Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex

Martin Meyer\textsuperscript{a,b,*}, Stefan Zysset\textsuperscript{b}, D. Yves von Cramon\textsuperscript{b}, Kai Alter\textsuperscript{b,c}

\textsuperscript{a}Department of Neuropsychology, University of Zurich, Treichlerstrasse 10, CH-8032 Zurich, Switzerland
\textsuperscript{b}Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
\textsuperscript{c}School of Neurology, Neurobiology, and Psychiatry, Newcastle Auditory Group, University of Newcastle, UK

Accepted 7 February 2005
Available online 29 March 2005

Abstract

In this event-related fMRI study, 12 right-handed volunteers heard human laughter, sentential speech, and nonvocal sounds in which global temporal and harmonic information were varied whilst they were performing a simple auditory target detection. This study aimed to delineate distinct peri-auditory regions which preferentially respond to laughter, speech, and nonvocal sounds. Results show that all three types of stimuli evoked blood–oxygen-level-dependent responses along the left and right peri-sylvian cortex. However, we observed differences in regional strength and lateralization in that (i) hearing human laughter preferentially involves auditory and somatosensory fields primarily in the right hemisphere, (ii) hearing spoken sentences activates left anterior and posterior lateral temporal regions, (iii) hearing nonvocal sounds recruits bilateral areas in the medial portion of Heschl’s gyrus and at the medial wall of the posterior Sylvian Fissure (planum parietale and parietal operculum). Generally, the data imply a differential regional sensitivity of peri-sylvian areas to different auditory stimuli with the left hemisphere responding more strongly to speech and with the right hemisphere being more amenable to nonspeech stimuli. Interestingly, passive perception of human laughter activates brain regions which control motor (larynx) functions. This observation may speak to the issue of a dense intertwining of expressive and receptive mechanisms in the auditory domain. Furthermore, the present study provides evidence for a functional role of inferior parietal areas in auditory processing. Finally, a post hoc conjunction analysis meant to reveal the neural substrates of human vocal timbre demonstrates a particular preference of left and right lateral parts of the superior temporal lobes for stimuli which are made up of human voices relative to nonvocal sounds.

© 2005 Elsevier B.V. All rights reserved.

Theme: Neural basis of behavior
Topic: Cognition

Keywords: Speech; Laughter; Vocal timbre; Functional Magnetic Resonance Imaging (fMRI); Peri-sylvian region; Auditory cortex

1. Introduction

Converging evidence obtained from a plethora of clinical and neuroimaging observations points to a residence of speech and auditory bottom-up functions in the superior temporal lobes of both cerebral hemispheres [5–7,11,12,23, 39,41,45,46,50,52–55,60,68,73,85,86,101,107,112]. Consistently, these studies demonstrate a sequence of sensory and cognitive processes mediated by brain regions stretching along from the primary auditory cortex (PAC)\textsuperscript{1} into different

\textsuperscript{*} Corresponding author. Department of Neuropsychology, University of Zurich, Treichlerstrasse 10, CH-8032 Zurich, Switzerland. Fax: +41 44 634 4342.

\textsuperscript{1} Notably, Heschl’s gyrus in its entirety does not exclusively accommodate primary auditory cortex [77]. Cytoarchitectonic and morphometric studies rather demonstrate that the primary auditory cortex (PAC) usually covers the most medially situated two thirds of Heschl’s gyrus [67]. However, human auditory areas are not confined to the PAC but are presumed to extend across the surface of the supratemporal plane (STP), including the planum polare (PPol) and the planum temporale (PT) even into the insula (INS) and the parietal operculum (PaOp) [78].
anterior and posterior parts of the peri-sylvian cortex and superior temporal sulcus (STS), either on the left or right hemisphere, depending on the acoustic or linguistic complexity of stimuli. Numerous recent neuroimaging studies compared speech and nonspeech stimuli (e.g., white noise bursts, sine wave tones, frequency-modulated tones, amplitude-modulated tones) to distinguish between brain regions that are more sensitive to speech and nonspeech stimuli [11,73,87]. However, by comparing spoken language with nonspeech auditory cues, one has to consider a multitude of aspects in which speech and nonvocal sounds may differ, for example, vocal timbre. Therefore, the present study uses human laughing as experimental condition. In terms of phonetics, human laughter is akin to speech as it also consists of syllable-like units but it provides less F0-variation, monotonous articulatory movements, and a different rhythmical structure. Perception of both speech and laughter therefore requires rapid temporal processing but the latter to a lesser degree. This makes laughter an interesting acoustic stimulus to examine, since the perception of laughter may activate the same brain regions as speech to a certain extent but can also be considered a holistic percept which may convey relevant paralinguistic and affective information in the context of communication. In this context, the brain’s right hemisphere enters the limelight as recent neuroimaging research has brought up some unspecific evidence which associates the right hemisphere with paralinguistic processes during speech perception [53,62]. The specific role the right hemisphere may play during processing vocal sounds has been vaguely indicated [6,7,56]. In particular, the studies conducted by Belin et al. uncovered an enhanced sensitivity to paralinguistic vocal information of the right superior temporal sulcus and gyrus while the contralateral structures appear to be more amenable to linguistic information during spoken language comprehension [7,26]. These findings are not in agreement with recent studies which propose a functional segregation of the supratemporal structures appearing in the premotor/frontal opercular area and leading through the motor cortex [105]. Clinical studies observed patients suffering from neurological disorders which caused ‘pathological laughter’ that is inappropriate, mood-independent, and uncontrollable laughter. However, it seems that pathological laughter originates from numerous pathological factors during brain disease rather than emerging from circumscribed lesions to particular brain sites [72]. Apparently, expressive laughter may be a phenomenon so complex that it cannot be assigned to any single brain area but depends on the temporal coordination of several areas constituting a functional network. This view receives partial support by recent brain imaging data which report a network incorporating left insula, bilateral amygdala and bilateral auditory cortices subserving the perception of human laughter [80]. Notably, activity in the right auditory cortices was stronger than in contralateral areas of the left hemisphere. According to the authors, this rightward asymmetry of supratemporal activity may be explained in terms of a right auditory cortex specialization for processing changes of pitch, timbre, and tonal patterns. While these authors used silence and human crying as baseline conditions in their study, we contrast human laughing with speech and nonvocal sounds to reveal a potential functional sensitivity of peri-sylvian regions.

In terms of phonetics, laughter and speech have essential acoustic underpinnings in common. First, speech and laughter share the same source of vocalization [75]. Like speech, laughter can also be described as a stream of acoustic cues which unfold in time. Typically, laugh bursts form a phrase-like sequence of repetitive laugh pulses that are initiated by an ingestion (see Methods section). Laughter involves the respiratory system and the vocal apparatus and is produced by changes in the vocal and sub-glottal, respiratory tracts as well as in the musculoskeletal system [19] that result in staccato-like syllabic vocalizations. Like speech, a sequence of laughter syllables can be combined into utterances. These utterances start with a vocalized inhalation and consist of a series of short laugh pulses with almost constant time intervals [66,76]. The reiteration of syllables with their specific segmental characteristics enables us to identify it as laughter. In contrast to speech, laughter contains less F0-variance. Laughter has a rich harmonic structure [74] and spectral qualities like formant frequencies can be evaluated [3]. Interestingly, both speech and laughter arise from laryngeal

1.1. What makes laughter so special?

Little is known about brain mechanisms of laughter. Generally, laughter can be considered a vocal expressive communicative signal and was there before man developed speech [79]. A recent review article suggested that expressive laughter can be associated with an ‘emotionally driven’ system involving the amygdala, thalamic/hypothalamic and subthalamic areas and with a ‘voluntary’ system originating in the premotor/frontal opercular area and leading through the motor cortex [105]. Clinical studies observed patients suffering from neurological disorders which caused ‘pathological laughter’ that is inappropriate, mood-independent, and uncontrollable laughter. However, it seems that pathological laughter originates from numerous pathological factors during brain disease rather than emerging from circumscribed lesions to particular brain sites [72]. Apparently, expressive laughter may be a phenomenon so complex that it cannot be assigned to any single brain area but depends on the temporal coordination of several areas constituting a functional network. This view receives partial support by recent brain imaging data which report a network incorporating left insula, bilateral amygdala and bilateral auditory cortices subserving the perception of human laughter [80]. Notably, activity in the right auditory cortices was stronger than in contralateral areas of the left hemisphere. According to the authors, this rightward asymmetry of supratemporal activity may be explained in terms of a right auditory cortex specialization for processing changes of pitch, timbre, and tonal patterns. While these authors used silence and human crying as baseline conditions in their study, we contrast human laughing with speech and nonvocal sounds to reveal a potential functional sensitivity of peri-sylvian regions.

In terms of phonetics, laughter and speech have essential acoustic underpinnings in common. First, speech and laughter share the same source of vocalization [75]. Like speech, laughter can also be described as a stream of acoustic cues which unfold in time. Typically, laugh bursts form a phrase-like sequence of repetitive laugh pulses that are initiated by an ingestion (see Methods section). Laughter involves the respiratory system and the vocal apparatus and is produced by changes in the vocal and sub-glottal, respiratory tracts as well as in the musculoskeletal system [19] that result in staccato-like syllabic vocalizations. Like speech, a sequence of laughter syllables can be combined into utterances. These utterances start with a vocalized inhalation and consist of a series of short laugh pulses with almost constant time intervals [66,76]. The reiteration of syllables with their specific segmental characteristics enables us to identify it as laughter. In contrast to speech, laughter contains less F0-variance. Laughter has a rich harmonic structure [74] and spectral qualities like formant frequencies can be evaluated [3]. Interestingly, both speech and laughter arise from laryngeal

Little is known about brain mechanisms of laughter. Generally, laughter can be considered a vocal expressive communicative signal and was there before man developed speech [79]. A recent review article suggested that expressive laughter can be associated with an ‘emotionally driven’ system involving the amygdala, thalamic/hypothalamic and subthalamic areas and with a ‘voluntary’ system originating in the premotor/frontal opercular area and leading through the motor cortex [105]. Clinical studies observed patients suffering from neurological disorders which caused ‘pathological laughter’ that is inappropriate, mood-independent, and uncontrollable laughter. However, it seems that pathological laughter originates from numerous pathological factors during brain disease rather than emerging from circumscribed lesions to particular brain sites [72]. Apparently, expressive laughter may be a phenomenon so complex that it cannot be assigned to any single brain area but depends on the temporal coordination of several areas constituting a functional network. This view receives partial support by recent brain imaging data which report a network incorporating left insula, bilateral amygdala and bilateral auditory cortices subserving the perception of human laughter [80]. Notably, activity in the right auditory cortices was stronger than in contralateral areas of the left hemisphere. According to the authors, this rightward asymmetry of supratemporal activity may be explained in terms of a right auditory cortex specialization for processing changes of pitch, timbre, and tonal patterns. While these authors used silence and human crying as baseline conditions in their study, we contrast human laughing with speech and nonvocal sounds to reveal a potential functional sensitivity of peri-sylvian regions.

In terms of phonetics, laughter and speech have essential acoustic underpinnings in common. First, speech and laughter share the same source of vocalization [75]. Like speech, laughter can also be described as a stream of acoustic cues which unfold in time. Typically, laugh bursts form a phrase-like sequence of repetitive laugh pulses that are initiated by an ingestion (see Methods section). Laughter involves the respiratory system and the vocal apparatus and is produced by changes in the vocal and sub-glottal, respiratory tracts as well as in the musculoskeletal system [19] that result in staccato-like syllabic vocalizations. Like speech, a sequence of laughter syllables can be combined into utterances. These utterances start with a vocalized inhalation and consist of a series of short laugh pulses with almost constant time intervals [66,76]. The reiteration of syllables with their specific segmental characteristics enables us to identify it as laughter. In contrast to speech, laughter contains less F0-variance. Laughter has a rich harmonic structure [74] and spectral qualities like formant frequencies can be evaluated [3]. Interestingly, both speech and laughter arise from laryngeal
modulations, modified to some degree by supralaryngeal activity, but in terms of laughter only to a small extent by articulation. Laughter also contains effects of one type of co-articulation at syllable-level between the aspiration onset and the following vowel (/h-a/), but it is more invariant than speech as vowel quality does not alter within laughing sequences. Differences between the two types of vocalization can be observed with regard to the organization of syllables contained in speech and laughing utterances. The isochronous reiteration of syllables in laughter is untypical for normal speech. To summarize, hearing speech requires the processing of rapid changes in frequency which reflect co-articulatory information between single segments, syllable, and word boundaries, while perceiving laughter involves the processing of rhythmical aspects and stereotyped changes in the fundamental frequency as melodic variation is more prominent in each laugh burst than in each speech syllable [79]. Notably, slow processing of prosodic and melodic modulations has been tied to right temporal lobe functions [56,63,64]. Here, we hypothesize that hearing human laughter is likely to preferentially drive right auditory fields as hearing laughter has formerly been observed to activate the right auditory cortex more strongly than contralateral areas [80]. Furthermore, when volunteers hear laughter, we may expect to find activity in the right anterior STS area as this region has been associated with nonspeech vocal sounds [6,7,26,99].

In summary, this study aimed to explore the hemodynamic responses within the peri-sylvian cortex to three different classes of auditory stimuli (human laughter, speech, and nonvocal sounds) while participants performed an auditory target detection task. The event-related design applied with the present study allowed us to investigate transient brain responses to single experimental stimuli without any confound by differential attentional levels as we used an auditory target detection which required the participants only to respond to specific target trials which did not enter the data analysis. In other words, volunteers were not explicitly required to indicate and classify speech, laughter, and nonvocal sounds, but to simply listen to the inflowing stimuli which was meant to factor out potential task effects.

In agreement with expertise provided so far, we hypothesize that hearing sentences activates anterolateral regions of the peri-sylvian cortex with the left hemisphere playing a prominent role as the left STG and the left IFG have been attributed to auditory sentence processing [21,22,31,61,63–65]. By regarding laughter and nonvocal sounds, we expect to find stronger bilateral or even rightward responses in the human auditory cortex since several recent studies pointed to a specialization of a right auditory cortical swathe for aspects of auditory processing outside the speech domain [42,89,104,110,114].

2. Materials and methods

2.1. Participants

12 neurologically healthy right-handed volunteers participated in this study (mean age 24 years old; range 20 to 32 years old; 6 females and 6 males). We obtained written consent from all participants. All volunteers were native speakers of German and had no hearing problem.

2.2. Stimuli and design

This study comprises 3 conditions and one target condition:

- Speech: short sentences containing a subject noun phrase and an intransitive verb such as ‘Peter sleeps’;
- Sounds: nonvocal sounds;
- Laughter: 10 female and 10 male laughing phrases;
- Target condition: reiterated sequence of non-sense syllables (‘lalala’).

2.3. Acoustic analyses

One representative sound file of each condition was subjected to acoustic analyses to demonstrate differential temporal and harmonic characteristics. The Praat software3 was used to do acoustic analyses of auditory events.

The wave form in Fig. 1A demonstrates that speech incorporates a continuous stream of syllables. Like speech, laughter also shows a syllable-like reiterated sequence of laugh bursts with an initial aspiration followed by a vowel /a/ and decreasing amplitude envelope as a function of time (Fig. 1B). Human vocal sounds such as speech and laughter contain more variation in the time signal due to typical speech characteristics such as co-articulation between single segments, syllable, and word boundaries and a decreasing amplitude envelope. The sound signal (Fig. 1C), however, shows less variance in the time and amplitude structures. On the other hand, both sounds and laughter show an equivalent distance between laugh bursts representing the temporal structure which differs from that for spoken languages.

Fig. 2 shows the wide band spectrograms (0–3 kHz) of speech (Fig. 2A), laughter (Fig. 2B), and sounds (Fig. 2C). The figure illustrates the resonance characteristics of the human vocal tract, that is, frequency bands with higher amplitude in periodic parts of the signals for speech and laughter. These frequency bands – the formants – are typical for human vocal production. These formants transitions reflect articulatory movements in the vocal tract which are more variable in speech than in laughter. In other words,

---

rapid changes in frequency reflect effects of co-articulation typical for speech and to a lesser extent for laughter. In contrast to human vocal sounds like speech and laughter, a less dynamic frequency spectrum can be observed for nonvocal sounds. Furthermore, sounds are composed of single frequency components without the typical speech like frequency and time variation.

To summarize, the acoustic observations of the material used in this study, the time wave form (Fig. 1) reflects temporal characteristics and differences between human vocal sounds and nonvocal sounds. In speech and laughter, segments and syllables are connected via co-articulation. Boundaries of syllables and even words can be detected. In nonvocal sounds, only the repetitive character of a re-iterant sound event is observable. The resonance characteristics, that is, the filtering properties of the resonance body of the human vocal tract for speech and laughter (Figs. 2A and B) reflect the presence of (at

![Fig. 1. Wave forms of acoustic stimuli. This figure shows time wave forms for (A) one speech event (‘Lotte pöbelt.’), (B) one laughing cycle, and (C) one sound event. Distances between auditory events reveal the rhythmical structure and normalized amplitude shows balanced intensity of auditory stimuli. See text for details.](image)

![Fig. 2. Wide band spectrogram of acoustic stimuli. This figure shows the wide band spectrogram (analysis width (s) = 0.005) generated from one speech event (A), one laughing cycle (B), and one sound event (C). Frequency changes in both the formant and harmonic structure are typical for vocal events (speech and laughter) and reflect articulatory dynamics in the vocal tract. See text for details.](image)

2.5. Task

It has been shown that actively calling a subject’s attention toward an auditory stimulus yields increased activation in auditory fields [51,111]. To avoid a confound between stimulus perception and attentional demands, we utilized a simple target detection task.

2.6. MR Imaging

MRI data were collected at 3 T 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 (echo-planar) images were obtained from 12 axial slices (6 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). The six inferior slices were positioned below the AC–PC plane and the remaining six slices extended dorsally. The whole range of slices covered all parts of the peri-sylvian cortex and extended dorsally to the supplementary motor area (SMA). After defining the slices’ position, a set of two-dimensional T1 weighted anatomical images (MDEFT sequence: TE = 20 ms, TR = 3750 ms, in-plane resolution 0.325 mm²) were collected in plane with the echo-planar images to align the functional images to the 3D-images [71,94]. A gradient-echo EPI sequence was used with a TE 30 ms, angle 90°, TR = 2000 ms. In a separate session, high-resolution whole-head 3D 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.

2.7. Data analysis

Data were analyzed with LIPSIA software [58]. During reconstruction of the functional data, the 5 (resp. 4) corresponding runs were concatenated into a single run. Data preparation proceeded as follows: slice-wise motion correction (time step 20 as reference), sinc interpolation in time (to correct for fMRI slice acquisition sequence); baseline correction (cut-off frequency of 1/80 Hz). To align the functional data slices onto a 3D 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-dimensional MDEFT and EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference data set. The resulting parameters, scaled to standard size, were then used to transform the functional slices using trilinear interpolation, so that the
Table 1

Laughter vs. Speech: in this table and in Table 2, results of direct comparison of laughter vs. speech and sounds are listed.

<table>
<thead>
<tr>
<th>Location</th>
<th>BA</th>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Z score</td>
<td>Cluster size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrCG</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PrCG</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HG/ROp/subCG</td>
<td>41</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PT</td>
<td>40</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>STP/HG/subCG</td>
<td>22</td>
<td>3.60</td>
<td>1218</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG (triangularis)/post IFS</td>
<td>9/45</td>
<td>–3.48</td>
<td>773</td>
</tr>
<tr>
<td>STG/STS</td>
<td>22/42/21</td>
<td>–4.86</td>
<td>11,714</td>
</tr>
<tr>
<td>Post MTG</td>
<td>39</td>
<td>–3.31</td>
<td>–</td>
</tr>
</tbody>
</table>

Z scores indicate the mean activation in a particular anatomical structure. Localization of coordinates correspond to the location of local maximal activation and is based on standard Talairach brain space. Distances are relative to the intercommissural (AC–PC) line in the horizontal (x), anterior–posterior (y), and vertical (z) directions. Functional activation was thresholded at \( |Z| \geq 3.09\) (uncorrected alpha-level 0.001). The table only lists activation clusters exceeding a minimal size of 225 mm\(^3\) (5 measured voxels).

\(^{a}\) Indicates that \( Z \) score refers to local maximum as this cluster is not distinct but overlaps with the cluster listed in the line above. Stereotactic coordinates are assigned to anatomical locations by using the ‘Automatic anatomical labeling’ tool [92] as implemented in the MRICRO software packages (http://www.psychology.nottingham.ac.uk/staff/cr1/mricro.html). Anatomical labels are abbreviated as follows: IFG = inferior frontal gyrus, IFS = inferior frontal sulcus, PrCG = precentral gyrus, subCG = subcentral gyrus, PaOp = parietal operculum, HG = Heschl’s gyrus, FOp = Frontal operculum, ROp = Rolandic operculum, PT = planum temporale, PPar = planum parietale, STG = superior temporal gyrus, STS = superior temporal sulcus, MTG = middle temporal gyrus, FG = fusiform gyrus, IOG = inferior occipital gyrus.

resulting functional slices were aligned with the stereotactic coordinate system. This linear normalization process was improved by a subsequent processing step that performs an additional non-linear normalization [91].

Statistical evaluation was based on a least-square estimation using the general linear model for serially auto correlated observations [2,14,32,109]. First, for each subject, statistical parametric maps were generated. The design matrix was generated with the standard hemodynamic response function, its first and second derivative and a response delay of 6 s. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 sec FWHM. Thereafter, contrast maps (i.e., estimates of the raw-score differences of the beta coefficients between specified conditions) were generated for each subject and the differences between conditions were calculated using the \( t \) statistic. Subsequently, \( t \) values were transformed into \( Z \) scores. As the individual functional data sets were all aligned to the same stereotactic reference space, a group analysis was subsequently performed. Group activations were calculated by the Gaussian test at corresponding voxels [14]. To protect against false positive activations, only regions with \( Z \) score greater than 3.1 (\( P < 0.001; \) uncorrected) and with a volume greater than 225 mm\(^3\) (5 measured voxels) were considered [15,27] (Tables 1 and 2).

We also computed an additional post hoc conjunction analysis to reveal the neural substrates of vocal timbre. For this purpose, we determined the significant clusters which demonstrated common activation in both the laughter vs. sounds and the speech vs. sounds contrast (Table 3).

Table 2

Laughter vs. Sounds: functional activation indicated separately for direct contrasts between Laughter and Sounds.

<table>
<thead>
<tr>
<th>Location</th>
<th>BA</th>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Z score</td>
<td>Cluster size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HG/STG</td>
<td>41/42/22</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fund STS</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FG</td>
<td>19</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

For explanations, see Table 1.
Furthermore, we performed a post hoc ‘region of interest’ (ROI) analysis which enabled us to compute condition-specific event-related time courses to uncover differential involvement of particular cortical sites. We collected BOLD signals from six bilateral ROIs along the peri-sylvian cortex. We selected these regions since they have been demonstrated to mediate cardinal speech and auditory functions, namely anterior superior temporal gyrus (LA aSTG: −58, −12, 1; RH aSTG: 55, −11, 2) \[11,30,45,82,84,85,88\], medial part of Heschl’s gyrus (LA mdHG: −43, −18, 9; RH mdHG: 40, −12, 5) \[42,80\], posterior STG (LA pSTG: −48, −40, 15; RH pSTG: 49, −34, 14) \[38,42,55,69,108\], planum parietale (LA Ppar: −54, −30, 25; RH Ppar: 56, −32, 26) \[33,40,49,55,60\], parietal operculum (RH PaOp: −40, −28, 23; RH PaOp: 40, −28, 26) \[33,40,49,55,73,78,90,108\], and subcentral gyrus (LA subCG: −49, −10, 12; RH subCG: 48, −11, 12) \[79\]. The signal was collected from the voxel with the highest regional activation in the average Zmap originating from separate contrasts for each single condition (laughter, speech, sounds) when compared to empty trials. The signal was averaged time-locked to the onset of each stimulus presentation and across all 12 subjects.

### 3. Results

**Fig. 3** displays the main fMRI results, the contrasts between Laughter vs. Speech, Laughter vs. Sounds, and the conjunction analysis Vocal timbre vs. Sounds\[^{5}\].

#### 3.1. Laughter vs. speech

**Fig. 3A** illustrates that several right peri-sylvian areas more strongly subserved laughter when compared to speech, namely the PT, HG/ROP, the subCG, and the ventrolateral precentral gyrus/sulcus (PrCG). Note that the medial strip of the supratemporal plane (STP) in the left hemisphere also shows stronger activity for laughter relative to speech. Hearing normal speech relative to laughter revealed considerable activity in the left peri-sylvian cortex along the superior temporal gyrus (lateral convexity) extending into the STS, STP, and in the pars triangularis of the frontal operculum (FOp). Furthermore, extra-sylvian activity was observed in the left posterior MTG. Stronger right hemisphere responses to speech relative to laughter were found in the middle part of the right STG/STS.

#### 3.2. Laughter vs. sounds

As apparent from **Fig. 3B**, stronger hemodynamic responses to laughter as compared to sounds were collected from right hemisphere sites, namely from the HG/STG, from the fund of the STS, and from the FG. Hearing sounds relative to laughter involves medial parts of the left and right Heschl’s gyrus (HG), the Ppar, the PT, the posterior insula, and the PaOp bilaterally. Furthermore, we observed activation in the left posterior MTG and in the left pars orbitalis of the IFG at the border to the anterior insula/posterior orbital gyrus\[^{6}\].

#### 3.3. Vocal timbre vs. nonvocal sounds

The post hoc conjunction analysis to identify regions which commonly activate during both speech and laughter processing revealed bilateral temporal and occipital sites with the right hemisphere demonstrating stronger involvement. As apparent from **Fig. 3C** and Table 3, hearing vocal sounds (speech and laughter) relative to nonvocal sounds more strongly activates bilateral sections of the anterior lateral STG. The right medial part of HG lining the posterior insula also shows starker engagement. Furthermore, hearing vocal timbre activated the left inferior occipital gyrus (IOG) as well as both the left and right fusiform gyrus (FG). Taken together, based on consideration of clusters’ size, we observed a considerable functional rightward asymmetry during perception of vocal timbre.

\[^{4}\] An additional Figure showing the position of ROIs projected on brain macroanatomy is available online (‘http://www.psychologie.unizh.ch/neuropsy/home_mmeyer/BRES-D-04-06566’).

\[^{5}\] Due to spatial restrictions we are not comprehensively discussing the Speech vs. Sounds contrast here. Brain scans, Tables, and results originating from the contrast Speech vs. Sounds are available at ‘http://www.psychologie.unizh.ch/neuropsy/home_mmeyer/BRES-D-04-06566’.

\[^{6}\] The latter cluster is depicted on a supplementary Figure available at (‘http://www.psychologie.unizh.ch/neuropsy/home_mmeyer/BRES-D-04-06566’).

### Table 3

Conjunction analysis [Laughter vs. Sounds ∩ Speech vs. Sounds]: clusters indicate regions which conjunctively activate during vocal timbre (speech and laughter vs. sounds)

<table>
<thead>
<tr>
<th>Location</th>
<th>BA</th>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Z score</td>
<td>Cluster size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>x    y    z</td>
</tr>
<tr>
<td>Medial HG</td>
<td>Speech vs. Sounds</td>
<td>41</td>
<td>−8</td>
</tr>
<tr>
<td>Anterior STG</td>
<td>Speech vs. Sounds</td>
<td>22</td>
<td>54</td>
</tr>
<tr>
<td>IOG</td>
<td>Speech vs. Sounds</td>
<td>19</td>
<td>378</td>
</tr>
<tr>
<td>FG</td>
<td>Speech vs. Sounds</td>
<td>37</td>
<td>135</td>
</tr>
</tbody>
</table>

For explanations, see Table 1.
3.4. ROI analyses

Data obtained from six bilateral ROIs for each of the three conditions evinces different local maxima in distinct regions of the peri-sylvian cortex (Figs. 4–6). Speech comprehension evoked the most considerable responses in the left temporal cortex, in particular in lateral parts of the anterior and posterior STG when compared to sounds and laughter. Left hemisphere dominance for speech was observed for temporal ROIs (aSTG, pSTG). A general rightward asymmetry of hemodynamic responses to laughter was found in four out of six ROIs (aSTG, pSTG, Ppar, subCG). The strongest responses to laughter relative to speech and sounds were collected from bilateral subCG with a salient peak in the right subCG. Notably, time courses for laughter overlapped with time courses for speech in eight out of twelve ROIs in both the left (mdHG, Ppar, PaOp, subCG) and the right hemisphere (aSTG, pSTG, Ppar, PaOp).

For sound processing, we consistently observed stronger RH activity in all ROIs (except mdHG). Compared to speech and laughter, hemodynamic responses to sounds were strongest in five out of six tempo-parietal ROIs (LH + RH Ppar, LH + RH PaOp, LH subCG), but only in two out of six supratemporal ROIs (LH + RH mdHG).

Summing up the major results, the data provide clear evidence that the three different conditions preferentially activated distinct areas. Hearing speech primarily involved areas in the lateral convexity of the left superior temporal region with the anterior and posterior STG obviously playing a prominent role. Hearing laughter preferentially recruited right hemisphere areas. Interestingly, areas in the transition zone between parietal auditory fields and motor cortex turned out to be considerably activated by laughter. Hearing sounds most strongly activated bilateral regions along the medial convexity of the posterior supratemporal plane (HG, PT, Ppar) up to the parietal operculum with the right hemisphere consistently producing stronger responses.

4. Discussion

This event-related fMRI study reveals distinct regions in the human auditory cortex during auditory perception of speech, vocal non-speech (laughter), and nonvocal sounds and indicates a functional division of peri-sylvian structures. Previous brain imaging studies have already been discussing a functional segregation of temporal regions during auditory processing but ignored consideration of vocal non-speech stimuli [9,11] or confounded vocal and non-vocal sounds.
which made it difficult to properly distinguish between peri-sylvian regions subserving speech and nonspeech perception. In order to address this issue more carefully, we presented three different types of auditory stimuli, namely sentential vocal speech, human laughter, and nonvocal sounds. All three types of auditory percepts evoked brain responses in the peri-sylvian cortex bilaterally with the level and spatial distribution of activation varying as a function of experimental condition.

The implications of our findings for auditory processing in distinct portions of the peri-sylvian cortex are discussed in turn.

4.1. Speech

Our data on perception of sentential speech are mostly confirmatory with published work in the field [29,47,84], so we refrain from elaborating on what has been comprehensively described elsewhere. In brief, our contrasts clearly show that perception of intelligible sentential speech particularly involves brain regions along the entire lateral STG and STS, and in the left lateral FOP. In general, this pattern of consistent brain activation is in agreement with a host of neuroimaging studies which observed bilateral recruitment of superior temporal areas while volunteers heard intelligible sentential speech [20,22,31,62,70,82,85].

More precisely, comprehending speech signals necessitates the decoding of information from differing linguistic domains, for example, meaning of words, thematic, and structural relations, as well as from simple sound-based signals comprised of multiple co-occurring frequencies called formants. Recent brain imaging studies identified distinct brain regions, especially in left peri-sylvian cortex subserving particular aspects of linguistic and non-linguistic aspects of speech comprehension [10,11,28,30,31,45]. In

Fig. 4. Averaged percent signal change collected from six bilateral regions of interest. The signal was collected from the voxel with the highest activation in the average Z map voxels. The signal was averaged time-locked to the onset of each stimulus presentation and across all 12 subjects. aSTG = anterior superior temporal gyrus, mdHG = medial Heschl’s gyrus, pSTG = posterior superior temporal gyrus, Ppar = planum parietale, PaOp = parietal operculum, subCG = subcentral gyrus.
other words, speech comprehension can be considered a hierarchical series of processing stages to translate sounds into meaning [22]. Both simple sound-based and abstract higher-level aspects of speech comprehension recruit cortical areas on the undulating surface of the STP as well as the insula and the deep frontal operculum [12,36,64] as well as areas stretching along the lateral convexity of the STG down to the temporal poles. Our ROI analyses uncovered two regions in the left superior temporal gyrus which have formerly been identified as subserving speech comprehension. Due to the present understanding, the posterior superior temporal gyrus supports the processing of rapidly changing acoustic cues which correspond to phonetically relevant acoustic features [55,73,108] while the anterior part maps acoustic-phonetic cues onto lexical and grammatical representations [82,88]. In summary, brain regions mediating higher-level aspects of speech being made up of complex sounds as well as meaningful semantic and even syntactic cues can be localised within left tempo-lo-lateral and predominantly anterior temporal and fronto-opercular cortex [29,73,81,83]. Based on the design of the present study, we are not able to ascribe any particular linguistic function to distinct subportions of the peri-sylvian cortex. However, all regions which exhibit stronger activity for sentences relative to laughter or sounds have been described elsewhere as playing a relevant role during auditory sentence comprehension [22,31,61].

4.2. Sounds

Even though the sound condition we presented in the current study was primarily considered a nonvocal control
condition, we discovered interesting and novel findings which are worth a detailed discussion. Processing nonvocal sounds involves posterior portions of the right and the left peri-sylvian region, namely the medial part of the primary auditory cortex, the adjacent PT, the Ppar, and the PaOp which all subserve auditory functions [1,38,40,43,49–55,81,87] and can be considered part of the auditory cortex [33,78,90]. Interestingly, in the context of the present study, the anterior parts of the STP did not reveal selective responses to nonvocal sounds, only the posterior structures including the PT and adjacent sections extending into the Ppar and the PaOp responded to the presentation of nonvocal sounds. The planum temporale has been recently described as a universal “computational hub” playing an integrative role in sound processing [37] and housing transient representation of temporally ordered sound sequences without distinguishing between words and other sounds [36]. Thus, the present data support a view which suggests the PT as governing intrinsic spectrotemporal features of sound [103] and which characterize the PT as responsive but not selective to speech as it is assumed to constitute an integrating mechanism mediating an ongoing updating between stored and present auditory information [69]. Alternatively, it should be mentioned that the posteromedial compartments of human auditory cortex have been associated with processing the location of sound sources [60,102,113]. A recent refinement of the model by Griffiths and Warren more specifically proposes that the PT accesses learned representations of temporarily stored sounds in higher cortical areas and gates spatial and object-related information [38,101]. Here, we only used monotonous nonvocal sounds which do not vary in

Fig. 6. Time course of mean event-related hemodynamic responses collected from three bilateral regions of interest in the inferior parietal lobe.
location and which are difficult to spontaneously associate with any environmental or meaningful origin. However, it cannot be ruled out that perception of nonvocal sounds automatically activated a neural ensemble which normally responds to changes in spatial and object-related cues to end up with a reasonable interpretation of nonvocal sounds. Corroborating evidence for our finding comes from a recent fMRI study which investigated brain regions involved in recognizing environmental sounds [57]. While the bilateral posterior MTG appears to preferentially subserve semantically related environmental sounds Lewis et al. [57] reported that bilateral regions in the posterior Sylvian cortex including the inferior parietal cortex responded to both environmental and unrecognizable backward sounds. Thus, our findings accord with recent studies which proposed a processing hierarchy in the human peri-sylvian regions with anterolateral regions such as lateral STG and PPol being preferentially activated by complex speech sounds while the posteriomedial portions of the Sylvian Fissure are more strongly attributed to the processing of the auditory structure of complex nonspeech sounds and simple speech sounds (phonemes) [38,55]. This suggestion overlaps with the observation that the temporoparietal cortex, particularly in the right hemisphere mediates abstract sound knowledge [25]. Our finding of bilateral activity covering the posterior peri-sylvian cortex, the tissue deeply buried in the posterior ascending Sylvian branch up to the parietal operculum may additionally speak to the existence of neurons sensitive to auditory stimulation in the inferior parietal lobe accommodating the Ppar and the PaOp. Little is known about the specific role of the Ppar and PaOp. Both Ppar and PaOp have been formerly attributed to auditory processing but have not yet been assigned to any specific function [40,49]. At least, the Ppar partly overlaps with the cytoarchitectonic area Tpt which has been classified as auditory cortex [33]. Thus, the present data add to the view that the posterior portions of the peri-sylvian region mediate processing of fine-grained, spectro-temporal, acoustic cues and can therefore be considered part of the auditory cortex in man [18] even though it is not possible to precisely identify the functional contribution of distinct subregions (PT, Ppar, PaOp) which brought on a signal increase for nonvocal sounds.

4.3. Laughter

Hearing laughter when compared to speech or sounds preferentially engages the right STG including the primary auditory cortex. This finding is in line with a recent fMRI study which also reported a stronger right STG involvement in processing human laughter [80]. As speech and laughter do not considerably differ in acoustic complexity, other issues must account for this finding. Numerous clinical studies pointed to a right hemisphere preference for affective moods [13,44]. However, contradictory evidence comes from a view which claims that the right hemisphere does not universally process affective moods [96]. It rather seems that the right hemisphere is more proficient in processing acoustic cues such as changes in pitch, amplitude, and frequency modulations if they load affective information. Vice versa, linguistically loaded changes in the same acoustic parameters may preferentially drive the left hemisphere. As laughter is an affective utterance, it is plausible that the right auditory cortex exhibits a starker involvement.

One may have assumed that laughter also may have evoked responses in other ‘emotionally driven’ system involving the amygdala, thalamic/hypo-, and subthalamic areas proposed by Wild et al. [105]. We believe that the laughter sequences we used in the present study were not sufficient to induce considerable emotional sensations for two reasons. First, the laughing samples were recorded from professional actors. After scanning, volunteers reported that they did not perceive the laughing samples as ‘contagious laughter’ in any case. However, Sander and Scheich [80] also used laughing recorded from professional actors as stimuli. Unlike our study, they observed clear brain responses in the amygdala to laughing and crying. Notably, Sander and Scheich used continuous streams of laughter which lasted ten times longer than the laughing sequences we applied in our study. Therefore, it is plausible that this difference in duration accounts for the differences in results.

Additionally, laughter produced activation in the right subcentral gyrus and precentral gyrus which has been described as part of the ‘voluntary’ system partly constituting the neural basis of expressive laughter in the brain. It is therefore conceivable that hearing laughter also excites brain regions which belong to the neural circuit subserving expressive laughter [79]. Analogously, brain regions which support the production of speech also appear to be involved in speech processing [24,48,64,106].

4.4. Vocal timbre—the potential role of the (right posterior) STG

The design of the present study also enabled us to perform a particular post hoc conjunction analysis to identify regions which commonly activate during both perception of speech and laughter, that is, human vocal timbre which could be considered the most prominent acoustic feature which distinguishes between vocal and nonvocal sounds. In the view of the present results, the right and left anterior lateral STG preferentially responded to stimuli made up of vocal timbre with the right hemisphere playing a considerably stronger role. Furthermore, the medial compartment of HG lining the posterior insular cortex also activated saliently during vocal timbre perception. This strong asymmetry conflicts with at least one recent PET study which concluded that temporal brain structures we observed to respond more strongly to vocal timbre (mid STG/HG) are functionally symmetrical [93].
However, this study by Tzourio-Mazoyer et al. did not explicitly investigate the significance vocal information may have for the right hemisphere. Corroborating evidence which lends credence to our view comes from a multitude of clinical and brain imaging studies. For decades, the study of neurologically impaired patients was the only source of information about the neural processing of voices. Evidently, our findings agree with the former neuropsychological research as these clinical observations suggest that lesions of the temporal lobe of either hemisphere may lead to deficits in the discrimination of unfamiliar voices [95,97,98]. Apparently, a lesion in the right hemisphere only leads to a deficit in the recognition of familiar voices [95]. Recently, the neural processing of voice information has also been particularly investigated by functional imaging experiments corroborating and complementing the clinical data. In a recent functional magnetic resonance imaging (fMRI) study [99], 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 superior temporal sulcus of the listener. In another series of studies [6,7], it was shown that temporal lobe areas in both hemispheres responded more strongly to human vocal timbre than to other vocal sounds, such as bells, dog barks, machine sounds with the right anterior polymodal STS being preferentially driven by human vocal information [4,8,36,56,100]. We also observed activation at the fund of posterior STS when volunteers heard vocal timbre relative to sounds. We presume that differences in baselines used in the different studies may account for the different STS locations.

Admittedly, the picture on vocal timbre perception remains fragmentary, but this post hoc conjunction analysis is at least in harmony with a host of clinical and neuroimaging studies which describe the superior temporal lobes as eminently relevant brain structures for the perception of human vocal information.

When compared to nonvocal sounds, vocal timbre produced activity in the left and right fusiform gyrus. Cross-modal mechanisms may account for this intriguing result. Previous observations in normal-hearing subjects and in patients with cochlear implants have established that visual cortex participates in speech processing [35], in particular when listening to meaningful sentences [34]. The precise location of activity in visual areas evoked by auditory stimuli varies across studies. Giraud et al. observed activity in the vicinity of the fusiform face area while cochlear-implant patients were listening to speech [34]. The authors suggest that patients probably translated auditory speech into facial expressions. Therefore, it is not unlikely that participants in the present study mentally associated a laughing or speaking face, while hearing speech and laughter. More corroborating evidence comes from an fMRI study which reports fusiform and lingual activity collected while participants heard sentences [99]. In the line of the authors, these results imply that participants translated auditory information into specific visual representations to better accomplish the task. Interestingly, the same cross-modal interaction works vice versa as it has been demonstrated that seeing speech recruits auditory areas [17,59]. Due to our results hearing nonvocal sounds did not evocate internal images and therefore did not produce activity in areas associated with imagery of faces.

Generally, the data allow us to draw the following picture sketching a functional division of the peri-sylvian cortex. Regions in the PAC and the peri-auditory cortex appear to be proficient at integrating complex spectro-temporal signals even though our data suggest a functional asymmetry varying with global and temporal information available in acoustic stimuli. Once the inflowing acoustic signal has been identified as spoken language, regions in the anterolateral and posteriolateral STG, particularly in the left hemisphere, initiate fine-grained processing at the phonetic, semantic, and syntactic level. If auditory input does not entail any speech, speech-like, or even vocal information regions in the posterior Sylvian region and the inferior parietal lobe become involved in acoustic analysis. Unlike vocal environmental sounds (dog barks, baby cries), the nonvocal sounds used in the present design cannot easily be associated with any meaning. It is therefore plausible to assume that enlarged activity for sounds reflect the search for a sensible interpretation. Recent observations underline this notion of the posterior peri-sylvian cortex as being highly sensitive to all categories of meaningful and meaningless complex sounds [87].

5. Conclusion

In conclusion, we have delineated distinct regions in the human peri-sylvian cortex which appear to preferentially support the processing of human laughter, speech, and nonvocal sounds. Hearing human laughter preferentially involves auditory and somatosensory fields primarily in the right hemisphere. Our findings deliver novel insight in that it appears that hearing laughter activates the same areas which have formerly been attributed to expressive laughter. Hearing spoken sentences activates left anterior and posterior lateral temporal and left inferior frontal regions which is confirmatory with numerous former findings. Hearing nonvocal sounds most strongly recruits bilateral areas in the medial portion of Heschl’s gyrus and at the medial wall of the posterior Sylvian fissure (planum parietale) even including the inferior parietal lobe (parietal operculum) and therefore adds functional evidence to the ongoing debate on the extension of auditory fields in the human brain. A post hoc conjunction analysis to identify regions which commonly activate during both speech and laughter revealed a functional rightward preference of supratemporal sites for human vocal timbre. In summary, the data add to recent work which provided evidence for a hierarchical processing system in the peri-sylvian cortex.
with the posteriomedial region mediating acoustic properties of nonvocal sounds and anteriolateral parts of the STG and STS supporting the perception of human vocal timbre.

Acknowledgments

The authors wish to thank Lutz Jancke, Adam McNamara, and two anonymous reviewers for helpful comments on the manuscript.

During the preparation of the manuscript, Martin Meyer was supported by Schweizer National Fonds (Swiss National Foundation) SNF 46234101, ‘Short-term and long-term plasticity in the auditory system’. This study was also supported by the project AL 357/2-1 by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) as well as by the Human Frontier Science Program (HFSP RGP5300/2002-C102) awarded to Kai Alter.

References


