signal regressors: (**A**) pitch, (**B**) formant structure, and (**C**) prototypicality (|Z| 3.09, α-level 0.001). Functional inter-subject activation (N = 16) is superimposed onto a normalised 3-D reference brain.

TABLE III. Formant structure*

Location	BA	Left hemisphere Z score	x	y	z	Right hemisphere Z score	x	y	z
Male voices > female voices									
IFG	46/45	3.65	−44	35	6				
Female voices > male voices									
HeG	41					−4.32	40	−31	12
STG	22	−4.11	−47	−34	15				
IPL	40					−4.3	57	−35	26
SCG/	6/43/4					−3.73	56	−13	18
ROP		−3.75	−29	−10	15	−3.85	23	−7	15
Tha		−3.68	−8	−10	3				

* Functional activation indicated separately for contrasts between conditions. For further explanation, see Table I.
IFG, inferior frontal gyrus; HeG, Heschl's gyrus; STG, superior temporal gyrus; IPL, inferior parietal lobe; SCG, subcentral gyrus; ROP, Rolandic operculum; Tha, thalamus.

voice spectra was found in the left inferior frontal gyrus (rostrodorsal part of the pars triangularis).

The role of voice prototypicality

In a final contrast, the role of voice prototypicality was investigated. Balancing fundamental frequency and formant spectra, i.e., the acoustic parameters, we contrasted the brain responses to natural voices (F, M) with the activation patterns evoked by the odd voices (F−, M+) (Table IV, Fig. 2C).

Odd voices produced activation in the right anterior superior temporal gyrus (STG). However, this particular contrast revealed temporal lobe activation that is even more anterior relative to the other results mentioned above. No selective responses to normal voices were found.

DISCUSSION

The aim of the present experiment was to investigate the processing of voice information by means of functional resonance imaging and to dissociate areas that are sensitive to different acoustic parameters in the voice signal. We suggest a functional segregation of the right superior temporal areas involved in voice processing based on the signal parameters, which we are going to discuss in turn.

We first investigated the brain activation in response to natural male and female voices (contrast: F vs. M). The activation pattern shows a stronger response to the female voice. The activation was bilateral with a clear right hemispheric dominance in temporal lobe regions including the anterior and posterior part of the supratemporal plane (STP) and the insula. We can only speculate upon the reason for the stronger brain activation in response to the female voices. One explanation could be that the signal is perceptually more salient than the male voice. However, a more cognitively loaded explanation focuses on the social and biological relevance of female/high-pitched voices. For instance, it is well known that neonates already prefer female voices over male voices [Fifer and Moon, 1988] and also the resemblance of high voices to children's voices might enhance the adults' physical response to the high-pitched signals. Moreover, an increase in pitch signals a speaker's increasing stress and, hence, a potential danger. All these examples suggest that the perceptual system is more aroused by high-pitched/female voices, which might explain the stronger brain activation in this experimental condition. The observed spatial distribution of the activation resembles clinical data, reporting an impairment of voice discrimination following brain damage to

TABLE IV. Prototypicality*

Location	BA	Left hemisphere Z score	x	y	z	Right hemisphere Z score	x	y	z
Odd voices > natural voices									
aSTG	22					−4.74	55	−1	0

* Functional activation indicated separately for contrasts between conditions. For further explanation, see Table I.
aSTG, anterior superior temporal gyrus.

both hemispheres, and a deficit in the recognition of familiar voices after right hemisphere brain damage [Van Lancker and Kreiman, 1987]. Note that the present categorization task involves low level perception as well as a comparison between the incoming signal and prototypical memory traces ("templates") of natural voices and is, therefore, comparable to a recognition task. The first imaging study that proclaimed the existence of voice selective areas [Belin et al., 2000], identified cortical sites along the STS, bilaterally with a right hemispheric preponderance, albeit not as clearly lateralized as observed in the present study. While these temporal areas exhibited a voice selectivity with respect to between-domain processing (i.e., voice signals vs. object sound), we suggest that right supratemporal areas and the brain regions nearby play a crucial role in the more fine-grained classification processes within the voice domain. Fundamental frequency (F0) and the spectral formant configuration are the most salient acoustic features in gender classification of natural voices. The values of these parameters are not independent from each other but correlate in real life, i.e., a formant configuration related to a female vocal tract is typically correlated with a relatively high F0. Although F0 and other formants are physically very similar, we hypothesized distinct functional areas to play a role in perception for two reasons. First, we derived our predictions from recent studies on simple auditory, speech, and music perception. Converging evidence obtained from these studies suggests dissociation in the (right) hemisphere [Hall et al., 2003] or even a hemispheric difference for the two types acoustic cues with the right hemisphere preferentially driving pitch processing and the left hemisphere particularly mediating the processing of fast spectral formant transitions [Zatorre et al., 2002]. Second, our hypotheses are based on the notion of differential functional relevance of these parameters in the closely related domains of voice and speech perception. In speech perception the F0 contour transfers information on the sentence structure (e.g., phrase boundaries), of the talker's attitude (e.g., ironic), and other paralinguistic information. In contrast, the spectral formants help decode the linguistic message, from phonemes to words. The perceptual system treats these types of information distinctly. Therefore, it is plausible to assume that functionally and neurally distinct areas mediate F0 level and spectral formant configuration, although they are correlated in real life. In the present study, we considered and tested fundamental frequency (F0) and the spectral formant configuration independently in order to learn more about the processing of the single features as well as of the configuration of the given parameters.

A very salient feature distinguishing male and female voices is their different pitch level, i.e., in terms of the acoustic substance, the fundamental frequency [Lass et al., 1975; Van Dommelen, 1990]. Consequently, the activation observed in the first contrast (M/F− vs. F/M+) might largely be explained by the different pitch levels of the male voice and the female voice, independent of other factors. In order to investigate this assumption, a second contrast of high- and low-pitched voices was performed (F/M+ vs. M/F−). Note that both conditions consist of voices with female and male vocal timbre; i.e., in terms of the spectral formant structure, the contrast is balanced and pitch is the only relevant variable. Perception of low pitch (M/F− > M +/F) leads to activation in the subcallosal area, which is an unexpected result. The subcallosal area is a small area of cortex on the medial surface of each cerebral hemisphere and is the ventralmost portion of the "affective division" of the anterior cingulate cortex (ACad) [Bush et al., 2000]. Little is known about the specific function of the subcallosal area so we can only speculate on the functional relevance of this finding. Generally, ACad primarily is recruited in assessing the salience of emotional information [Bush et al., 2000]. High pitch is often correlated with positive emotional valence whereas low pitch is more likely to signal sad emotional valence. Thus, it may be possible that the uncertainty about emotional information in low-pitch voices yields an enhancement of activity in the ACad. Alternatively, activity of the subcallosal area might imply that participants had unpleasant sensations when perceiving low-pitch voices, in particular the pitch-shifted female voice, which indeed sounds unaesthetic. Perception of high pitch (M/F − < M +/F) uncovers stronger responses in the right peri-auditory cortex. As can be seen in Tables I and II, the activation pattern evoked by high-pitch voices partly overlaps with responses to natural female voices in the right hemisphere STP anteriorly adjacent to Heschl's gyrus and in the right insula. However, the activation posterior to Heschl's gyrus observed in the first contrast did not occur, and must, hence, be attributed to other than F0-related differences.

A plausible hypothesis could be that the posterior part of the right temporal lobe is involved in the processing of spectral/formant information. To investigate this hypothesis, we contrasted the brain activation in response to all female voices with all voices comprised of a male formant spectrum (F/F− vs. M/M+; note that this contrast is balanced in terms of fundamental frequency). Voices with male formant configurations (M/M+), when compared to the female voices, activated the triangular part of the left inferior frontal gyrus, which is an intriguing finding. This region has been previously associated with auditory perception of simple and complex sentences [Caplan et al., 1999; Müller et al., 1998]. In the present study, participants also heard simple sentences; however, sentences were equal in complexity so that our findings cannot be explained in terms of complexity. The mean Z-value of this cluster was 3.23 ($\sigma = 0.12$), which is just above the significance threshold ($Z > 3.09$) we applied. Therefore, given the data at hand, we are not in the position to provide a suitable interpretation and we con-

clude that this issue needs to be clarified in follow-up studies.

In general, a clearly stronger activation was found in response to the voices with female vocal tract characteristics (F/F−). This activation was observed in the posterior part of the right supratemporal plane (planum temporale, planum parietale), extending even into the parietal lobe, which may also speak to a particular sensitivity of this region for specific aspects of processing vocal information. Interestingly, this contrast also yielded activation in left hemisphere brain areas, particularly in the superior temporal gyrus. The planum temporale (PT) of the left hemisphere is known to play an important role in speech processing [Hickok and Pöppel, 2000; Jäncke et al., 2002; Meyer et al., 2004; Pöppel, 2003]. This region is involved in the phonetic analysis, particularly in the perception of fast formant transitions; for example, activation of the PT is observed in response to CV syllables but not to steady vowels [Jäncke et al., 2002]. The present response pattern can be interpreted in two ways: First, activation of the left hemispheric PT could be explained by the processing of speaker-specific formant transitions due to articulatory characteristics; second, it could be a correlate of the analysis of the constant aspects of the formant spectrum that reflect each talker's vocal tract and, hence, his or her unique voice timbre. Further investigations need to clarify this issue. However, brain-imaging studies on speech- and voice-related parameters suggest that the perception of speech and voice during on-line speech processing cannot be considered independent. Even though some studies report a neurophysiological dissociation of speech perception and the recognition of familiar voices [Assal et al., 1981], there is meanwhile considerable evidence indicating that the perception/discrimination of unfamiliar voices interacts with speech perception: understanding what is said facilitates speaker recognition [Goggin et al., 1991], talker familiarity facilitates speech perception [Nygaard et al., 1994], and even in terms of clinical [Ziegler et al., 1999] and neurophysiological [Knösche et al., 2002] data, there is no evidence in favour of a neat dissociation of speech and voice perception.

Having explored the impacts of single acoustic parameters, next we investigated the role of voice prototypicality. Balancing fundamental frequency as well as formant spectra, we contrasted the brain responses to natural voices with the activation patterns evoked by the unusual voices (F/M vs. F−/M+). This contrast revealed stronger vascular responses only to unusual voices in the right anterior temporal lobe. However, this temporal lobe activation is anterior to what has been observed in the earlier contrasts reported here. Thus, the anterior part of the right superior temporal gyrus (aSTG) appears to be less sensitive to feature-based acoustic processing of F0 and formants, but it is responsible for the processing of the combination of the single parameters [Lattner et al., 2003]. The anterior part of the right hemisphere, therefore, reflects the highest level of processing during voice recognition as investigated in the present study. Results of other recent neuroimaging studies that report a particular role of the right anterior STS in response to voice-related processing [von Kriegstein et al., 2003], adaptation to a particular speaker [Belin and Zatorre, 2003], or even voice-specific perception across domains [Belin et al., 2000] clearly support this conclusion even though the acoustic substance of the stimuli was comparable or even identical across conditions, thus requiring a cognitively higher analysis or categorization processes. Therefore, our data concur with Belin's notion of cortical regions selective to sounds of voice [Belin et al., 2004].

A final aspect that we investigated was the role of the listener's gender. Other fMRI studies report gender-related differences in the laterality of brain activation to cognitively higher, language-related processes [Jaeger et al., 2000; Kansaku and Kitazawa, 2001; Phillips et al., 2000; Shaywitz et al., 1995], although not unequivocally, e.g., [Frost et al., 1999]. In addition, gender-related differences were also reported for pitch memory [Gaab et al., 2003], where male listeners exhibited a stronger lateralisation of the activation, as well as for magnetic brain responses in early pure tone processing [Salmelin et al., 1999]. The role of voice quality as a feature of sexual attraction and its potential role with respect to the selection of a potential partner also support the hypothesis that there would be gender-related differences in neurophysiological processing of various voice signals. However, the present study revealed no significant differences in the activation patterns of male and female listeners in response to male and/or female voices but rather supports studies reporting behavioural null effects in gender-related voice perception in children [Mann et al., 1979] as well as in adults [Mullenix et al., 1995].

CONCLUSIONS

The present study is the first neuroimaging experiment to look at the processing of speaker information within the voice domain and takes the speaker's and the listener's gender into account. The gender of the listener does not seem to influence the way of processing of voice-related information. However, a general functional segregation of right hemisphere areas dependent on the processing of different signal features is suggested: (1) pitch is predominately processed by mid STG areas, anterior but close to Heschl's Gyrus; (2) spectral voice properties are predominately processed by overlapping areas in the posterior STG and IPL (planum temporale/planum parietale); and (3) prototypicality is processed by areas of the right anterior STG. However, while we show that these regions exhibit a sensitivity to the acoustic parameters relevant in voice processing, we do not argue that they are restricted to voice processing, i.e., that they are "voice selective" in a strict sense. Instead, we suggest a close overlap between voice and speech processing. Based on this observation, we argue in favour of a more explicit and more careful consideration of the role of voice-related perceptual processes in current models of speech perception.

ACKNOWLEDGMENTS

We thank Lutz Jäncke and two anonymous reviewers for helpful comments on the manuscript. Martin Meyer is supported by Schweizer National Fonds (Swiss National Foundation, SNF 46234101, "Short-term and long-term plasticity in the auditory system").

REFERENCES

Assal G, Aubert C, Buttet J (1981): Asymétrie cérébrale et reconnaissance de la voix. Rev Neurol (Paris) 137:255–268.

Belin P, Zatorre RJ (2003): Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P, Zatorre RJ, Ahad P (2002): Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17–26.

Belin P, Fecteau S, Bédard C (2004): Thinking the voice: neural correlates of voice perception. TICS 8:129–135.

Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM (1998): Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. NeuroReport 9:3735–3739.

Bush G, Luu P, Posner MI (2000): Cognitive and emotional influences in anterior cortex. TICS 4:215–222.

Caplan D, Alpert N, Waters G (1999): PET studies of syntactic processing with auditory sentence presentation. NeuroImage 9:343–351.

Fant G (1960): Acoustic theory of speech production. The Hague: Mouton.

Fifer W, Moon C (1988): Auditory experience in the fetus. In: Smotherman W, Robertson S, editors. Behavior of the fetus. West Cadwell, NJ: Telford Press. p 175–188.

Friston KJ (1994): Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

Frost JA, Binder JR, Springer JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999): Language processing is strongly left lateralized in both sexes. Evidence from functional MRI. Brain 122: 199–208.

Gaab N, Keenan JP, Schlaug G (2003): The effect of gender on the neural substrates of pitch memory. J Cog Neurosci 15:810–820.

Goggin J, Thompson C, Sturbe G, Simental L (1991): The role of language familiarity in voice identification. Mem Cognit 19:448–458.

Hall DA, Hart HC, Johnsrude IS (2003): Relationships between human auditory cortical structure and function. Audiol Neurootol 8:1–18.

Hickok G, Pöppel D (2000): Towards a functional neuroanatomy of speech perception. TICS 4:131–138.

Holmes AP, Friston KJ (1998): Generalisability, random effects, and population inference. NeuroImage 7:754.

Jaeger J, Lockwood A, Van Valin R, Kemmerer D, Murphy B, Wack D (2000): Sex differences in brain regions activated by grammatical and reading tasks. NeuroReport 9:2803–2807.

Jäncke L, Wüstenberg T, Scheich H, Heinze H (2002): Phonetic perception and the temporal cortex. NeuroImage 15:733–746.

Kansaku K, Kitazawa S (2001): Imaging studies on sex differences in the lateralization of language. Neurosci Res 41:333–337.

Knösche T, Lattner S, Maess B, Schauer M, Friederici A (2002): Early parallel processing of auditory word and voice information. NeuroImage 17:1493–1503.

Lass N, Hughes K, Bowyer D, Waters L, Bourne V (1975): Speaker sex identification from voiced, whispered, and filtered isolated vowels. J Acoust Soc Am 59:675–678.

Lattner S, Maess B, Schauer M, Alter K, Friederici A (2003): Dissociation of human and computer voices in the brain: evidence for a preattentive gestalt-like perception. Hum Brain Mapp 20:13–21.

Lavner Y, Gath I, Rosenhouse J (2000): The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun 30:9–26.

Lohmann G, Müller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (2001): Lipsia: a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imag Graph 25:449–457.

Mann V, Diamond R, Carey S (1979): Development of voice recognition: parallels with face recognition. J Exp Child Psychol 27: 153–165.

Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004): Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89:277–289.

Mullenix J, Johnson K, Topcu-Durgun M, Farnsworth L (1995): The perceptual representation of voice gender. J Acoust Soc Am 98:3080–3095.

Müller RA, Chugani DC, Behen ME, Rothermel RD, Muzik O, Chakraborty PK, Chugani HT (1998): Impairment of dentato-thalamo-cortical pathway in autistic men: language activation data from positron emission tomography. Neurosci Lett 245: 1–4.

Norris DG (2000): Reduced power multi-slice MDEFT imaging. Magn Reson Imag 11:445–451.

Nygaard L, Sommers MS, Pisoni DB (1994): Speech perception as a talker-contingent process. Psychol Sci 5:42–46.

Patterson M, Werker JF (2002): Infants' ability to match dynamic phonetic and gender information in the face and voice. J Exp Child Psychol 81:93–115.

Phillips M, Lowe M, Lurito J, Dzemidzic M, Mathews V (2000): Temporal lobe activation demonstrates sex-based differences during passive listening. Radiology 220:220–207.

Pöppel D (2003): The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun 41:245–255.

Salmelin R, Schnitzler A, Parkkonen L, Biermann K, Helenius P, Kiviniemi K, Kuukka K, Schmitz F, Freund H (1999): Native language, gender, and functional organization of the auditory cortex. Proc Natl Acad Sci USA 96:10460–10465.

Shaywitz BA, Shaywitz SE, Pugh KR, Constable RT, Skudlarski RK, Fulbright RK, Bronen RA, Fletcher JM, Skudlarski P, Katz L, Gore JC (1995): Sex differences in the functional organisation of the brain for language. Nature 373:607–609.

Talairach J, Tournoux P (1988): Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Ugurbil K, Garwood M, Ellermann J, Hendrich K, Hinke R, Hu X, Kim SG, Menon R, Merkle H, Ogawa S, R. S (1993): Magnetic fields: Initial experiences at 4T. Magn Reson Q 9:259.

Van Dommelen W (1990): Acoustic parameters in human speaker recognition. Lang Speech 33:259–272.

Van Lancker D, Kreiman J (1987): Voice discrimination and recognition are separate abilities. Neuropsychologia 25:829–834.

Van Lancker D, Cummings J, Kreiman J, Dobkin B (1988): Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex 24:195–209.

Van Lancker D, Kreiman J, Cummings J (1989): Voice perception deficits: Neuroanatomical correlates of phonagnosia. J Clin Exp Neuropsychol 11:665–674.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003): Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48–55.

Worsley KJ, Friston KJ (1995): Analysis of fMRI time series revisited, again. NeuroImage 2:173–181.

Zarahn E, Aguirre G, D'Esposito M (1997): Empirical analysis of BOLD-fMRI statistics. I. Spatially smoothed data collected under null-hypothesis and experimental conditions. NeuroImage 5:179–197.

Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. TICS 6:37–46.

Ziegler W, Jochmann A, Zierdt A (1999): Assessment of auditory word comprehension in aphasia. In: Maassen B, Groenen P, editors. Pathologies of speech and language: advances in clinical phonetics and linguistics. London: Whurr Publishers Ltd. p 229–235.