Spectrotemporal features of the auditory cortex

Langers, DRM; Backes, WH; van Dijk, P

Published in:
Neuroimage

DOI:
10.1016/S1053-8119(03)00258-1

Publication date:
2003

ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 20 (2003) 265–275

www.elsevier.com/locate/ynimg

Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples

Dave R.M. Langers,[a,b,c,*] Walter H. Backes,[a,b] and Pim van Dijk[b,d]

[a] *Department of Radiology, Maastricht University Hospital, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands*
[b] *Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
[c] *Institute for Brain and Behaviour, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*
[d] *Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Hospital, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands*

Received 7 October 2002; revised 5 February 2003; accepted 29 April 2003

Abstract

Functional MRI was used to investigate the characteristics of the human cerebral response to dynamic ripples. Dynamic ripples are sound stimuli containing regular spectrotemporal modulations, which are of major importance in speech processing; however, in contrast to speech, dynamic ripples can be characterized fully by a limited number of parameters. Extensive activation consisting of multiple separate regions was found bilaterally in the auditory cortex, particularly along the Heschl's gyri. This agrees with the presence of a structural cortical subdivision into functional fields. The level and the extent of activation were measured and correlated highly ($R^2 = 0.97$). Both measures depended strongly on the spectral density, temporal frequency, and amplitude of the modulations and matched the perceptual discernibility of the spectrotemporal modulations. The largest responses occurred for parameter values near the optimal human sensitivity. The drift direction of the modulations did not influence the activation. No quantitative differences were found between the two hemispheres. Average brain activation levels proved to be separable with regard to the spectral density and temporal frequency of the modulations. Topographic mappings of the modulation density and frequency onto the cortical surface were shown, approximately in posterolateral-to-anteromedial and lateral-to-medial directions, respectively. Posterolateral regions were most sensitive to spectrotemporal features at a scale similar to phonemes. Anteromedial regions, however, were also relatively sensitive to smaller scale acoustic features. This spatially dependent sensitivity suggests a functional topographic and hierarchical organization of the auditory cortex.

Keywords: Functional magnetic resonance imaging (fMRI); Human cerebral cortex; Auditory cortex; Dynamic ripples; Topographic organization

Introduction

In this study we focused on sound properties related to speech, the processing of which obviously is an important function of the human brain (Belin et al., 2000, 2002; Jäncke et al., 2002; Wong et al., 2002; Wise et al., 2001; Scott et al., 2000). In systematic studies, the structural complexity of vocal sounds is a significant problem when defining their exact properties (Hickok and Poeppel, 2000). Therefore, simpler sound signals have been derived that can be defined with mathematical exactitude, but that still contain some essential characteristics of vocal sounds. An important category of such signals comprises modulated noise. The most basic of these modulated noise signals contain either temporal amplitude modulations (Giraud et al., 2000; Wakefield and Viemeister, 1990) or static spectral modulations, also known as rippled noise (Supin et al., 1999; Calhoun and Schreiner, 1998; Versnel and Shamma, 1998). However, modulations in the spectrum of speech are dynamic. Similar dynamic behavior can be incorporated into modulated noise by combining temporal and spectral modulations, resulting in a spectral ripple drifting in time along the frequency axis (Fig. 1). Such signals have a similar

* Corresponding author. Maastricht University Hospital, Department of Radiology, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands. Fax: +31-43-38-76-909.
E-mail address: dlan@rdia.azm.nl (D.R.M. Langers).

1053-8119/03/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1053-8119(03)00258-1

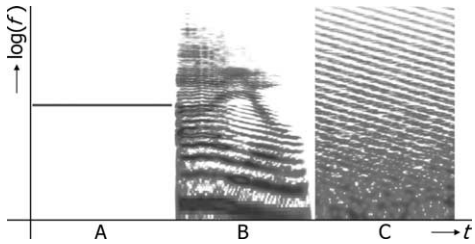


Fig. 1. A spectrogram showing the characteristics of a pure and continuous tone (A), a fragment of speech (B, vowel transitions in the Dutch acronym "aio"), and a dynamic ripple with downward spectral drift (C). The shading indicates the presence of spectral power as a function of frequency vs time. Note the similarities between speech and dynamic ripples on an acoustic and phonetic level, both containing spectrotemporal modulations.

spectrotemporal envelope as speech and can model formant transitions between vowels and changes in pitch due to intonation. However, the resemblances between dynamic ripples and vocal sounds are limited to acoustic and phonetic properties. Dynamic ripples are not associated with speech on a phonological or linguistical level, involving the recognition of syllables or the comprehension of words.

Dynamic ripples can be characterized by four parameters (Fig. 2): the spectral modulation density Ω (in cycles per octave, c/o), the temporal modulation frequency ω (in cycles per second, c/s), the modulation amplitude A (%), and the drift direction d (downward or upward). Any sound pattern can be regarded as a superposition of such dynamic ripples, since they form a complete orthonormal basis of functions. In electrophysiological animal studies, dynamic ripples play a useful role in characterizing neuronal spectrotemporal response fields (Miller et al., 2002; Depireux et al., 2001; Klein et al., 2000; Kowalski et al., 1996). Selective neuronal responses to the speed and direction of spectral changes have been reported using a range of stimuli (Liang et al., 2002; deCharms et al., 1998; Tian and Rauschecker, 1998), as well as corresponding topographic mappings on the brain's cortical surface (Versnel et al., 1995; Mendelson et al., 1993; Shamma et al., 1993). However, in humans electrophysiological experiments are unfeasible due to the invasive nature of this technique. Psychoacoustical experiments have been performed to measure thresholds for the perceptual detection of spectrotemporal modulations; near an optimum at $\Omega = 0.5$ c/o and $\omega = 4$ c/s, dynamic ripples with amplitudes A as low as 5% can be discerned from noise without modulations (Chi et al., 1999).

We used functional MRI (fMRI), which is a volume imaging technique that is suitable for humans, with the additional advantage of a high spatial resolution. Although the poor time resolution prevented us from directly detecting any synchronization of neuronal responses with the temporal sound modulations, sustained changes in overall activity due to selective neural activation were measurable (Seifritz et al., 2002). Given that the presence of tonotopy in the auditory cortex has previously been demonstrated using fMRI (Engelien et al., 2002; Le et al., 2001; Talavage et al., 2000; Bilecen et al., 1998; Wessinger et al., 1997), it may also be able to reveal any topographic organization associated with dynamic ripples. In contrast to animal studies, the activation of the human auditory cortex in response to dynamic ripples has not yet been studied systematically; in this paper we investigate the magnitude and spatial distribution of the cortical activation as a function of the previously mentioned dynamic ripple characteristics.

Materials and Methods

Subjects

Nine healthy subjects with normal hearing were recruited (eight male, one female; aged 23–59 years, mean 37 years) and gave written informed consent to participate in this study. Of all subjects, four were right-handed, three left-handed, and two ambidextrous (Oldfield, 1971). Subjects received an explanation and a short demonstration beforehand, but were not trained to the task.

Stimulus paradigm

A sound generation setup was used to present the auditory stimuli. Each stimulus presentation consisted of a series of two 3.25-s dynamic ripples (Fig. 3). The first sound consisted of a reference stimulus with certain parameters (Ω, ω, d) and $A = 60\%$. This was immediately followed by a second stimulus with equal parameters (Ω, ω, d), but A equal to 30 or 100%. Values used for Ω were $\frac{1}{4}$, 1, and 4 c/o; for ω 2, 8, and 32 c/s; d was taken downward or upward. All possible combinations of Ω, ω, A, and d were included in the experiment, giving a total of 36 dynamic ripple stimulus series. In addition, 4 identical stimulus series were added that consisted of two consecutive noise stimuli without any modulations. Effectively this condition corresponds to $A = 0\%$. These noise stimulus series were used as the baseline condition in the subsequent image analysis. The experiment

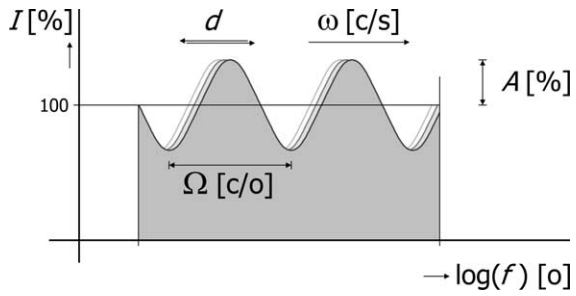


Fig. 2. The spectrum of a dynamic ripple is based on pink noise, but contains a modulation that drifts in time along the logarithmic frequency axis in either direction d. The modulation density Ω equals the number of spectral peaks present in a frequency band of one octave. The modulation frequency ω is equal to the number of spectral peaks that passes per second. The modulation amplitude A is expressed relative to the noise level.

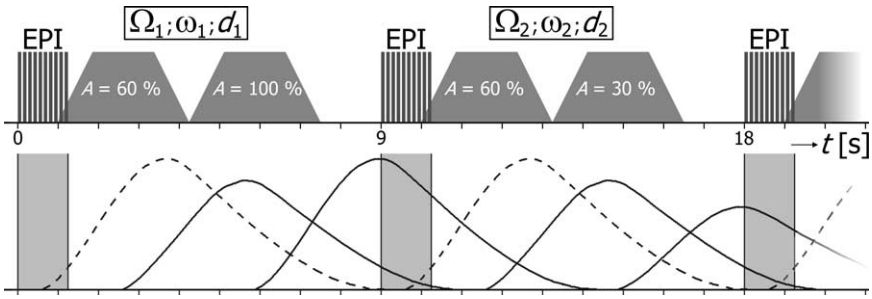


Fig. 3. A schematic representation of the acquisition and stimulus paradigms. Echo-planar imaging (EPI) scans were spaced 9 s; the silent periods in between were used to consecutively deliver two similar dynamic ripple stimuli, only slightly differing in modulation amplitude (*A*). Below, the courses of the hemodynamic responses are outlined. The dashed line represents the response to the scanner noise.

was performed in six blocks, each containing the total of 40 stimulus series in a different random order.

Task

Subjects were asked to compare both consecutively delivered auditory stimuli in each series and to press one of two fiber optic button devices to indicate whether the perceived modulation amplitude of the second stimulus was larger or smaller than the first. The main purpose of this two-alternative forced-choice task was to let subjects focus actively on the auditory stimuli. The results were used only to verify that the subjects understood the task and that they were able to discern the spectrotemporal modulations. Given the fact that they judged 85% of all stimulus series correctly, this was indeed the case.

Stimulus generation

Stimuli were generated using a real-time signal processing system (System II 16-bit DA converter, Tucker Davis Technologies) at a 32.8-kHz sampling rate. The baseline stimulus was computed by inverse Fourier transformation of a 2^{15}-point complex array z_k according to

$$z_k = 1/\sqrt{|k|} \cdot \exp(\mathrm{i}\varphi_k), \quad (1)$$

with quasi-random phases φ_k in the range $-\pi < \varphi_k \leq \pi$. The amplitude spectrum defined in Eq. 1 ensures that the logarithmic density of the spectral power $|z_k|^2$ is constant (i.e., pink noise). A spectrotemporal modulation was added by transforming the complex spectrum z_k to obtain a spectrum s_k according to

$$s_k = z_k + \frac{1}{2} A \cdot z_{k+\omega} \cdot \exp(-2\pi\mathrm{i} \cdot {}^2\log(|k + \omega|) \cdot \Omega) + \frac{1}{2} \mathrm{A} \cdot z_{k-\omega} \cdot \exp(2\pi\mathrm{i} \cdot {}^2\log(|k - \omega|) \cdot \Omega). \quad (2)$$

The drift direction *d* is controlled by choosing the sign of ω and Ω. The ^{2}log functions in Eq. 2 result in a constant number of modulations per octave. Because the duration of the waveform obtained after inverse Fourier transformation (1 s) is a multiple of the period of all dynamic ripples that were used, the waveforms could be repeated periodically without introducing artifacts. Signals were scaled to obtain equal intensity and loudness and windowed with an arbitrary rise/fall time (100 ms) to obtain the stimuli.

The stimuli were delivered by means of headphones consisting of piezoelectric speakers built into a noise protection headset, with a lower cutoff frequency of 0.2 kHz. In addition to the headphones, subjects wore earplugs, resulting together in 40-dB damping of the scanner acoustic noise of 105 dB peak sound pressure level (SPL). The auditory stimuli were presented binaurally at 80 dB SPL and were damped by 20 dB by the earplugs alone. The spectral power at different frequencies in the range 0.2–16.3 kHz varied smoothly within 10 dB.

Imaging

Subjects were placed in a supine position in the bore of a 1.5 T MRI system (Philips Intera, Philips Medical Systems), which was equipped with a standard quadrupolar head receiver coil. A 3D T_1-weighted fast-field echo scan with 50 contiguous sagittal slices was acquired for anatomical orientation (TR 9.85 ms; TE 3.5 ms; flip angle 15°; matrix 256 × 256; voxel size 0.94 × 0.94 × 3.0 mm^3). On these high-resolution images an imaging plane was positioned parallel to the Sylvian fissure of both hemispheres, containing the superior temporal gyrus of both temporal lobes. An additional 3D T_1-weighted fast-field echo scan (TR 10.37 ms; TE 3.5 ms; flip angle 15°; matrix 256 × 256; voxel size 1.25 × 1.25 × 1.75 mm^3) was made of 20 slices parallel to this imaging plane in order to serve as an anatomical reference for the subsequent functional scans. The functional scans were made using a scan volume identically oriented to the anatomical reference scan. These acquisitions consisted of a single-shot T_2^*-sensitive echo-planar imaging (EPI) sequence over 10 slices (TR 9 s; TE 50 ms; flip angle 90°; matrix 96 × 96; voxel size 2.5 × 2.5 × 3.5 mm^3). For each subject the functional session consisted of six blocks of 40 acquisitions each; all dynamic ripple con-

ditions were acquired 6 times in total; the baseline condition was acquired 24 times.

A considerable disadvantage of fMRI is the influence of acoustic noise produced by the scanner itself, evoking auditory responses that interfere with the responses to the stimuli being presented. This problem was reduced by using a sparse clustered volume scanning technique (Amaro et al., 2002; Le et al., 2001; Yang et al., 2000; Edmister et al., 1999; Hall et al., 1999; Talavage et al., 1999). The functional acquisitions were performed at a rate of one per 9.0 s and took 1.25 s per image volume, leaving 7.75 s of scanner silence between acquisitions. An electronic trigger signal was used to start the presentation of the 6.5-s stimulus series precisely 1.0 s after the start of an acquisition (Fig. 3). The 9-s acquisition spacing allowed the hemodynamic response to the scanner noise to be largely faded at the time the next acquisition took place (Backes and van Dijk, 2002; Belin et al., 1999). Insofar as there was a remaining contribution of scanner noise it was the same in all acquisitions, and thus it was assumed to cancel in the subsequent image subtraction analysis. The immediate influence of the scanner noise on the measurement itself was negligible because all slices were acquired within less than 2s (Talavage et al., 1999).

Data analysis

In the data analysis we made use of the MatLab environment (The MathWorks Inc.) and routines from the SPM99 software package (Wellcome Department of Cognitive Neurology). The functional images were motion-corrected and spatially smoothed using a 5-mm Gaussian kernel. Images were thresholded to omit voxels outside the brain. Areas responding to the auditory stimulation were identified by analyzing the data of each of the voxels individually. We set up a general linear regression model containing 36 conditions representing all different dynamic ripple parameter combinations. To correct for drifts of the scanner signal, the baseline was modeled by a first order polynomial for each of the six acquisition blocks separately. Because the acquisitions were spaced at as much as 9 s, and because the order of the presented series of stimuli was randomized in each block of acquisitions, we assumed that consecutive acquisitions were uncorrelated to simplify the analysis. We performed *t* statistics for contrasts between subsets of dynamic ripple conditions and the baseline condition to obtain statistical parametric maps (SPMs). These subsets were made to include all conditions having the value of one dynamic ripple parameter in common. For instance, all dynamic ripples with $\Omega = \frac{1}{4}$ c/o were contrasted together against the baseline condition to obtain one SPM. With the values mentioned for the parameters Ω, ω, A, and d, this leads to a total of 10 contrasts. Furthermore, one SPM was constructed using the contrast of all dynamic ripple conditions against the baseline condition. Within the SPM formalism, the contrast vector was constructed such that the conditions with $A = 100\%$ were weighted twice as strongly as conditions with $A = 30\%$.

Two different measures of the cerebral activation were employed. First, we defined the activation extent (AE) as the number of voxels in the brain becoming significantly active, using a threshold of $P < 0.001$. The AE was determined as a function of each dynamic ripple parameter separately. For each of the 10 contrasts corresponding with a certain value of a dynamic ripple parameter, the significantly activated voxels in the brain were counted in the respective SPM, without constraints on location or cluster size. Second, the activation level (AL) was defined as the percentage of signal change due to the hemodynamic response, averaged over all significantly activated voxels in the brain. In this calculation, the contrast of all dynamic ripple conditions against the baseline condition was used to determine one single set of voxels that significantly responded in general. For each of the 36 dynamic ripple conditions the percentage of signal change caused by the hemodynamic response was averaged over this set of voxels to obtain AL. In this way we were able to determine AL as a function of all possible combinations of parameter values.

Separability

Both in electrophysiological (Kowalski et al., 1996) and in psychoacoustical (Chi et al., 1999) experiments it has been found that the response to dynamic ripples is separable with regard to Ω and ω. This means that there is no mutual dependence of these parameter characteristics, and the overall response is simply a product of the response characteristics of both of these modulation parameters:

$$R(\Omega, \omega) = R_{\Omega}(\Omega) \cdot R_{\omega}(\omega). \tag{3}$$

To investigate the separability of AL in our experiment, the method of singular value decomposition (SVD) was applied. In order to use this method the data were shaped into a matrix **M,** which was subsequently factored as $\mathbf{M} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^{\dagger}$. In this expression $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1 \ldots \lambda_n$ in decreasing order, and **U** and **V** are unitary matrices containing the corresponding eigenvectors in both dimensional directions of the matrix. The degree of separability α_{SVD} is defined as

$$\alpha_{\mathrm{SVD}} = \lambda_1^2 / (\sum_i \lambda_i^2). \tag{4}$$

This equals the proportion of **M**'s total power which is accounted for by its best separable approximation (Depireux et al., 2001).

Topographic mappings

The presence of any topographic organization in the auditory cortex with respect to the parameters Ω, ω, A, and d was studied by determining which parameter values were

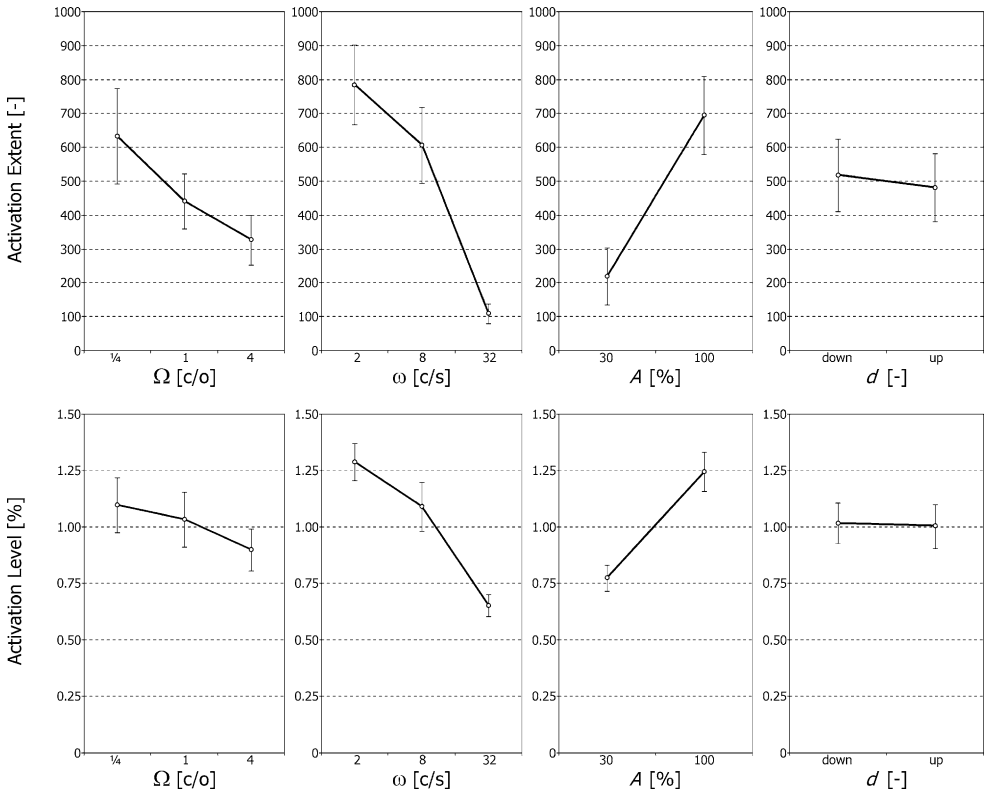


Fig. 4. The dependency of the activation extent (top panels) and mean activation level (bottom panels) on the dynamic ripple parameters Ω, *ω*, *A*, and *d*. Results were averaged over all subjects, with error bars visualizing the standard error of the mean. The changes in activation as a function of the parameters are all significant, except for those in *d*.

able to elicit the largest response in individual voxels. Because the brain's overall sensitivity to the presence of dynamic modulations strongly depends on the values of these parameters, a normalization step was carried out first. For each dynamic ripple parameter, the median response level of all significantly activated voxels was calculated as a function of the parameter value. Hereafter, the measured response levels of individual voxels were divided by this median brain response level to obtain a normalized response level. In each voxel, the parameter value resulting in the largest normalized response was determined for each separate parameter. In order to determine the presence of a topographic gradient, a linear function

$$f(x, y, z) = a \cdot x + b \cdot y + c \tag{5}$$

of the lateral-to-medial and posterior-to-anterior coordinates (*x* and *y*) was fitted to the normalized response patterns. As can be seen in Eq. 5, the inferior-to-superior coordinate *(z)* was not included in the fit, because the superior temporal surface hardly extends in that direction. For each activated voxel, *f* refers to the parameter value which gave the largest normalized response. For example, when investigating the topography of *ω*, discrete values equal to 1, 2, or 3 were fitted for *f*, indicating that the corresponding voxel gave a largest normalized response for $\omega = 2$, 8, or 32 c/s, respectively. Fits were carried out for the largest cluster of significantly activated voxels in the left and in the right auditory cortices separately. The gradient vectors *(a, b)* of these fitted functions were used to indicate the direction of the topographic mapping of the parameter (e.g., *ω*) onto the cortex and to determine its significance ($P < 0.05$).

Results

Activation extent and activation level

The cerebral activation caused by the spectrotemporal modulations in dynamic ripples depended on the parameters of the stimulus. Fig. 4 shows the results with respect to the AE, averaged over all subjects. We found that AE is maximal for low modulation densities ($\Omega = \frac{1}{4}$ c/o) and small modulation frequencies ($\omega = 2$ c/s) and decreases monotonically for both Ω and *ω*. With regard to the modulation amplitude we found that AE increased with *A*. In contrast, we did not find any significant changes in AE with the drift direction *d*. In Table 1 we computed the ratio of AE for the extreme values of the relevant parameters, with *P* values indicating the level of significance of the differences. Similar results were found with respect to the AL. Again, AL decreased with increasing values of Ω and *ω* and increased with *A*. The relative differences were smaller compared to AE, but still

Table 1
A comparison of the activation for different values of various dynamic ripple parameters, as well as for the left and right hemispheres

Comparison	Activation extent (AE)		Activation level (AL)	
	Ratio	*P* value	Ratio	*P* value
Ω: $\frac{1}{4}$ c/o vs 4 c/o	1.933	0.05	1.255	0.06
ω: 2 c/s vs 32 c/s	19.95	0.03	2.049	0.03
A: 100% vs 30%	3.557	0.03	1.647	0.03
d: Down vs up	1.124	0.5	1.015	0.8
Hemisphere: right vs left	1.072	0.9	1.024	0.9

Note. The *P* values were determined by the Wilcoxon matched-pairs signed-ranks test, after which Hochberg correction for multiple comparisons was applied (Hochberg, 1988), and indicate whether the values for AE as well as AL for the given parameter values differ significantly.

significant. Again there was no dependency on *d*. To confirm the similarity between AE and AL, the correlation coefficient between the subject averages of AE and AL for the various parameter values was computed. The result amounted to $R^2 = 0.97$, indicating a strong linear correlation between AE and AL ($P < 10^{-7}$).

We applied the SVD method to analyze the separability of the AL data with regard to Ω and ω. To obtain a matrix of values depending on Ω and ω only, the AL's were averaged over *A* and *d*. For the individual subjects, the degree of separability α_{SVD} was in the range 0.974–0.996, with a mean of 0.989. When the data were averaged over all subjects before the separability analysis, α_{SVD} equaled 0.998, indicating almost perfect separability.

Additionally, we compared AE as well as AL in the left and right hemispheres separately. We found no significant quantitative hemispherical difference for any dynamic ripple stimulus or any set of stimuli (Table 1).

Activation pattern

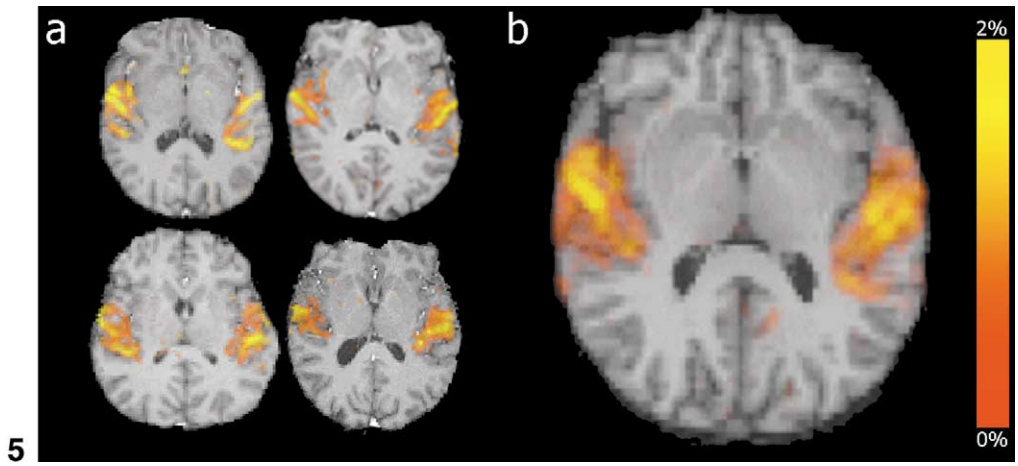
To investigate the spatial distribution of cerebral activation, the signal change from baseline in response to the set of all dynamic ripple conditions was calculated for each activated voxel. A few representative images are given in Fig. 5a. The anatomical T_1-weighted background of these images was extracted from the slice in the scan volume that contained the largest mean activation, and it therefore shows the anatomical structures that correspond with the main regions of activation. The maximal activation level over all 10 slices was color coded and projected perpendicularly onto this anatomical plane. In practice, the activation was largely confined to about 3 slices in the temporal lobe.

For each subject substantial regions of activation could be discerned bilaterally, with an organization into multiple separate regions. These regions mainly extend in the posteromedial-to-anterolateral direction, approximately parallel to the Heschl's gyri. Generally at least two main regions could be identified in each hemisphere, stretching out along the borders of Heschl's gyrus. Usually more regions were observed, sometimes located anteriorly in the planum polare (in three subjects) but more often posteriorly in the planum temporale (in five subjects). However, there is considerable variability in the pattern of activation between subjects, and within subjects there is quite some variability between hemispheres. Fig. 5b shows the signal change averaged over subjects in Talairach coordinate space. Two separate regions of activation extending along Heschl's gyrus are clearly visible in both hemispheres, while in the auditory cortex as a whole activation is present on a lower average level.

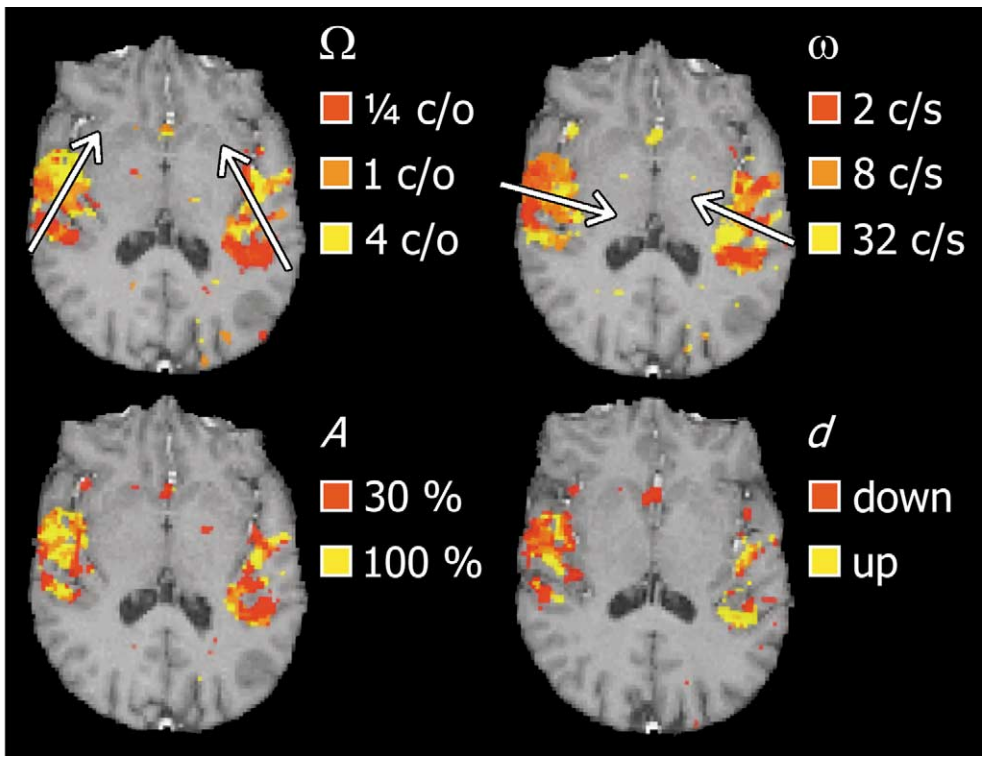
Similar regression analyses using SPMs were carried out to identify regions in the auditory cortex where the presence of dynamic modulations decreased the baseline activation caused by noise without modulations. No regions with significant signal reductions were found.

Topographic organization

In Fig. 6 the parameter values to which each voxel responded best after normalization to correct for differences in sensitivity of the whole brain are illustrated. In this figure the optimal parameter values for the voxels that showed significant activation were color coded, as indicated in the legends, and projected perpendicularly onto the anatomical slice with the highest number of activated voxels. The top left panel of the figure illustrates that the posterior (or posterolateral) part of the auditory cortex is more responsive to dynamic ripples with low spectral densities, whereas the anterior (or anteromedial) part responds relatively well to high values of Ω. This is visible in the form of a yellow-to-red color gradient. This feature was noted by three independent observers in five of nine subjects; in the other four subjects this pattern was visible retrospectively. To quantify this observation, the gradient of the 2D linear fit to the spatial distribution of optimal Ω values was calculated. The magnitude of this gradient was significantly different from 0 for six subjects (for both hemispheres). Its average direction is indicated in the figure by the white arrows and deviated from the posterior-to-anterior axis by approximately 30 $\pm$ 50° (left hemisphere) and 30 $\pm$ 20° (right) toward the lateral-to-medial direction (mean $\pm$ SD). Furthermore, it was observed in six of nine subjects that the activated voxels in a lateral position generally responded best to low modulation frequencies, while voxels located more medially responded relatively well to higher ω. Quantitatively, the magnitude of the gradient was significant in five (left) and six (right) subjects. The average direction deviated approximately 65 $\pm$ 60° (left) and 105 $\pm$ 30° (right) from the anterior-to-posterior axis, as indicated in Fig. 6. The subjects that showed less clear parameter mappings for both Ω and ω were precisely those that showed the smallest amount of activation. Both topographic mappings stretch out over the entire region of activation in the auditory cortex, which amounts to approximately 4 cm in the pos-



5



6

Fig. 5. (a) The spatial distribution of the signal change from baseline level for four representative subjects, indicated by the color coding. (b) The signal change, mapped into Talairach space and averaged over subjects, overlaid on a canonical brain. The activation is concentrated in the auditory cortex and shows a bilateral parcellation into multiple regions. All images are shown in radiological coordinates (the left hemisphere is on the right).

Fig. 6. The spatial distribution of the optimal modulation density Ω, modulation frequency ω, modulation amplitude A, and drift direction d, respectively. The colors represent the parameter values that lead to an optimal response in a voxel, after normalization to correct for differences in sensitivity of the whole brain. When multiple voxels project to the same pixel, the corresponding colors are mixed. Ω and ω show a topographic mapping on the auditory cortex in posterolateral-to-anteromedial and lateral-to-medial directions, respectively, which is visible as a red-to-yellow color gradient. Results are shown for the significantly activated voxels in one subject that showed both mappings. The white arrows indicate the direction of the gradient averaged over multiple subjects. The images concerning A and d show a smoothed random distribution of optimal sensitivity, and no representative gradient is present. All images are shown in radiological coordinates (the left hemisphere is on the right).

terolateral-to-anteromedial direction for the dynamic ripple density mapping and to 3 cm for the lateral-to-medial mapping of dynamic ripple frequency.

With regard to A and d, the observers found no systematic evidence of a topographic organization subjectively. For these two parameters the normalized optimal response appeared to show a smoothed random distribution. Still, up to three subjects quantitatively showed a gradient that differed from 0. However, the directions of these gradients varied greatly between subjects (with spreads in excess of 90° in both directions), and no representative average could be calculated.

Discussion

Activation dependencies

We considered two quantitative measures, the extent (AE) and the level (AL) of activation. In principle it is possible for these two measures to show different dependencies on the dynamic ripple parameter values. After all, large volumes that show marginally significant activation might well be imagined, as well as confined regions with high activation levels. Nevertheless, our data show an almost perfect correspondence between both measures ($R^2 = 0.97$); AE shows the same general trends as AL. This suggests that at least in this experiment both are adequate measures of cerebral activation. The origin of the stimulus dependence of AE is not obvious, but has also been shown in other experiments, e.g., for stimulus loudness (Brechmann et al., 2002; Hart et al., 2002; Lockwood et al., 1999; Mohr et al., 1999; Jäncke et al., 1998). It may partly be related to physiological phenomena, such as the recruitment of a larger neuronal population, or to an increased response contribution from draining veins and also to the spatial smoothing of the data (Hall et al., 2001). In the following discussion we refer to both measures simultaneously.

Our results show that the perception of dynamic ripples is accompanied with significant amounts of bilateral activation in an extensive part of the superior surface of the temporal lobes. The activation differs considerably as a function of the parameters Ω and ω, although the spatial locations of maximal activation stay approximately the same. Significantly fewer voxels reach the activation threshold as Ω and ω take on larger values. Our findings indicate that the human auditory processing on a cortical level mainly focuses on spectrotemporal modulations that are quite sparse in the spectral sense ($\Omega = \frac{1}{4}$ c/o) and relatively slow in the temporal sense ($\omega = 2$ c/s). This agrees with the psychoacoustical data that show that the sensitivity in humans to dynamic ripples with these parameters is close to the optimum (Chi et al., 1999). These values also fall within the range of modulations that are present in speech and correspond best with formant transitions. For even lower values of Ω and ω, we would expect the brain's sensitivity to the presence of spectrotemporal modulations to diminish again.

Because we observed a similar Ω dependence of the cortical activation regardless of ω, and a similar ω dependence regardless of Ω, we reasoned that the activation might be separable with regard to Ω and ω. To confirm this impression we applied the SVD method to our AL data. The high degree of separability proved that the deviations from complete separability were negligable. Despite the limited amount of data, this is an important finding because it enables one to determine the brain response to all possible parameter combinations by measuring only the dependencies on Ω and ω separately, holding the other parameter constant. However, from our data this conclusion is only justified for average responses in the brain as a whole, or at least large regions containing a multitude of neurons. We cannot extrapolate this to single neuron response characteristics.

Activation pattern

The large extent of activation shows that the processing of the spectrotemporal dynamics of sound is not confined to one specialized region in the auditory cortex. Instead numerous regions of activation were shown to exist. These are spread over the entire auditory cortex, mainly along Heschl's gyrus, but also extend posterolaterally onto the planum temporale and to a lesser extent anteriorly onto the planum polare. There is considerable variability in the precise location and number of regions involved between subjects and hemispheres. Nevertheless, the average activation persistently showed two regions of activation flanking Heschl's gyrus, in both hemispheres (Fig. 5b). The anterior region extends along Heschl's gyrus, somewhat on its anteromedial side toward the circular sulcus. The posterior region also extends along Heschl's gyrus, but is located on its posterolateral side, bordering Heschl's sulcus. Both regions primarily occupy the lateral part of the auditory cortex. This pattern may correspond to functional areas in the cortex and is quite consistent with the stripe-like patterns of activation described by others and denoted as T1b and T2 (Brechmann et al., 2002; Di Salle et al., 2001; Scheich et al., 1998). These regions might correspond with the most lateral portion of the auditory core region, Te1.0 and Te1.2, and the auditory belt region, Te2 (Morosan et al., 2001).

There are strong indications from the literature that differences exist between hemispheres concerning the processing of sound, especially at higher processing levels (Gage et al., 2002; Palomaki et al., 2002; Zatorre et al., 1992, 2002; Binder et al., 1997). Nevertheless, we did not find any statistically significant differences in the amount of activation between the left and right hemisphere. This was valid for the responses to the individual ripple stimuli, as well as for the mean response to the combined set of all stimuli. The precise patterns of activation differ between hemispheres within subjects, but this variability exists between subjects as well. Because dynamic ripples contain similar spectrotemporal modulations as speech fragments, a left hemisphere preference might have been expected. However, the lack of phonological content apparently results in similar activity of both hemispheres. Current evidence indeed suggests that there are functional hierarchies, such that the early stages of processing take place in auditory core areas bilaterally. High cognitive functions like speech recognition and comprehension depend on these areas, but also involve more extensive regions predominantly in the left hemisphere (Zatorre et al., 2002; Binder et al., 2000; Hickok and Poeppel, 2000; Demonet et al., 1992)

Topographic organization

The auditory cortex contains mappings of certain sound characteristics (Schreiner, 1998). Tonotopy is best known and refers to the fact that the frequency of the tone that evokes the largest response depends on the cortical location of the neurons. In this experiment mappings of dynamic ripple parameters onto the auditory cortex were considered. It would have little meaning to determine the parameter values for which the absolute response of a voxel is largest, because we have already shown that the mean response in the brain is highly parameter dependent (Fig. 4). The overall differences in sensitivity would overshadow the differences that might exist between voxels. In fact almost all voxels did indeed respond best in an absolute sense to the same set of parameter values ($\Omega = \frac{1}{4}$ c/o , $\omega = 2$ c/s, $A = 100\%$). Therefore we corrected for the brain's overall responsiveness by normalizing. Then, if no topographic organization of dynamic ripple parameters were present, approximately identical distributions of activation would be found for all dynamic ripple conditions. For each voxel it would be a matter of chance which dynamic ripple evoked the largest normalized response. Using this normalized definition of topicity we found strong indications that a topographic organization indeed exists for Ω and ω, but no systematic topicity was clear for d or A.

With respect to the temporal modulation frequency ω, we found a topographic mapping approximately in the lateral-to-medial direction. Regarding the spectral modulation density Ω there is a mapping approximately in the posterolateral-to-anteromedial direction. This topicity was found in both hemispheres, although the results for the right hemisphere were slightly more consistent between subjects. With respect to temporal modulations, similar large-scale topographic organizations were found in humans (Giraud et al., 2000) and nonhuman vertebrates (Miller et al., 2002; Schreiner and Urbas, 1988). The mapping regarding spectral modulations can be related to findings concerning a spatially dependent selectivity to the spectral broadness of stimuli (Rauschecker et al., 1995). Both topographic organizations span the entire auditory cortex. This suggests a hierarchical organization of the human auditory cortex as a whole with regard to modulation density and frequency. The anteromedial region responds relatively well to sounds with small-scale spectral and temporal features (i.e., high Ω and ω), representing acoustic details like specific voice characteristics or the presence of echoes. This is in line with the primary auditory cortex being located in this region (Morosan et al., 2001; Rademacher et al., 2001; Galaburda and Sanides, 1980). The posterior and lateral regions, on the other hand, respond best to the slowly changing large-scale features in the spectrum (i.e., small Ω and ω), in both the absolute and the relative sense, by performing some spectrotemporal integration (Thivard et al., 2000). This indicates a specialization toward the processing of formants and phonemes. This agrees with the fact that speech-related areas lie in these regions (Wise et al., 2001; Hickok and Poeppel, 2000).

In electrophysiological experiments, neuronal preferences were shown for the direction (d) of spectral drifts (deCharms et al., 1998; Tian and Rauschecker, 1998; Shamma et al., 1993). Several reasons can explain why our results did not demonstrate any differences in the average brain response or any topographic mapping with respect to d. First of all, the number of responding neurons may be small. Note that only a small fraction of the neurons might respond to a stimulus with a given set of parameter values (Ω, ω, d and A) when selectivities exist for combinations of these parameters. Functional MRI is a technique with a limited signal-to-noise ratio and will not be sensitive enough to distinguish a sufficiently small fraction from the background activation of the complete cell population. Second, a population bias toward either drift direction may be absent in humans. If neurons with preferences for upward drifts have the same spatial density as neurons that preferentially respond to downward moving ripples, then the measured macroscopic response patterns may be similar. The literature suggests that neurons with directional preferences are clustered at small scales (<0.2 mm) within cortical fields (Nelken and Versnel, 2000), while at larger scales directional preferences do not differ between multiple fields (Tian and Rauschecker, 1998; Kowalski et al., 1995). This scale at which clustering is found is too small to resolve using fMRI at 1.5 T.

In summary, we found a shift in the brain's sensitivity from small-scale to large-scale spectrotemporal features as signals are mediated from primary to secondary auditory cortex. Such a notion is of fundamental importance in the understanding of the hierarchical processing as it takes place in the auditory cortex.

References

Amaro Jr., E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., Brammer, M.J., McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: Current strategies and future prospects. J. Magn. Reson. Imaging 16, 497–510.

Backes, W.H., van Dijk, P., 2002. Simultaneous sampling of event-related BOLD responses in auditory cortex and brainstem. Magn. Reson. Med. 47, 90–96.

Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Brain Res. Cogn. Brain Res. 13, 17–26.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. Neuroimage 10, 417–429.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Bilecen, D., Scheffler, K., Schmid, N., Tschopp, K., Seelig, J., 1998. Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI. Hearing Res. 126, 19–27.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., Prieto, T., 1997. Human brain language areas identified by functional magnetic resonance imaging. J. Neurosci. 17, 353–362.

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

Brechmann, A., Baumgart, F., Scheich, H., 2002. Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. J. Neurophysiol. 87, 423–433.

Calhoun, B.M., Schreiner, C.E., 1998. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. Eur. J. Neurosci. 10, 926–940.

Chi, T., Gao, Y., Guyton, M.C., Ru, P., Shamma, S., 1999. Spectro-temporal modulation transfer functions and speech intelligibility. J. Acoust. Soc. Am. 106, 2719–2732.

deCharms, R.C., Blake, D.T., Merzenich, M.M., 1998. Optimizing sound features for cortical neurons. Science 280, 1439–1443.

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

Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 85, 1220–1234.

Di Salle, F., Formisano, E., Seifritz, E., Linden, D.E., Scheffler, K., Saulino, C., Tedeschi, G., Zanella, F.E., Pepino, A., Goebel, R., Marciano, E., 2001. Functional fields in human auditory cortex revealed by time-resolved fMRI without interference of EPI noise. Neuroimage 13, 328–338.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., 1999. Improved auditory cortex imaging using clustered volume acquisitions. Human Brain Mapp. 7, 89–97.

Engelien, A., Yang, Y., Engelien, W., Zonana, J., Stern, E., Silbersweig, D., 2002. Physiological Mapping of Human Auditory Cortices with a Silent Event Related fMRI Technique. Neuroimage 16, 944.

Gage, N.M., Roberts, T.P., Hickok, G., 2002. Hemispheric asymmetries in auditory evoked neuromagnetic fields in response to place of articulation contrasts. Brain Res. Cogn. Brain Res. 14, 303–306.

Galaburda, A., Sanides, F., 1980. Cytoarchitectonic organization of the human auditory cortex. J. Comp. Neurol. 190, 597–610.

Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A., 2000. Representation of the temporal envelope of sounds in the human brain. J. Neurophysiol. 84, 1588–1598.

Hall, D.A., Haggard, M.P., Summerfield, A.Q., Akeroyd, M.A., Palmer, A.R., Bowtell, R.W., 2001. Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. J. Acoust. Soc. Am. 109, 1559–1570.

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

Hart, H., Palmer, A., Hall, D., 2002. Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex. Hearing Res. 171, 177.

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

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–802.

Jäncke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage 15, 733–746.

Jäncke, L., Shah, N.J., Posse, S., Grosse Ryuken, M., Muller Gartner, H.W., 1998. Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia 36, 875–883.

Klein, D.J., Depireux, D.A., Simon, J.Z., Shamma, S.A., 2000. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J. Comput. Neurosci. 9, 85–111.

Kowalski, N., Versnel, H., Shamma, S.A., 1995. Comparison of responses in the anterior and primary auditory fields of the ferret cortex. J. Neurophysiol. 73, 1513–1523.

Kowalski, N., Depireux, D.A., Shamma, S.A., 1996. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J. Neurophysiol. 76, 3503–3523.

Le, T.H., Patel, S., Roberts, T.P., 2001. Functional MRI of human auditory cortex using block and event-related designs. Magn. Reson. Med. 45, 254–260.

Liang, L., Lu, T., Wang, X., 2002. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J. Neurophysiol. 87, 2237–2261.

Lockwood, A.H., Salvi, R.J., Coad, M.L., Arnold, S.A., Wack, D.S., Murphy, B.W., Burkard, R.F., 1999. The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0 kHz tones at varied intensities. Cereb. Cortex 9, 65–76.

Mendelson, J.R., Schreiner, C.E., Sutter, M.L., Grasse, K.L., 1993. Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. Exp. Brain Res. 94, 65–87.

Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., 2002. Spectro-temporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 87, 516–527.

Mohr, C.M., King, W.M., Freeman, A.J., Briggs, R.W., Leonard, C.M., 1999. Influence of speech stimuli intensity on the activation of auditory cortex investigated with functional magnetic resonance imaging. J. Acoust. Soc. Am. 105, 2738–2745.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684–701.

Nelken, I., Versnel, H., 2000. Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex. Eur. J. Neurosci. 12, 549–562.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Palomaki, K.J., Tiitinen, H., Makinen, V., May, P., Alku, P., 2002. Cortical processing of speech sounds and their analogues in a spatial auditory environment. Brain Res. Cogn. Brain Res. 14, 294–299.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., Zilles, K., 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13, 669–683.

Rauschecker, J.P., Tian, B., Hauser, M., 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111–114.

Scheich, H., Baumgart, F., Gaschler Markefski, B., Tegeler, C., Tempelmann, C., Heinze, H.J., Schindler, F., Stiller, D., 1998. Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. Eur. J. Neurosci. 10, 803–809.

Schreiner, C.E., 1998. Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. Audiol. Neuro-Otol. 3, 104–122.

Schreiner, C.E., Urbas, J.V., 1988. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hearing. Res. 32, 49–63.

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

Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J.G., Bilecen, D., Tedeschi, G., Scheffler, K., Di Salle, F., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297, 1706–1708.

Shamma, S.A., Fleshman, J.W., Wiser, P.R., Versnel, H., 1993. Organization of response areas in ferret primary auditory cortex. J. Neurophysiol. 69, 367–383.

Supin, A., Popov, V.V., Milekhina, O.N., Tarakanov, M.B., 1999. Ripple depth and density resolution of rippled noise. J. Acoust. Soc. Am. 106, 2800–2804.

Talavage, T.M., Edmister, W.B., Ledden, P.J., Weisskoff, R.M., 1999. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. Human Brain Mapp. 7, 79–88.

Talavage, T.M., Ledden, P.J., Benson, R.R., Rosen, B.R., Melcher, J.R., 2000. Frequency-dependent responses exhibited by multiple regions in human auditory cortex. Hearing Res. 150, 225–244.

Thivard, L., Belin, P., Zilbovicius, M., Poline, J.B., Samson, Y., 2000. A cortical region sensitive to auditory spectral motion. Neuroreport 11, 2969–2972.

Tian, B., Rauschecker, J.P., 1998. Processing of frequency-modulated sounds in the cat's posterior auditory field. J. Neurophysiol. 79, 2629–2642.

Versnel, H., Shamma, S.A., 1998. Spectral-ripple representation of steady-state vowels in primary auditory cortex. J. Acoust. Soc. Am. 103, 2502–2514.

Versnel, H., Kowalski, N., Shamma, S., 1995. Ripple analysis in ferret primary auditory cortex. III. Topographic distribution of ripple response parameters. Auditory Neurosci. 1, 271–286.

Wakefield, G.H., Viemeister, N.F., 1990. Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise. J. Acoust. Soc. Am. 88, 1367–1373.

Wessinger, C.M., Buonocore, M.H., Kussmaul, C.L., Mangun, G.R., 1997. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. Human Brain Mapp. 5, 18–25.

Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124, 83–95.

Wong, D., Pisoni, D.B., Learn, J., Gandour, J.T., Miyamoto, R.T., Hutchins, G.D., 2002. PET imaging of differential cortical activation by monaural speech and nonspeech stimuli. Hearing Res. 166, 9–23.

Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E., Silbersweig, D.A., 2000. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. Magn. Reson. Med. 43, 185–190.

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

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