Of cats and women: Temporal dynamics in the right temporoparietal cortex reflect auditory categorical processing of vocalizations

NeuroImage 62 (2012) 1877–1883

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Of cats and women: Temporal dynamics in the right temporoparietal cortex reflect auditory categorical processing of vocalizations

Hanna Renvall [a,b,*], Noël Staeren [a], Nicolette Siep [a], Fabrizio Esposito [a], Ole Jensen [c], Elia Formisano [a]

[a] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[b] Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University, P.O. Box 15100, FI-00076 Aalto, Finland
[c] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Accepted 10 June 2012
Available online 19 June 2012

Keywords:
Categorical perception
Imaging

ABSTRACT

Understanding the temporal dynamics underlying cortical processing of auditory categories is complicated by difficulties in equating temporal and spectral features across stimulus classes. In the present magnetoencephalography (MEG) study, female voices and cat sounds were filtered so as to match in most of their acoustic properties, and the respective auditory evoked responses were investigated with a paradigm that allowed us to examine auditory cortical processing of two natural sound categories beyond the physical make-up of the stimuli. Three cat or human voice sounds were first presented to establish a categorical context. Subsequently, a probe sound that was congruent, incongruent, or ambiguous to this context was presented. As an index of a categorical mismatch, MEG responses to incongruent sounds were stronger than the responses to congruent sounds at ~250 ms in the right temporoparietal cortex, regardless of the sound category. Furthermore, probe sounds that could not be unambiguously attributed to any of the two categories ("cat" or "voice") evoked stronger responses after the voice than cat context at 200–250 ms, suggesting a stronger contextual effect for human voices.

Our results suggest that categorical templates for human and animal vocalizations are established at ~250 ms in the right temporoparietal cortex, likely reflecting continuous online analysis of spectral stimulus features during auditory categorizing task.

Introduction

The ability to rapidly recognize and categorize sounds is essential, not only for understanding and reacting to our surroundings, but for daily communication and social interaction. Studies in macaque monkeys have suggested that auditory information relevant for sound recognition in general is processed in a specialized and anatomically segregated stream of cortical areas (Kaas and Hackett, 1999; Rauschecker and Tian, 2000; Romanski et al., 1999). Correspondingly in humans, sound recognition activates regions located laterally to the Heschl's gyrus and extending along the posterior–anterior direction of the superior temporal gyrus (STG) and sulcus (STS) (Alain et al., 2001; Warren and Griffiths, 2003). Within these areas, sound categories are encoded in a spatially distributed manner (Formisano et al., 2008; Staeren et al., 2009).

In humans, both animal and human vocalizations constitute rapidly and effortlessly recognizable auditory categories that are learned early in childhood and share many spectrotemporal features. Vocalizations activate specific auditory networks: Regions in the bilateral STS and STG exhibit a larger blood-oxygenation-level-dependent response to vocal than to non-vocal human sounds (Belin et al., 2000, 2004; Warren et al., 2006), and the middle portions of the STG are bilaterally more activated during the categorization of animal vocalizations than tool sounds (Lewis et al., 2005). Furthermore, subregions at these areas show species-specific reactivity to vocalizations (Fecteau et al., 2004).

In functional magnetic resonance imaging (fMRI) studies, minimizing the low-level acoustic differences between stimuli abolishes conventional univariate differences between responses to different sound categories (Staeren et al., 2009). Exemplars of separate categories differ from each other temporospectrally, and time-sensitive electroencephalographic (EEG) and magnetoencephalographic (MEG) responses are especially sensitive to such deviations. In a recent EEG study, responses to human voices differed from those to bird songs and environmental sounds at ~200 ms bilaterally at the frontotemporal electrodes, but the results were speculated to be at least partly due to differences between the experimental stimuli (Charest et al., 2009). Another EEG study, in which the sound spectrograms and power spectra did not statistically significantly differ between sound categories, demonstrated stronger activity to human than animal vocalizations at 169–219 ms over the right temporal areas (De Lucia et al., 2010). However, the same ~200-ms time window has

* Corresponding author at: Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University, P.O. Box 15100, 00076 Aalto, Finland. Tel.: +358-40-7036161
E-mail address: hanna@neuro.hut.fi (H. Renvall).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.06.010

been related to general processing of spectral fine structure of any complex sound (Altmann et al., 2008), and the nature of auditory categorical processing has remained unclear.

Here we used MEG in combination with acoustically well-controlled human and cat vocalizations to study cortical processing of auditory categories beyond the processing of low-level features. As an important addition to previous studies, the temporal profiles of our stimuli were equated for their harmonic structures. This manipulation ensures that the sounds have a similar "perceptual pitch" profile over time, behaviorally relevant for sound categorization (Staeren et al., 2009). Furthermore, we used an adaptation paradigm in which exact same stimuli could be presented in different contexts. Based on a predictive coding account of auditory adaptation (Friston, 2005; Garrido et al., 2007, 2008; Jääskeläinen et al., 2004; Wacongne et al., 2011), we hypothesized that sounds incongruent to the preceding context, would produce — in the superior temporal cortex — stronger responses than congruent sounds as a marker of a categorical mismatch. Finally, we probed and compared these categorical adaptation effects for the two different contexts ("voice" and "cat") with acoustically identical target sounds that could not be unambiguously attributed to any of the two categories.

Materials and methods

Subjects

We studied, with informed consent, 8 adults (mean ± SEM age 28 ± 1 years; 3 females, 5 males; 7 right-handed and one ambidextrous). None of the subjects had a history of hearing or neurological impairments, and the study received a prior approval by the Ethical Committee of the Faculty of Psychology, Maastricht University.

Auditory stimuli and experimental design

One cat (*meowing*) and one voice sound (singing female) were selected from the stimulus set used in Staeren et al. (2009), on the basis of their close resemblance in harmonics-to-noise ratios (Boersma, 1993; Lewis et al., 2005; Murray et al., 2006) and power spectra. To further minimize the spectrotemporal differences between the stimuli, the time-varying fundamental pitch of the cat sound was extracted at 25 time points (in ~30 ms steps) within the stimulus with Praat software (Boersma, 2001) and applied to the voice sound using Adobe Audition™. Sounds were then low-pass (LP) filtered at 13 cut-off frequencies; the LP frequencies varied in steps of 100 Hz between 500 and 900 Hz, and in steps of 200 Hz between 900 and 2500 Hz. To add more variation to the stimuli, they were transposed to five different fundamental frequencies between 230 and 260 Hz. These procedures resulted in 65 stimuli for each of the two categories (5 pitch levels × 13 frequency ranges). The stimuli lasted for 780 ms, and they were equalized for their mean intensities with MATLAB 7.0.1™ (The MathWorks, Inc., Natick, MA, USA). Differences in stimulus amplitude envelopes between cat and voice stimuli were minimized by using 10-ms moving-average windows, to an extent not to disturb original sound quality.

The stimuli were tested behaviorally in 14 subjects who did not participate in the final experiment. In these behavioral tests, subjects were first familiarized with six easily recognizable representatives from both categories together with visual information about the sound category (Presentation 9.3™, Neurobehavioral Systems, Inc., Albany, CA, USA). Then, they were instructed to carefully listen to the sounds presented at 2 s interstimulus intervals (ISI), and to decide whether the sound was a voice or a cat stimulus. Subjects were asked to be as accurate and fast as possible, and their ratings were reported through button presses. After a few practise trials, each stimulus (65 per category) was presented nine times.

At the largest bandwidths, the stimuli sounded very natural and, correspondingly, they were easily recognized as representatives of their category, while narrowing the bandwidth gradually affected the behavioral response. On the basis of the results, nine cat/voice stimulus pairs with similar recognition accuracies and reaction times between categories were selected as "easy". These sounds consisted of LP levels 1500 Hz (at two different pitch levels), 1900 Hz (three pitch levels), and 2300 Hz (four pitch levels). In addition, the voice sounds that were LP-filtered at 500 Hz (four pitch levels) resulted in behavioral responses at chance level, and they were selected as "ambiguous". Examples of the stimuli and their spectrograms are presented in Fig. 1 (for auditory examples, see Supplement auditory material S1–S3).

Despite the efforts to minimize the spectrotemporal differences between stimulus categories, the easily recognizable female voice stimuli contained more energy at ~1000 — 1500 Hz than the cat vocalizations throughout the stimulus duration (see Fig. 1a and b). The remaining amplitude differences between cat and female voice stimuli were tested by analyzing the sound intensities in 20-ms steps at 0–220 ms from the beginning of the stimuli: the stimulus intensities did not differ statistically significantly between the cat and voice stimuli ($p > 0.09$). Although the ambiguous stimuli were modified from the voice stimuli by LP filtering at 500 Hz and thus their resembled more closely the voice stimuli in their amplitude behavior, their spectrotemporal structure was rather flat at 0–500 Hz and did not contain the upper harmonics that were characteristics for both the easy voice and cat stimuli.

During the MEG session, the behavioral responses were too scarce for statistical inference. Therefore, in a separate session prior to the MEG experiment, all subjects underwent a short behavioral test (Presentation 9.3™). First the subject listened twice to all nine "easy" cat and voice stimuli presented with an ISI of 2 s, together with visual information on the stimulus category. Subsequently, the same stimuli were presented randomly three times without visual aid and interspersed with the ambiguous stimuli, and the subject was asked to respond with a button press whether the stimulus was a cat or a female voice. Finally, the subjects listened to the sounds as they would be presented in the MEG experiment, i.e. four sounds in a row, and they were asked to respond after each trial whether the all four sounds belonged to the same category (*yes*/*no*).

The percentage of correct cat and voice sound recognition was ≥97 ± 2% (mean ± SEM). Subjects' responses to the ambiguous sounds were at the chance level: The percent correct (the subject responded 'voice') was 39 ± 12% when the sounds were presented after cat sounds, and 63 ± 14% after voice sounds ($p > 0.35$ compared

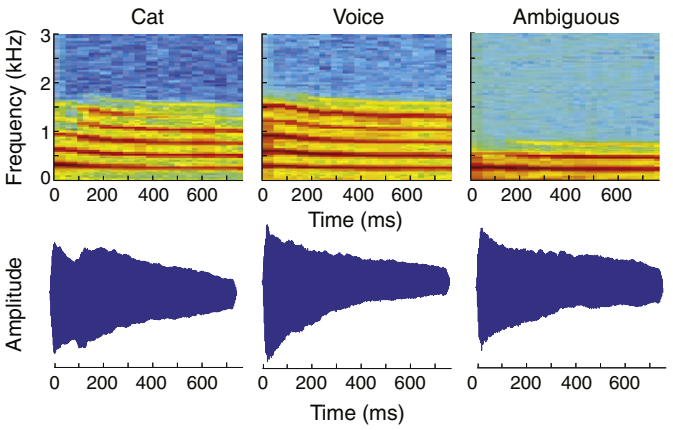

Fig. 1. Spectrograms of exemplary cat and voice stimuli (both low-pass filtered at 1900 Hz), and of ambiguous stimuli (voice sound, low-pass filtered at 500 Hz). The time-varying fundamental frequency of the cat sound was extracted and imposed onto the voice stimuli. All the harmonics of the voice sounds were modified accordingly.

with 50%), and the responses did not differ statistically significantly from each other ($p = 0.15$).

MEG experiment

In the MEG experiment, the sounds were delivered to the subjects binaurally at a comfortable listening level through plastic tubes and ear pieces. They were presented in trains of four, and the subject's task was to attend to all sounds carefully, and decide whether the sounds belonged to the same category (cat or voice). The experiment is described schematically in Fig. 2. The stimuli within a train were presented with ISIs of 600 ms (from offset to onset), resulting in a trial duration of 4920 ms, and they were followed by an inter-trial interval of 2700 ms.

The experiment consisted of six conditions utilizing the stimuli described above (nine voice sounds, nine cat vocalizations and four ambiguous sounds). In the congruent conditions, four voice (or cat) sounds were presented in a row. In the incongruent conditions, three voice (cat) sounds were followed by a cat (voice) sound. In the ambiguous conditions, three voice (or cat) stimuli were followed by an ambiguous stimulus. To minimize build-up of purely acoustic memory traces during the trials and to avoid mismatch responses elicited by infrequent sounds among otherwise monotonous stimulation (Näätänen, 1992), the three first stimuli in a train were selected each from a different filtering level. The last sound in a row could be either from the same or different filtering level as the preceding third sound; MEG responses were pooled across the different filtering and pitch levels. The different stimulus trains were presented in a random order, and the same condition was not allowed to occur more than twice in succession.

In 7% of the trials, a question mark appeared 1 s after the last stimulus, and the subject was required to respond by lifting her/his index or middle finger whether the sounds belonged to a same category (*yes*/*no*). The subsequent trials were discarded from the analysis. The response hand was alternated across subjects, and to minimize possible motor contamination on the data, subjects were instructed to keep their hand relaxed during the experiment. To prevent subjects' deciding on the last stimulus only, 7% of the trials were "catch trials" in which the incongruent stimulus occurred at the first, second or third stimulus position. These responses were also removed from the data analysis.

Auditory evoked fields were recorded in a magnetically shielded room using a whole-head MEG system (VSM/CTF Systems Inc., Port Coquitlam, Canada) with 275 axial gradiometers. Three head position indicator coils were attached at anatomical landmarks (the left and right ear canals and the nasion). The head position with respect to the sensor array was determined by feeding current to the marker coils and measuring their positions with respect to the sensory array before and after the measurements.

The MEG signals were low-pass filtered at 300 Hz and digitized at 1200 Hz, and averaged offline with two time scales: (i) from 200 ms before the onset of the whole stimulus block to 1000 ms after the onset of the last (4th) stimulus, and (ii) from 200 ms before the onset of each stimulus to 1000 ms after it. The averaged signals were digitally low-pass filtered at 40 Hz, and a prestimulus baseline of 200 ms was applied.

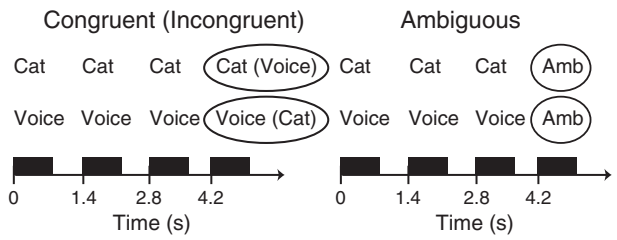

Fig. 2. Schematic presentation of the incongruent and ambiguous experimental trials. Note that the stimuli within a trial varied both in their LP filtering and pitch levels (see text).

The experiment was conducted in 5 blocks, each lasting ~10 min. During the experiment each of the six conditions (two congruent, two incongruent and two ambiguous conditions) was repeated 70 times. Horizontal and vertical electro-oculograms were recorded to discard data contaminated by eye blinks and movements; ~60–70 artifact-free responses were averaged per condition.

MEG sensor-level signals

For an initial estimate of the experimental effects, the responses to whole stimulus blocks were first analyzed at the sensor level. To simplify the analysis, a planar gradient was estimated for each channel from the neighbouring channels (Medendorp et al., 2007). Planar gradients give the maximum signal just above the source area (Hämäläinen et al., 1993). Root mean square of the horizontal and vertical planar gradient fields was then calculated (combined planar gradient). Subsequently areal mean averages were calculated over the central, left and right temporal, left and right frontal, and left and right occipito-parietal regions.

Source analysis: equivalent current dipole modelling

For source analysis, the head was modelled as a homogeneous spherical volume conductor. The model parameters were optimised for the intracranial space obtained from MR images that were available for all subjects. The neurophysiological responses were analyzed by first segregating the recorded sensor-level signals into spatiotemporal components, by means of manually guided multidipole current modelling (equivalent current dipole, ECD; Hämäläinen et al., 1993). The analysis was conducted separately for each subject using Elekta Neuromag (Elekta Oy) software package, following standard procedures (Hansen et al., 2010; Salmelin et al., 1994). The parameters of an ECD represent the location, orientation, and strength of the current in the activated brain area. The ECDs were identified by searching for systematic local changes, persisting tens of milliseconds, in the measured magnetic field pattern. ECD model parameters were then determined at those time points at which the magnetic field pattern was clearly dipolar. The software identifies the sensor measuring the strongest signal at the channels covering the field pattern, and uses a location below this sensor as a seed point for the following ECD model parameter estimation. The parameter fit does not depend on the exact selection of the seed point in the local neighbourhood of the maximum signal. Only ECDs explaining more than 85% of the local field variance during each dipolar response peak were accepted in the multidipole model. Based on this criterion, 3–4 spatiotemporal components were selected into the individual subjects' models. The analysis was then extended to the entire time period, and all MEG channels were taken into account: The previously found ECDs were kept fixed in orientation and location while their strengths were allowed to change.

For optimizing the accuracy of the spatial fits, the orientation and location of the ECDs were estimated in each individual in the condition with the strongest signals in the time windows of the main experimental effects suggested by the sensor-level data. However, the variability in the signal-to-noise ratios between conditions was very small, and, on the basis of visual inspection and on the calculated goodness-of-fit values obtained by comparing the original data and the data predicted by the fitted sources, the same sources explained well the responses in the other conditions.

Due to the variability of the response shape across individuals, the 250-ms response amplitudes were estimated as an average over a 50-ms (for ambiguous sounds) or 100-ms window (separately for congruent and incongruent conditions) around the individual response peaks. For

consistency, 100-ms response amplitudes were estimated from 50-ms time windows around the individual response peaks.

The ECD source waveforms (average strengths and peak latencies of the responses) were statistically tested using ANOVA and paired *t* tests (two-sided, Bonferroni corrected). Effect sizes μ were estimated as the difference between two condition means divided by a standard deviation of the data across both conditions.

Source analysis: minimum norm estimates

In the auditory modality, ECD models have been shown to coincide well with distributed modelling approaches (Vartiainen et al., 2009). For verifying the spatial distribution of activity obtained with ECD modelling, the cortical generators were additionally visualized with a distributed source model, using MNE Suite software package (M. Hämäläinen, Martinos Center for Biomedical Imaging, Massachusetts General Hospital). MNE implements the L2 minimum norm estimate of the source distribution, which seeks for current distribution that explains the measurements and has the smallest L2-norm. MNE analysis results in distributed models of the cortical activation, but provides little information of the shape or extent of the activated area.

For MNE analysis, the cortical surface of each subject was reconstructed from the corresponding MR images with the Freesurfer software (Dale and Sereno, 1993; Fischl et al., 1999). Each hemisphere was covered with ~5000 potential source locations. Currents oriented normal to the cortical surface were favoured by weighting the transverse currents by a factor of 0.3 (Lin et al., 2006), and depth-weighting was used to reduce the bias towards superficial sources. Noise-normalized MNEs (dynamical Statistical Parametric Maps, dSPMs) were calculated over the whole cortical area to estimate the signal-to-noise ratios in each potential source location (Dale et al., 2000). Noise covariance matrix was estimated from the 200-ms prestimulus baseline periods in the raw data.

For group-level visualization, the MNEs of individual subjects were first normalized to the maximum value of that subject and subsequently morphed, with spatial smoothing, to one subject's brain. The statistical analysis of MNEs was performed, by means of paired two- sided *t* tests, on each subject's normalized values within a region of interest (ROI) centered around the Heschl's gyrus that contained both the MNE maxima and the ECD models of all subjects.

Results

Congruent vs. incongruent sounds: sensor-level results

The initial sensor-level analysis revealed that all four stimuli within the stimulus blocks evoked strong responses bilaterally over the temporal areas, peaking at about 100 ms and at 250–700 ms after the onset of each sound. Fig. 3 depicts the areal averages of the sensor-level signals (for the whole-head sensor-level data, see Supplementary Fig. 1). The 100-ms (N100m) responses were attenuated for the stimuli at positions 2nd − 4th compared with the first stimulus, similarly in all conditions. An additional response at around 250 ms was observed in both incongruent conditions.

Congruent vs. incongruent sounds: source-level results

Despite the careful acoustic matching of stimuli, the N100m responses to the first stimuli in a block were statistically significantly smaller for the cat than voice sounds in the left hemisphere (LH) as modelled by the ECDs (*t* test $p<0.02$, effect size $\mu=0.7$), whereas the N100m responses to other stimulus positions did not differ significantly between cat and voice stimuli in either hemisphere.

For the last stimulus, the incongruent sounds evoked prominent responses at ~250 ms after the stimulus onset in the right hemisphere

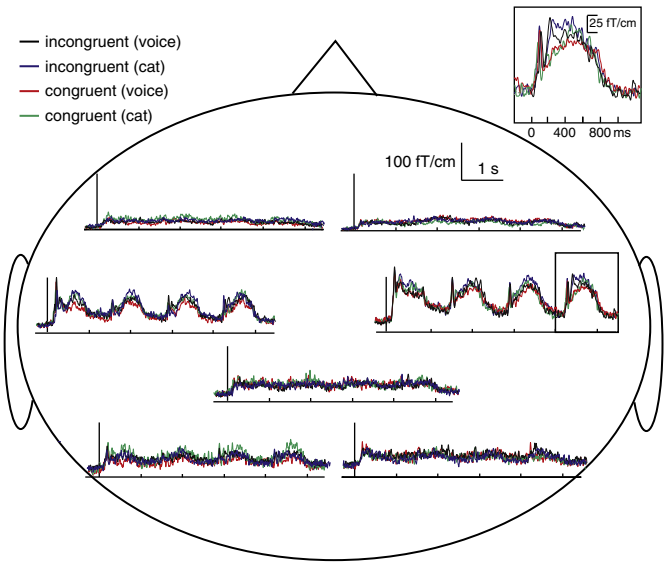

Fig. 3. MEG signals. Areal responses over all subjects to congruent and incongruent trials. Incongruent (voice) refers to an experimental condition in which the three first sounds were voice sounds, and the last sound was a cat sound. The insert shows enlarged responses to the last sounds, recorded over the right temporal cortex.

(RH), without statistically significant differences between the cat and voice contexts (ECD analysis, Congruency × Category type interaction, $F_{1,7}=0.64$; $p=0.43$), suggesting that the effect was not specific to female voices nor cat vocalizations. Thus for the subsequent analysis of the congruent/incongruent sounds, the responses to cat and voice sounds were averaged together.

Fig. 4 depicts the ECDs, the corresponding source waveforms, and the MNE dSPMs of one subject to the last sounds in the incongruent and congruent conditions, superimposed on her reconstructed cortical surface.

In agreement with previous studies (for a review, see Hari, 1990), the N100m responses were adequately explained by two ECDs, one in the left and one in the right supratemporal auditory cortex (indicated by white dipoles). The same sources explained also the sustained activity peaking >300 ms. In the RH, another source with more supero-posterior location was needed to explain the responses around ~250 ms (indicated by a blue dipole). The ECD and MNE analyses suggested rather similar sequence of cortical activation: Both methods indicated right-hemispheric temporoparietal activation ~230 − 250 ms that was stronger in the incongruent than congruent stimulus condition.

Fig. 5 illustrates the ECDs, the corresponding source waveforms, and the MNE dSPMs over all subjects to the last sounds in the incongruent and congruent conditions, morphed and superimposed on one subject's reconstructed cortical surface.

The ECD models for the different subjects consisted typically of two ECDs in the RH, and one ECD in the LH. In three subjects, ECDs explaining the field patterns around 100 ms and 250 ms in the RH were located close to each other and had very similar orientations, and to prevent interactions between these ECDs, the same ECD was used to model both responses. In one subject, a 4th ECD was needed in the LH to explain the magnetic field variations at ~250 ms (Fig. 5B). While the N100m responses were consistently located in the vicinity of planum temporale in both hemispheres in all subjects, the location of the 250-ms responses showed more interindividual variability.

The N100m responses peaked in the LH at 108 ± 8 ms and at 113 ± 7 ms (mean ± SEM), respectively, in the incongruent and congruent conditions, and in the RH at 113 ± 5 ms in both conditions, without significant differences in the ECD peak latencies or mean response amplitudes between conditions. At the LH, the responses at ~200 ms

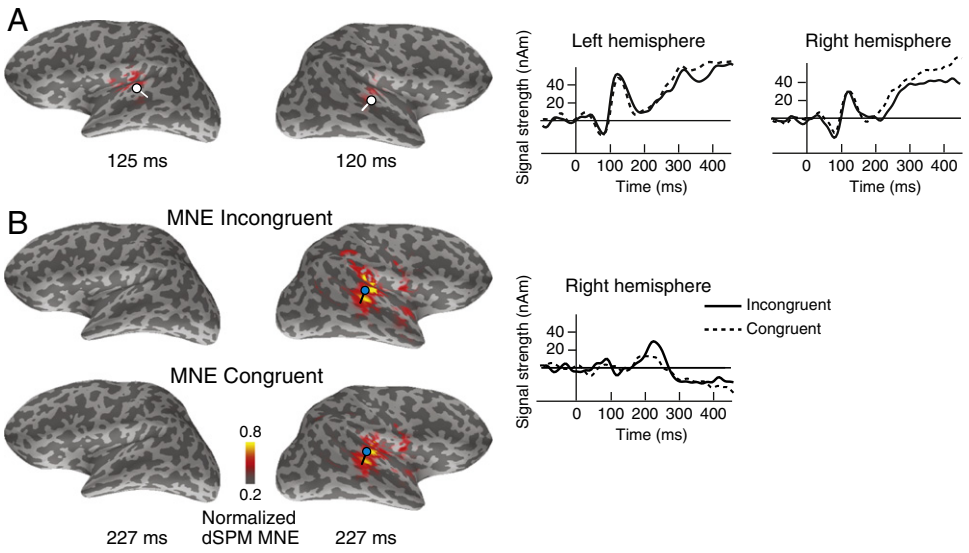

Fig. 4. MEG source analysis in one subject. The locations (dots) and orientations (tails) of the ECDs used to model the N100m responses (A, white dots), and of the right-hemispheric 250-ms responses in incongruent and congruent conditions (B, blue dots) in one subject, superimposed on the subject's MNE dSPM distributions. The inserts (right) depict the corresponding ECD time courses in a time window of −100 to 450 ms with respect to the stimulus onset.

explained by the same ECDs tended to be stronger for incongruent than congruent sounds, but this difference did not reach statistical significance (estimated individually from a 50-ms time window around the maximum difference between conditions, *t* test $p = 0.15$).

The RH 250-ms responses peaked at 230 ± 10 ms in the incongruent condition, and at 231 ± 12 ms in the congruent condition. The responses were statistically significantly stronger for the incongruent than congruent sounds as modelled by the ECDs (estimated from a 100-ms time window around the individual peak responses, *t* test $p < 0.01$, effect size $\mu = 0.9$; for individual source waveforms, see Supplementary Fig. 2). ROI analysis of the maximum MNE maps over the right temporoparietal region gave consistent results (average over the time window of 195–245 ms in the incongruent vs. congruent conditions, *t* test $p < 0.03$, effect size $\mu = 0.7$; Fig. 5D).

"Ambiguous" sounds

For testing the categorical adaptation effects in two different contexts ("voice" and "cat"), we used acoustically identical target sounds that were derived from the voice sounds (see Methods). Whereas the N100m responses to these ambiguous sounds presented after cat and voice stimuli did not differ from each other, the right-hemispheric responses peaking at 265 ± 28 ms were statistically significantly stronger to the target sounds presented after the voice than cat stimuli as

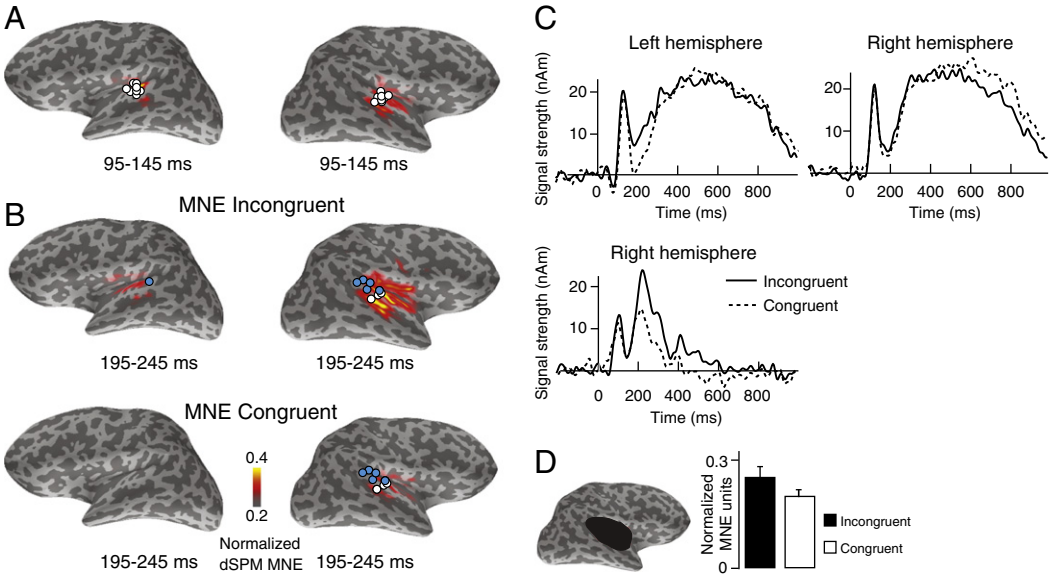

Fig. 5. MEG group-level data. The locations of the ECDs used to model the N100m responses (A, white dots), and the 250-ms responses (B, blue dots) in congruent and incongruent conditions in all subjects, morphed and superimposed on the average MNE dSPM distributions. Note that in three subjects, the same ECD was used to model both the N100m and the 250-ms response in the right hemisphere (white dots in B). (C) ECD time courses from −200 to 1000 ms with respect to the stimulus onset. (D) MNE ROI analysis on the mean activation over the marked cortical area in the time window of 195–245 ms.

modelled by the ECDs (estimated from a 50-ms time window around the individual peak responses, t test $p<0.02$, effect size $\mu=1.1$; see Fig. 6).

Discussion

In the present study, we investigated the temporal processing of auditory categories by utilizing carefully matched human and cat vocalizations. In particular, we used a paradigm that enabled us to compare responses to physically identical stimuli presented in different categorical contexts. Our results demonstrate that, when the low-level auditory stimulus differences are minimized, responses specifically at the right temporoparietal cortex react vigorously to auditory categorical violation regardless of the stimulus category at ~200 − 250 ms after the stimulus onset.

Although our experimental stimuli were matched for several temporospectral acoustic characteristics, for the easily recognizable stimuli, the overall harmonic structures still differed enough to provide cues needed for successful online categorization of the sounds. The conspicuous auditory N100m responses can be evoked by any sound onset or change in the auditory environment, but they also indicate stimulus-specific neural activity (Hari, 1990). Indeed, in the left hemisphere the first N100m responses for a stimulus block were stronger for voice than cat sounds, probably reflecting the remaining acoustic differences between the sounds. This effect may be partly explained by the female voice stimuli containing more energy at the frequency level of 1000–1500 Hz than the cat vocalizations, although effect of stimulus bandwidth on cortical responses has been shown to be highly stimulus specific (Seithler-Preisler et al., 2003; Shahin et al., 2005; Soeta et al., 2005). Thus, the use of a paradigm that allowed us to present the exact same stimuli in different categorical contexts can be considered crucial for the interpretation of the results. The differences between the congruent and incongruent sounds at ~200 − 250 ms after sound onset in the right hemisphere, present regardless of the sound category, suggest that at this time window, auditory processing has proceeded to a stage at which categorical templates have been established. Previously, right-lateralized auditory cortical fMRI activation in response to species-specific vocalizations has been reported in humans and monkeys, mainly in the STG/STS region (Belin and Zatorre, 2003; Belin et al., 2002; Formisano et al., 2008; Petkov et al., 2008), and right-hemispheric STG/STS has recently been related to speaker-related changes in pitch that are needed for recognizing speech among changing speakers (von Kriegstein et al., 2010). Several earlier neuroimaging studies have pointed to functional asymmetries in the auditory areas, with the left and right auditory cortices being predominantly sensitive to temporal and spectral changes, respectively (e.g., Obleser et al., 2008; Zatorre and Belin, 2001). Our MEG results for categorizing vocalizations—for which rapid analysis of spectral information is crucial—are in agreement with these results, and further suggest the observed activity to support categorical processing at ~200 − 250 ms after sound onset.

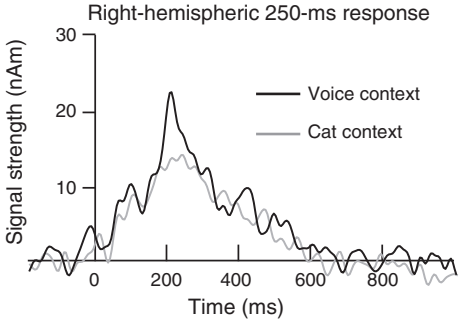

Fig. 6. The mean time courses of the right-hemispheric 250-ms responses for ambiguous sounds in cat and voice contexts, averaged over all subjects.

Recently, human vocalizations were demonstrated to evoke stronger responses than animal vocalizations at 169–219 ms after sound onset within the anterior right STG and STS, without topographical differences between stimuli (De Lucia et al., 2010). In the current experiment, after rather strict stimulus control for both acoustical features and attentional demands, auditory MEG responses to human voices and cat sounds did not statistically significantly differ from each other at around 200–250 ms. Rather, our results suggest the right posterior temporoparietal cortex to be especially activated in response to auditory categorical violation, regardless of the actual auditory stimulus. In 5 out of 8 subjects, the source for this response was separable from the source of the N100m response that has repeatedly been localized to the posterior part of the planum temporale (Hari, 1990). However, taking the relatively large interindividual variability in the 250-ms response source locations, they are likely to reflect anatomically more widespread synchronous activity, possibly including also the planum temporale that has earlier been suggested to be engaged in segregating and matching spectrotemporal patterns crucial for auditory object recognition (Griffiths and Warren, 2002). Combining electrophysiological measures e.g., with fMRI could in the future provide more detailed spatial information on these responses.

The 250-ms responses in the present study had a fairly similar polarity to the N100m responses, and their cortical sources were located at the near vicinity of those of N100m with right-hemisphere dominance. These sources are unlikely to reflect the well-established, broad positive component at ~300 ms (P300) evoked by infrequent task-relevant stimuli in EEG recordings, likely reflecting widespread activity with bilateral sources at occipitotemporal, centrotemporal, parietal and precuneal areas (Anurova et al., 2005). Rather, our 250-ms responses seem to overlap temporally and spatially with activity that has been observed, although bilaterally, in earlier auditory MEG studies on processing syllables, spoken words, and environmental sounds (Bonte et al., 2006; Renvall et al., 2012; Uusvuori et al., 2008). These responses do not seem to react to, e.g., phonetic or semantic task manipulations (Bonte et al., 2006; Uusvuori et al., 2008). Future studies are needed to explore whether these responses are related e.g., to accessing templates for different auditory categories regardless of stimulus type, possibly with different hemispheric emphasis for speech-like sounds.

The careful stimulus control can also be considered the main limitation of our present study: The stimuli were simple and they were constructed as continua from two exemplars. Even though their variability was increased by filtering and transposing them to different pitches, their ecological validity remains limited, compared with e.g., spoken words or environmental sounds. In future studies, the representation of auditory categories should be addressed also using more realistic auditory scenes, for example by modifying stimulus recognizability with varying level of superimposed noise (Renvall et al., 2012) and using a wider range of stimulus categories.

Although at the behavioral level the categorical context did not statistically significantly affect the categorization of ambiguous sound stimuli, the cortical responses to these sounds differed greatly depending whether they were presented after cat or voice sounds. Specifically, the right-hemispheric 250-ms responses were statistically significantly greater to sounds presented in the voice than cat context although the ambiguous sounds were acoustically closer to the voice stimuli. This finding could suggest that human voices as potentially more meaningful stimuli for the listener generated a stronger contextual effect, and thus resulted in a greater categorical mismatch for sounds that could not be unambiguously attributed to one of the two categories. This suggests a more established status for processing of human voices in the human auditory cortex than e.g., animal vocalizations (Fecteau et al., 2004). However, further studies are evidently needed for establishing the

complex interactions between context and target sounds. Specifically if the target sounds such as the ambiguous sounds here do not belong to any natural category, different cortical mechanisms may also apply.

In conclusion, our present results suggest that, after careful matching of acoustic stimulus features and behavioral demands, auditory categories for vocalizations are accessed by ~250 ms, preferably in the right posterotemporal cortex. This activity may reflect the detailed spectral analysis needed in the auditory categorical distinction of vocalizations.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.06.010.

Acknowledgments

We thank Niclas Kilian-Hütten and Jasper van den Bosch for help with the behavioral measurements, Mia Illman for the surface reconstructions, and Jan Kujala, Miiamaaria Kujala, Lauri Parkkonen and Tiina Parviainen for comments on the manuscript. This work was supported by the Academy of Finland (National Centers of Excellence Programme 2006–2011, and grant numbers 213828 and 127401 to HR), Netherlands Organisation for Scientific Research, Helsingin Sanomat Centennial Foundation, Emil Aaltonen Foundation and The Ella and Georg Ehrnrooth Foundation.

References

Alain, C., Arnott, S., Hevenor, S., Graham, S., Grady, C., 2001. “What” and “where” in the human auditory system. Proc. Natl. Acad. Sci. U. S. A. 98, 12301–12306.
Altmann, C., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y., Kakigi, R., 2008. Temporal dynamics of adaptation to natural sounds in the human auditory cortex. Cereb. Cortex 18, 1350–1360.
Anurova, I., Artchakov, D., Korvenoja, A., Ilmoniemi, R.J., Aronen, H.J., Carlson, S., 2005. Cortical generators of slow evoked responses elicited by spatial and nonspatial auditory working memory tasks. Clin. Neurophysiol. 116, 1644–1654.
Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105–2109.
Belin, P., Zatorre, R., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Belin, P., Zatorre, R., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.
Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.
Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. Proc. Inst. Phon. Sci. 17, 97–110.
Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot Int. 5, 341–345.
Bonte, M., Parviainen, T., Hytönen, K., Salmelin, R., 2006. Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. Cereb. Cortex 16, 115–123.
Charest, I., Pernet, C., Rousselet, G., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J., Belin, P., 2009. Electrophysiological evidence for an early processing of human voices. BMC Neurosci. 10, 127.
Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.
Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67.
De Lucia, M., Clarke, S., Murray, M., 2010. A temporal hierarchy for conspecific vocalization discrimination in humans. J. Neurosci. 30, 11210–11221.
Fecteau, S., Armony, J., Joanette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage 23, 840–848.
Fischl, B., Sereno, M., Tootell, R., Dale, A., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.
Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322, 970–973.
Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360, 815–836.
Garrido, M.I., Killner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: A reproducibility study. Neuroimage 36, 571–580.
Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Killner, J.M., 2008. The functional anatomy of the MMN: A DCM study of the roving paradigm. Neuroimage 42, 939–944.
Griffiths, T., Warren, J., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.
Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497.
Hansen, P.C., Kringelbach, M.L., Salmelin, R. (Eds.), 2010. MEG - An introduction to methods. Oxford UP, New York.
Hari, R., 1990. The neuromagnetic method in the study of the human auditory cortex. In: Grandori, F., Hoke, M., Romani, G.L. (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials. Karger, Basel, pp. 222–282.
Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. U. S. A. 101, 6809–6814.
Kaas, J., Hackett, T., 1999. ‘What’ and ‘where’ processing in auditory cortex. Nat. Neurosci. 2, 1045–1047.
Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Janik, J.J., DeYoe, E.A., 2005. Distinct cortical pathways for processing tool versus animal sounds. J. Neurosci. 25, 5148–5158.
Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31, 160–171.
Medendorp, W., Kramer, G., Jensen, O., Oostenveld, R., Schoffelen, J., Fries, P., 2007. Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. Cereb. Cortex 17, 2364–2374.
Murray, M., Camen, C., Gonzalez Andino, S., Bovet, P., Clarke, S., 2006. Rapid brain discrimination of sounds of objects. J. Neurosci. 26, 1293–1302.
Näätänen, R., 1992. Attention and brain function. Lawrence Erlbaum Associates, Hillsdale, NJ.
Obleser, J., Eisner, F., Kotz, S., 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J. Neurosci. 28, 8116–8124.
Petkov, C., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N., 2008. A voice region in the monkey brain. Nat. Neurosci. 11, 367–374.
Rauschecker, J., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806.
Renvall, H., Formisano, E., Parviainen, T., Bonte, M., Vihla, M., Salmelin, R., 2012. Parametric merging of MEG and fMRI reveals spatiotemporal differences in cortical processing of spoken words and environmental sounds in background noise. Cereb. Cortex 22, 132–143.
Romanski, L., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P., Rauschecker, J., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2, 1131–1136.
Salmelin, R., Hari, R., Lounasmaa, O.V., Sams, M., 1994. Dynamics of brain activation during picture naming. Nature 368, 463–465.
Seithler-Preisler, A., Krumbholz, K., Lütkenhöner, B., 2003. Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery? Audiol. Neurootol. 8, 322–337.
Shahin, A., Roberts, L.E., Pantev, C., Trainor, L.J., Ross, B., 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport 16, 1781–1785.
Soeta, Y., Nakagawa, S., Tonoike, M., 2005. Auditory evoked magnetic fields in relation to bandwidth variations of bandpass noise. Hear. Res. 202, 47–54.
Staeren, N., Renvall, H., De Martino, F., Goebel, R., Formisano, E., 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr. Biol. 19, 498–502.
Uusvuori, J., Parviainen, T., Inkinen, M., Salmelin, R., 2008. Spatiotemporal interaction between sound form and meaning during spoken word perception. Cereb. Cortex 18, 456–466.
Vartiainen, J., Parviainen, T., Salmelin, R., 2009. Spatiotemporal convergence of semantic processing in reading and speech perception. J. Neurosci. 29, 9271–9280.
von Kriegstein, K., Smith, D., Patterson, R., Kiebel, S., Griffiths, T., 2010. How the human brain recognizes speech in the context of changing speakers. J. Neurosci. 30, 629–638.
Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., Dehaene, S., 2011. Evidence for hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. U. S. A. 108, 20754–20759.
Warren, J., Griffiths, T., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J. Neurosci. 23, 5799–5804.
Warren, J., Scott, S., Price, C., Griffiths, T., 2006. Human brain mechanisms for the early analysis of voices. Neuroimage 31, 1389–1397.
Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953.