NeuroImage 43 (2008) 321–328

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise

Stephan M.A. Ernst [a], Jesko L. Verhey [a], Stefan Uppenkamp [b,*]

[a] AG Neurosensorik, Carl von Ossietzky Universität, 26111 Oldenburg, Germany
[b] Medizinische Physik, Carl von Ossietzky Universität, 26111 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 16 November 2007
Revised 22 July 2008
Accepted 25 July 2008
Available online 3 August 2008

Keywords:
Auditory cortex
Hearing
Masking
fMRI
Loudness

ABSTRACT

Functional magnetic resonance imaging has been used to investigate the signal representation in human auditory cortex for a sinusoidal signal in the presence of a noise masker. This paradigm is widely used in auditory research to study auditory processing. Five-note tonal melodies were presented in a masking noise for signal-to-noise ratios (S/N) from −18 dB to+24 dB in 6 dB-steps. For small S/N (−18 dB, −12 dB, −6 dB) the overall level of the sound is nearly constant, but the audibility of the tone varies with S/N. For S/N of 0 dB and above, the tone is always clearly audible, and the perceived change is mainly the increase in overall level. This interaction between S/N, overall level and perception is reflected by a spatial dissociation of the respective activation in auditory cortex. Brain regions mainly sensitive to level changes were found in various parts of the superior temporal lobes, including primary auditory cortex and Planum temporale, while those regions mainly sensitive to S/N changes were located at or close to lateral Heschl's gyrus. The overlap between these two regions is small. The results are interpreted as indicating that the coding of overall level and, thus, loudness is different from the coding of audibility of a periodic signal. The S/N-sensitive region largely overlaps with the pitch-sensitive regions in lateral Heschl's gyrus found in previous studies. The results from the present study further suggest that the audibility of a tone in noise is related to the overall pitch strength.

Introduction

One important aspect of communication with acoustic signals is the temporal fluctuation of the sound level. Background noise determines to a great extent the audibility of any acoustic signal, the audibility of any modulation of the signal carrying much of the information like in running speech, and also e.g. the phrasing in a musical performance, since it masks parts of the signal of interest. In psychoacoustics, masking denotes the effect that the detection threshold for a target signal, for example a pure tone or a modulation of a certain carrier signal, is increased in the presence of another acoustic signal, e.g. a masking noise. The phenomenon of masking has been widely used in auditory research to study both peripheral and central auditory processing in human listeners (for an overview, see e.g. Moore, 2003, chapter three). For example, the detection threshold for a sinusoidal signal in the presence of a masking noise of varying bandwidth can be used to characterize the frequency selectivity of the auditory system (Fletcher, 1940). On the basis of experiments with bandpass-noise maskers centred at the signal frequency, Fletcher proposed the energy detector model. The model assumes that the incoming sound is analysed by a bank of overlapping bandpass filters and that the filter with the largest ratio of signal to masker energy, commonly referred to as signal-to-noise ratio (S/N), determines the detection threshold. The model thus assumes that signal detection is simply determined by the relative intensity of the sounds. Consequently, detection thresholds are usually expressed as S/N at threshold.

When listening to a target signal, e.g. a sinusoid, in a broadband masking noise, the target is audible even when it is up to 18 dB lower in level than the masker. For any signal-to-noise ratio of −3 dB or less, the overall intensity is largely determined by the level of the masker, while the perceived change with S/N is simply the audibility of the target tone, irrespective of overall intensity. For S/N of 0 dB and above, the tone is always clearly audible, and the perceived change with S/N is mainly a change of the overall loudness of the sound due to the increase in the level of the target tone. In order to detect the signal at low S/N, the auditory system needs to distinguish the tone from the masker. A sinusoidal signal produces the sensation of pitch, whereas a broadband noise usually has no pitch. Thus, at low S/N, when the tone is just audible, an additional feature is added, and the perceived pitch strength then increases with increasing S/N (Zwicker and Fastl, 1999, p. 144, Fig. 5.37d). This may also serve as a detection cue.

It is still not completely understood how the physical parameter of sound intensity in general and the effect of masking in particular are exactly transformed into the perceptual parameter of loudness and audibility of a signal, and how these perceptual quantities are represented in the human auditory cortex. Functional magnetic

* Corresponding author. Fax: +49 441 7983902.
E-mail address: stefan.uppenkamp@uni-oldenburg.de (S. Uppenkamp).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2008.07.046

resonance imaging has been used in several studies to investigate the relationship between sound intensity in the auditory brainstem and in primary and secondary auditory cortex (Jäncke et al., 1998; Mohr et al., 1999; Hall et al., 2001; Brechmann et al., 2002; Bilecen et al., 2002; Hart et al., 2002; Scarff et al., 2004; Yetkin et al., 2004; Sigalovsky and Melcher, 2006; Langers et al., 2007). Most studies have demonstrated a systematic increase of brain activation in auditory areas as the level increases. A quantitative comparison between studies, however, appears to be difficult due to the great variability in the type of stimuli and range of levels employed. A complete intensity-activation-function can only be studied using a wide range of stimulus levels, as has been done by Langers et al. (2007). They found a better correlate of the growth of brain activation with loudness level rather than intensity level, and also a significantly steeper slope in the growth of brain activation with level for hearing impaired listeners showing loudness recruitment. These results were interpreted to mean that the activation represents the subjective strength of the stimulus percept, that is its loudness, rather than stimulus intensity.

So far, only little attention has been paid to fMRI correlates of the psychacoustical effect of masking. van Dijk and Backes (2003) investigated the differences between simultaneous and backward masking, using a sparse imaging fMRI protocol. For regions beyond primary auditory cortex, they reported a different activation pattern for backward and simultaneous masking tasks. Backward masking showed more activation in anterior regions when compared to simultaneous masking, while simultaneous masking gave more response in posterior regions (van Dijk and Backes, 2003, p.13). They interpreted their results as a difference in the cognitive brain resources used, with variable involvement of “what” and “where” processing streams, as suggested by Romanski et al. (1999). Although the concept of “what” and “where” streams in human auditory cortex is a matter of debate (cf. Kaas and Hackett, 1999; Belin and Zatorre, 2000; Alain et al., 2001; Maeder et al., 2001; Zatorre et al., 2002), there is still strong evidence for a posterior pathway including Planum temporale for location-specific processing (e.g. sound-source motion: Warren et al., 2002). Griffiths and Warren (2002) have suggested a model of human Planum temporale as a general centre for the processing of spectrotemporal patterns and sound source segregation. In line with this hypothesis, Binder et al. (2004) found activation in Planum temporale that was related to task performance in the discrimination of two phonemes in the presence of noise. The two phonemes only differed in the frequency trajectory of their formants, i.e. a subtle change in their spectro-temporal patterns.

In the current study, we investigate the cortical representation of pure tones in the presence of a masking noise for varying signal-to-noise ratio. We expect (1) that there are regions in auditory cortex that will show increasing activation as a function of increasing overall level, as reported in previous studies, and (2) that there are regions in auditory cortex that will show increasing activation as a function of increasing S/N, that is, with increasing audibility of the tone in the masking noise irrespective of overall sound intensity. Regions according to expectations (1) and (2) do not need to be the same.

Using magnetoencephalography, Gutschalk et al. (2002) showed a spatial separation for the processing of sound level and temporal regularity in the auditory cortex. They found a shift from more periodicity-specific regions in the lateral part of Heschl's gyrus to more intensity-specific regions in Planum temporale for regular and irregular click trains of varying pitch and intensity. We hypothesize that the perceptual dissociation of 'changing audibility of the tone within the noise for negative S/N' and 'changing loudness level for positive S/N' may be reflected by a similar shift of activation from more periodicity-specific regions in the lateral part of Heschl's gyrus for low signal-to-noise ratios to regions in Planum temporale, which are more sensitive to changes in overall intensity.

Materials and methods

Stimuli

The tones used in this study were sinusoids of 750 ms duration with 50 ms on- and offset ramps and with frequencies between 440 and 587 Hz. They were presented as five-note diatonic melodies together with the masking noise for S/N from −18 dB to +24 dB in 6 dB-steps. The silent interval between two successive notes was 250 ms. The level of the masking noise was kept constant at 65 dB SPL. The frequencies for the sinusoids within one stimulus presentation (trial) were selected randomly from the five predefined notes (440 Hz to 587 Hz). The selection was with replacement, i.e. the frequency selected for the previous note was left in the pool for the next selection. Thus it is possible that the same note occurred twice in succession.

The task for the subject was to press a button whenever there were two consecutive identical notes within one trial, which is true for approximately half of the trials. This task was not related to the effects of interest, i.e. the overall level and S/N. Its only role was to keep listeners attentive to the stimuli. In addition to the eight different S/N conditions, noise only and silence were run as control conditions. Stimulus conditions changed randomly from trial to trial. Overall each condition was repeated 32 times, giving a total of 320 trials for each listener. All stimuli were bandpass filtered between 250 Hz and 4 kHz and presented diotically to the subjects via MR-compatible, dynamic headphones mounted into ear defenders (Baumgart et al., 1998, MR confon, Magdeburg,).

For the small S/N −18, −12, and −6 dB the overall level of the sound is nearly constant (see Table 1), but the audibility of the tone varies. For S/N of 0 dB and above, the tone is always clearly audible, and the perceived change is mainly an increase of overall loudness.

Subjects

Twelve listeners between 22 and 28 years of age (mean 23.9 a, standard deviation 1.7 a) were scanned (6 male, 6 female, all but one were right-handed). None of the volunteers had a history of neurological illness, head injury, or hearing impairment. Written informed consent was obtained from all volunteers. The study was approved by the local ethics committee of the University of Oldenburg and by the ethics committee at the Medical School of the University of Göttingen.

Scanning procedure

Functional MRI data were acquired using a Siemens Sonata 1.5 T MRI system. In total, 320 echo-planar imaging (EPI) volumes were acquired over four runs of 80 volumes each. A T_1-weighted high-resolution anatomical image (176 sagittal slices, TR = 2.11 s, TE = 4.38 ms) was also collected for each subject. For the functional data, twenty-one axial EPI slices (in-plane resolution 3×3 mm; thickness 5 mm, echo time TE = 63 ms to maximize BOLD contrast) were acquired covering most of the cortex, including the whole of the temporal lobes. Sparse imaging with clustered volume acquisition was used (Edmister et al., 1999; Hall et al., 1999). The total volume acquisition time TA was 2.5 s. On each trial, there was a 5.2-s stimulus interval followed by the 2.5-s scanning interval, making a total repetition time of TR = 7.7 s. The sparse imaging technique minimizes the interference of scanner noise on brain activity because the

Table 1
Overview of presentation levels during different S/N conditions. The third row gives the overall sound pressure level of the signals as a function of S/N from −18 dB to +24 dB

Condition	1	2	3	4	5	6	7	8	9	10
S/N (dB)	Silence	−∞	−18	−12	−6	0	6	12	18	24
Overall level (dB HL)	Background noise	62	62	62	62	65	68	74	80	86