

Parametric Merging of MEG and fMRI Reveals Spatiotemporal Differences in Cortical Processing of Spoken Words and Environmental Sounds in Background Noise

Hanna Renvall^{1,2}, Elia Formisano¹, Tiina Parviainen^{2,3}, Milene Bonte¹, Minna Vihla² and Riitta Salmelin²

¹Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, the Netherlands, ²Brain Research Unit, MEG Core and Advanced Magnetic Imaging Center, Low Temperature Laboratory, Aalto University, FI-00076 Aalto, Finland and ³Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK

Address correspondence to Hanna Renvall, Brain Research Unit, Low Temperature Laboratory, Aalto University, PO Box 15100, FI-00076 Aalto, Finland. Email: hanna@neuro.hut.fi.

There is an increasing interest to integrate electrophysiological and hemodynamic measures for characterizing spatial and temporal aspects of cortical processing. However, an informative combination of responses that have markedly different sensitivities to the underlying neural activity is not straightforward, especially in complex cognitive tasks. Here, we used parametric stimulus manipulation in magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) recordings on the same subjects, to study effects of noise on processing of spoken words and environmental sounds. The added noise influenced MEG response strengths in the bilateral supratemporal auditory cortex, at different times for the different stimulus types. Specifically for spoken words, the effect of noise on the electrophysiological response was remarkably nonlinear. Therefore, we used the single-subject MEG responses to construct parametrization for fMRI data analysis and obtained notably higher sensitivity than with conventional stimulus-based parametrization. fMRI results showed that partly different temporal areas were involved in noise-sensitive processing of words and environmental sounds. These results indicate that cortical processing of sounds in background noise is stimulus specific in both timing and location and provide a new functionally meaningful platform for combining information obtained with electrophysiological and hemodynamic measures of brain function.

Keywords: auditory, multimodal, speech

Introduction

Although electrophysiological and hemodynamic brain imaging techniques measure different aspects of brain function, they are mainly assumed to reflect the same neuronal activity. Indeed, magnetoencephalography (MEG) evoked responses and functional magnetic resonance imaging (fMRI) blood-oxygenation-level-dependent (BOLD) responses have been demonstrated to spatially coincide in early sensory and sensorimotor processing (Korvenoja et al. 1999; Sharon et al. 2007; Raij et al. 2010). Approaches to combine electrophysiological and hemodynamic responses have so far concentrated on deducing the timing information from MEG (or electroencephalography [EEG]) and the location from fMRI (Dale et al. 2000). However, MEG and fMRI are likely to be sensitive to different features of the underlying neuronal activity, and discrepancies can be expected especially in complex cognitive tasks (Vartiainen et al. 2011). Meaningful combined use of MEG and fMRI in such studies is thus greatly

needed for understanding the correspondences between the 2 methods and their optimal merging.

In the current study, we applied parametric noise manipulation on 2 categories of natural sounds—spoken words and environmental sounds—and used the same experimental design in MEG and fMRI on the same subjects. In daily life, communication takes place among various sources of noise. Yet, the underlying neural mechanisms of auditory perception in such conditions remain poorly understood. For speech sounds, background noise compromises recognition, and it diminishes and delays electroencephalographic auditory cortical responses (Martin et al. 1997, 1999; Whiting et al. 1998; Cunningham et al. 2001; Müller-Gass et al. 2001; Kozou et al. 2005; Martin and Stapells 2005; Kaplan-Neeman et al. 2006), even when speech discrimination scores remain high (Martin et al. 1997, 1999; Whiting et al. 1998; Martin and Stapells 2005; Kaplan-Neeman et al. 2006). The interrupting effects of noise have largely been attributed to peripheral mechanisms. However, word recognition in noise is facilitated by contextual information (Bradlow and Alexander 2007) and musical experience (Parbery-Clark et al. 2009), suggesting the involvement of central processes as well. Few studies have addressed the effect of varying signal-to-noise ratio (SNR) on auditory evoked responses. During continuous background noise, decreasing SNR uniformly delayed responses to both speech sounds and pure tones (Kaplan-Neeman et al. 2006; Billings et al. 2009), whereas effects on response amplitudes were less systematic.

In MEG, different stages of speech sound processing are reflected in the same general area in the supratemporal cortex, in several time windows (for a review, see Salmelin 2007). Acoustic-phonetic feature analysis shows at the nonprimary auditory cortices ~50 to 100 ms after the stimulus onset (Poeppel et al. 1996; Obleser et al. 2004). At ~150 to 200 ms, responses to native versus nonnative phonemes differ from each other (Näätänen et al. 1997), indicating language-specific phonetic-phonological analysis. From ~200 ms onwards, a prominent sustained response sensitive to lexical-semantic manipulation can be recorded from the left temporal cortex (Helenius et al. 2002; Bonte et al. 2006).

The inherent limitations of spatial information in MEG can be complemented by fMRI to provide more detailed insights to the spatial architecture of the activity. fMRI studies have demonstrated sensitivity to speech in the left superior and middle temporal gyri (STG, MTG) (Binder et al. 2000; Benson et al. 2001; Vouloumanos et al. 2001; Davis and Johnsrude 2003) and superior temporal sulcus (STS) (Benson et al. 2006). Furthermore, anterior areas in the left STS are activated by intelligible

speech (Scott et al. 2000), syllables (Liebenthal et al. 2005), and vowels (Obleser et al. 2006). It has also been suggested that speech perception emerges from integrated activation of areas processing both nonspeech and speech sounds (Price et al. 2005) and with relevant contribution from the “early” auditory areas (Formisano et al. 2008).

Environmental sounds are meaningful but nonlinguistic complex auditory stimuli that share many spectrotemporal characteristics with speech (Gygi et al. 2004; Dick et al. 2007). Comprehension of the 2 stimulus types seems to develop similarly during childhood (Cummings et al. 2008, 2009). Environmental sounds can prime semantically related words (Van Petten and Rheinfielder 1995), and—similarly to words—their processing is modulated by contextual cues (Ballas and Howard 1987). The use of speech and environmental sounds can thus reveal not only cortical activity that is common to processing of different real-life sounds but also activations that are specific to each stimulus category. Earlier electrophysiological studies on responses to speech versus environmental sounds have concentrated on the late time windows (400–600 ms) of semantic processing and suggested largely overlapping networks for processing of the 2 types of sounds (Van Petten and Rheinfielder 1995; Cummings et al. 2006).

The present study served 2 purposes. First, we explored new means of combining information from electrophysiological and hemodynamic measures of brain function. The parametric stimulus manipulation allowed us to combine the results in an optimal fashion: The random-effect general linear models (GLMs) for the fMRI analysis were specified on the basis of individual neurophysiological reactivity to the added noise as measured by MEG. Second, as a step toward increasingly natural experimental paradigms, we investigated the effect of noise on cortical processing of real-life sounds, that is, spoken words and environmental sounds.

Materials and Methods

Subjects

We studied, with informed consent, 10 Finnish-speaking adults (mean \pm standard error of the mean [SEM] age 28 ± 1 years; 4 females, 6 males; all right-handed). None of the subjects had a history of hearing or neurological impairments. All subjects participated in both MEG and fMRI experiments. The study had a prior approval of the Ethical Committee of the Helsinki Uusimaa Hospital District.

Stimuli

Our stimuli consisted of 50 different spoken words and 50 environmental sounds (mean [SEM] durations 824 ± 13 and 832 ± 12 ms, respectively; the stimuli are listed in Supplementary Table S1). The spoken words were consonant-initial, 8-letter common nouns, pronounced by 4 speakers (2 males and 2 females), and they were recorded in an acoustically shielded room. The environmental sounds were collected from the internet and comprised, for example, sounds of tool use, animal cries, and traffic sounds. All the stimuli were sampled at 22 050 Hz (16-bit, mono) and they had rise and fall times of 25 ms. During the MEG and fMRI experiments, the sounds were delivered to the subjects binaurally at a comfortable listening level through plastic tubes and ear pieces.

All the stimuli were low-pass filtered at 4 kHz. In order to achieve SNRs of +18 dB, 0 dB, –9 dB, –12 dB, and –18 dB, the stimuli were embedded in white noise (low-pass filtered at 4 kHz), keeping the overall root-mean-square value unchanged. At the SNR level of +18 dB, the sounds were clearly distinguishable and at –18 dB nondistinguishable; the behavioral threshold fell between these levels.

Our pilot study showed that the most prominent changes in the cortical responses occurred when the SNR was decreased from +18 dB (n_1) via 0 dB (n_2) to –9 dB (n_3), and these 3 noise levels were included in the main experiment.

To characterize the acoustical properties of the 2 sound categories, sound intensity envelopes, harmonics-to-noise ratios (HNRs) and spectra were examined in more detail with the freely available Praat software (www.fon.hum.uva.nl/praat/). Sound intensities were analyzed in 20-ms steps at 0–140 ms from the beginning of the stimuli; the intensities did not differ statistically significantly between stimulus categories at any of the noise levels ($P > 0.14$ – 0.63). HNR reflects the degree of periodicity of the sound (Boersma 1993), and it is quantified as the energy of the periodic (harmonic) component of the signal over time relative to the remaining “noise” signal. For natural sounds with time-varying characteristics, calculation of the HNR value is not straightforward (Riede et al. 2001; Lewis et al. 2009). The obtained HNRs, although greater for spoken words due to the harmonic nature of vocalizations (Riede et al. 2001), were generally low for all stimuli because of the embedded noise. The spectra of the sounds were estimated in 50-Hz frequency steps. On average, the spectra for environmental sounds were rather flat at 0–4000 Hz at all noise levels, whereas the speech sounds contained more energy at 0–700 Hz than at higher frequencies; this difference was diminished with the increasing noise level (Supplementary Fig. S1). Although the speech and environmental stimuli, as such, could not be fully matched acoustically, the parametric addition of noise was identical for the 2 types of stimuli.

Experimental Setup

The MEG and fMRI experimental paradigms were adjusted to be as similar as possible. In both experiments, the sounds were presented in trains of 4 with stimulus-onset-asynchrony (SOA) of 3600 ms (Supplementary Fig. S2). In 8% of the trials, a question mark appeared after the last stimulus, and the subject was required to respond by lifting her index or middle finger (“the last 2 stimuli were the same/not the same”). The response hand was alternated across subjects, and the target trials were discarded from the analysis. The stimuli within a train were all either spoken words or environmental sounds at the same SNR level, resulting in altogether 6 experimental conditions. Stimulus trains were presented in a random order and the same condition did not repeat immediately. The order of the fMRI and MEG experiments was counterbalanced across subjects.

In our pilot MEG study with 10 subjects, speech and environmental sounds at 5 different noise levels were presented within the same sequence, in a random order. The interstimulus interval was 1500–1800 ms, and the subject was asked to respond with a button press when the same sound appeared twice in a row. The main results of this pilot study (Supplementary Fig. S3), using an event-related design, were similar to those obtained with the block design that was used in the main experiment.

Behavioral Session

During the MEG and fMRI sessions, the behavioral responses were too scarce for statistical inference. Therefore, in a separate behavioral session, the subjects listened to pairs of stimuli, and they were asked to respond with a button press whether the 2 sounds embedded in noise were the same. The stimuli were the same as in the fMRI/MEG experiments. Altogether 190 sound pairs were presented, with a SOA of 2 s and an interpair interval of ~3 s.

MEG Experiment and Signal Analysis

In the MEG experiment, all 50 spoken words and 50 environmental sounds were presented twice. Each trial was followed by a rest of 13 s, to mimic the baseline condition in the fMRI experiment.

Auditory evoked fields were recorded in a magnetically shielded room while the subject was sitting with the head supported against the helmet-shaped bottom of the 306-channel Vectorview (Elekta NeuroMag, Helsinki, Finland) neuromagnetometer. The device contains 102 identical triple sensors, comprising 2 orthogonal planar first-order gradiometers and one magnetometer, each of them coupled to a SQUID

(Superconducting QUantum Interference Device). Four head-position-indicator coils were attached to the scalp, and their positions were measured with a 3D digitizer; the head coordinate frame was anchored to the 2 periauricular points and the nasion. The head position with respect to the sensor array was determined by briefly feeding current to the marker coils before the actual measurement.

The MEG signals were band-pass filtered at 0.03–200 Hz, digitized at 600 Hz and averaged online from 300 ms before the stimulus onset to 1500 ms after it, setting as baseline the 200-ms interval immediately preceding the stimulus onset. The averaged signals were digitally low-pass filtered at 40 Hz. The horizontal and vertical electrooculograms were recorded to discard data contaminated by eye blinks and movements. The responses from the second to the fourth sounds in the trains were all averaged together; a minimum of 60 artifact-free responses were collected per condition. The experiment was conducted in 4–5 blocks, each lasting ~10 min.

For source-level analysis, the head was modeled as a homogeneous sphere. The model parameters were optimized for the intracranial space obtained from MR images that were available for all subjects. The neurophysiological effects of varying noise were analyzed by first segregating the recorded sensor-level signals into separable cortical-level spatiotemporal components, by means of guided current modeling (equivalent current dipole [ECD]; Hämäläinen et al. 1993), separately for each subject. The model parameters of an ECD represent the location, orientation, and strength of the net current in the activated brain area. Only ECDs explaining more than 85% of the local field variance during the response peaks were accepted in the model. Based on this criterion, 3–4 spatiotemporal components were selected in the models of each subject. The components were identified in the condition with the strongest signals, typically the *n1* stimuli, and the same ECDs explained well the responses in the other conditions. The components explaining the field patterns around 100 and 400 ms were very similar and, to prevent interactions between these ECDs, a single component was used to model both responses. The same component accounted for the opposite current flow at ~200 ms. The location parameters of the ECDs were transformed according to Talairach and Tournoux (1988) for comparison with the loci of the mean noise-sensitive fMRI activity.

Due to considerable interindividual variability in the shape and duration of the deflection at ~200 ms, its strength was estimated as the mean amplitude in the 100–300 ms time window after stimulus onset. For consistency, the 100-ms and 250-ms response amplitudes were estimated as mean values from +50% to –50% of the maximum activation (time windows [mean ± SEM] from 76 ± 4 to 149 ± 9 ms and from 158 ± 7 to 344 ± 22 ms, respectively). The same average time window per subject was used for all noise levels and both stimulus categories. The mean amplitude of the 400-ms response was calculated in the 300–1500 ms time window.

Statistical Analyses on the MEG Signals

A repeated-measures analysis of variance (ANOVA) with Stimulus type, Hemisphere and Noise level as within-subjects factors was used for the statistical analysis of the main effects for source strengths and response latencies. Subsequently, responses to spoken words and environmental sounds in different cortical areas and time windows were tested for linear trends with repeated-measures ANOVA (within-subjects factor Noise level). Paired *t*-tests were used for the statistical analysis of source locations and orientations.

fMRI Experiment and Signal Analysis

To minimize the possible masking effect of scanner noise, the sounds were presented during 1600-ms silent periods between 2000-ms scans (Shah et al. 2000; Jäncke et al. 2002; van Atteveldt et al. 2004; Staeren et al. 2009); the sounds started 400 ms after the beginning of silence (cf. Supplementary Fig. S2). Each trial was followed by a 14.4-s rest interval. A subset of 24 sound stimuli per category (duration 829 ± 15 and 835 ± 17 ms for speech and environmental sounds, respectively) were presented twice during 2 experimental runs, each consisting of 6 trials per condition (altogether 36 trials) and 3 target trials.

The imaging was performed with a whole-body General Electric Signa EXCITE 3.0-T scanner and an 8-channel high-resolution brain coil (GE Healthcare, Chalfont St. Giles, UK). In each subject, 2 runs of 319 volumes were acquired with a gradient-echo echo planar imaging sequence (field of view = 200×200 mm², time of repetition = 2000 ms, time to echo = 30 ms, flip angle = 75°, 34 slices with slice thickness of 3.0 mm with no gap, number of excitations 1, the acquisition matrix size 64×64). Structural MRIs were obtained from each subject with a standard spoiled-gradient-echo sequence after the functional runs.

The functional and anatomical images were analyzed with BrainVoyager QX (Brain Innovation, Maastricht, the Netherlands). Preprocessing consisted of slice scan-time correction, linear trend removal, temporal high-pass filtering (cutoff 5 cycles per time course), and 3D motion correction. No spatial smoothing was applied to the fMRI data. Functional slices were coregistered to the anatomical data, and both data sets were normalized to the Talairach space (Talairach and Tournoux 1988).

Statistical Analyses on the fMRI Signals

The fMRI time series were analyzed using 4 differently specified multisubject 2-level random-effect GLMs. In the first GLM, all noise levels were grouped together resulting in predictors for spoken words and environmental sounds. In the second GLM, stimulus SNR levels (*n1*, *n2*, and *n3*) were used as additional predictors. The GLM consisted of the 2 previous predictors explaining the overall responses to spoken words and environmental sounds, and 2 additional normalized and orthogonalized predictors (with values 1, –0.2, and –0.8) explaining the SNR-based parametric variation between responses at different noise levels (Bandettini et al. 1993; Riecke et al. 2007), separately for each stimulus category. This procedure allowed for studying the conjunction of response to spoken words/environmental sounds and to the parametrically varying noise. In the third GLM, the 2 noise-sensitive predictors were formed on the basis of the group MEG results, and in the fourth GLM, on the basis of individual MEG responses. The predictors were obtained by transforming the statistically significant group/individual MEG response strengths, separately for each subject and hemisphere and separately for the spoken word and environmental sound data, to 3 parameters corresponding to the different noise levels in a similar manner as the SNR-based parametrization above.

For the second to fourth GLM, maps of group-level repeated-measures *t* statistics were calculated as a conjunction of the “response to spoken word/environmental sound” predictor and the “stimulus noise-level/MEG response strength” predictor using the fitted model parameters (i.e., individual β values) that were derived from the first-level analysis. The time courses of the predictors were adjusted for the hemodynamic delay by convolving them with a hemodynamic response function (Boynton et al. 1996).

Differences in noise sensitivity between spoken word and environmental sound stimuli were tested as a conjunction of “response to both spoken words and environmental sounds” and contrast “spoken words (environmental sounds) > environmental sounds (spoken words)” of the noise-sensitive predictors that resulted in the highest *T* values in the preceding analysis.

The statistical analysis of the data used a cluster-threshold method (Goebel et al. 2006). A voxel-level threshold was first set to $t_0 = 3.7$ ($P < 0.005$, uncorrected) for the main analysis and to $t_0 = 3.3$ ($P < 0.01$, uncorrected) for testing the noise sensitivity. Subsequently, all the maps were corrected on the basis of their spatial smoothness and a Monte Carlo simulation (1000 iterations) that estimated the cluster-level false positives. After this procedure, the minimum cluster size that yielded a false-positive rate $\alpha = 0.05$ was applied to the statistical maps. The multiple comparison-corrected maps were superimposed on one subject’s Talairach-transformed volume.

Results

Behavioral Results

Prior to the MEG and fMRI experiments, all subjects underwent a short behavioral session during which they listened to pairs of

stimuli and decided whether the 2 sounds embedded in noise were the same. The reaction times were prolonged with the increasing noise level (ANOVA repeated measures, test for linear trend, $F_{1,9} = 7.5$, $P < 0.03$); no significant differences were found between the stimulus types (Stimulus type \times Noise level interaction, $F_{1,9} = 0.4$, $P > 0.5$). The percent correct for all spoken word categories and for the $n1$ and $n2$ environmental sound categories was $\geq 98 \pm 1\%$ (mean \pm SEM), whereas the $n3$ environmental sounds were statistically significantly more difficult (percent correct $87 \pm 3\%$; $P = 0.003$ compared with 100%). The observed dissociation of reaction time and recognition accuracy is in agreement with the results of Kaplan-Neeman et al. (2006) on speech sounds embedded in noise.

MEG: Sensor Signals

For both spoken words and environmental sounds, the strongest responses occurred bilaterally over the auditory cortices (Fig. 1). A strong transient response at about 100 ms was followed by a longer-lasting activation, with the maximum

at about 400 ms after the sound onset. The responses continued until 1200–1400 ms. When the spoken words contained more noise, responses were diminished in amplitude at around 100–500 ms after the stimulus onset, bilaterally. In the left hemisphere (LH), the 400-ms response to spoken words also peaked later and was of longer duration to the noisier $n3$ than the less noisy $n1$ stimuli. In the right hemisphere (RH), responses to environmental sounds showed a striking reduction with increasing noise at about 200–500 ms after the stimulus onset.

MEG: Cortical Sources

The ECDs for responses to spoken words and environmental sounds did not differ statistically significantly from each other either in location ($P > 0.3$ in all directions; mean differences 2, 4, and 1 mm in x , y , and z directions, respectively) or orientation ($P > 0.2$). The responses to both types of stimuli were thus well accounted for by the same set of ECDs.

The final models consisted typically of 2 ECDs in each hemisphere (Fig. 2). In agreement with previous studies (for

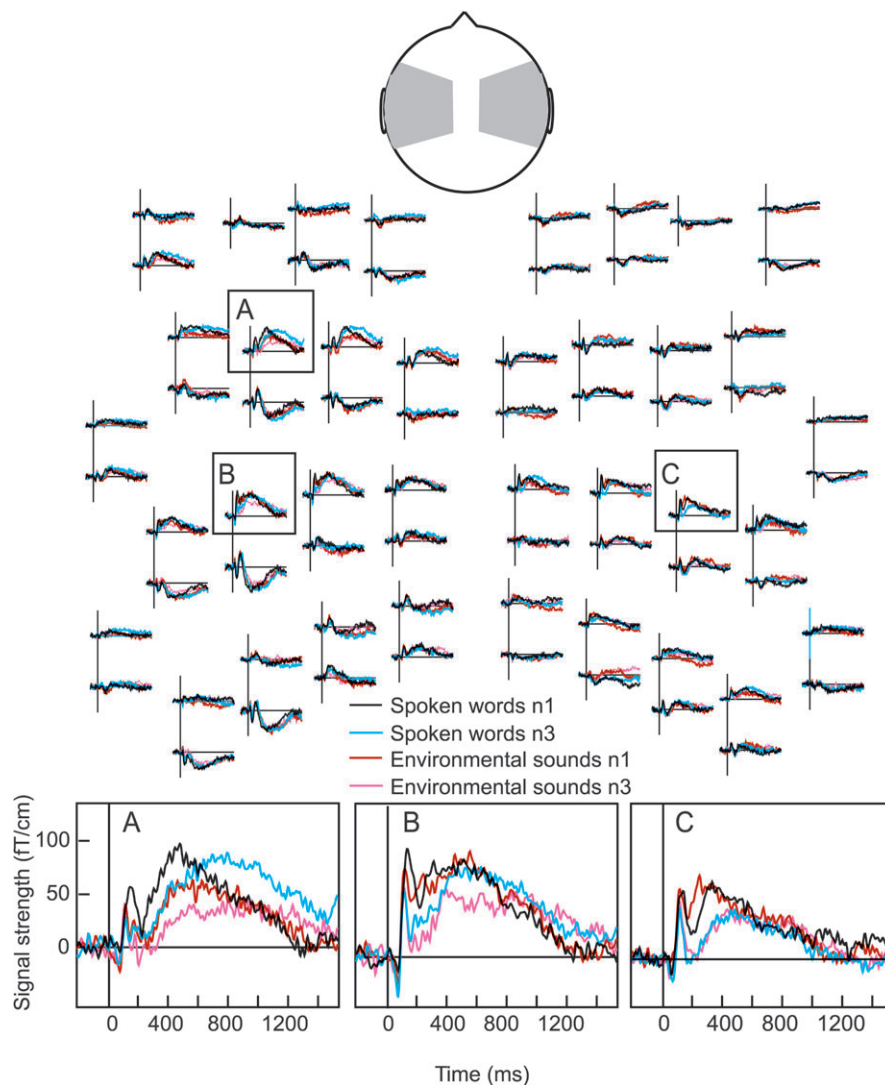


Figure 1. MEG signals. MEG responses of subject S4 to spoken words and environmental sounds at the noise levels $n1$ (+18 dB) and $n3$ (−9 dB). The upper and lower traces in each location correspond, respectively, to the latitudinal and longitudinal gradients of the radial magnetic field component B_z . The insets show enlarged responses from the sensors that detected the maximum signal over each hemisphere.

a review, see Hari 1990), the 100-ms responses (N100m) were adequately explained by one ECD in the left and another in the right supratemporal auditory cortex, with superior-inferior orientation. The same sources explained also the sustained activity peaking at around 400 ms that corresponds to the N400m responses reported in MEG studies of spoken word processing (Helenius et al. 2002; Bonte et al. 2006; Uusvuori et al. 2008). The N100m/N400m sources additionally accounted for the deflection at ~200 ms. In all subjects, the N100m/N400m ECDs were located in the anterior part of the planum temporale (PT) in both hemispheres (Fig. 3A).

In both hemispheres, the responses suggested a further current source around 250 ms, often with posterior-anterior orientation (cf. Bonte et al. 2006; Uusvuori et al. 2008). Across subjects, the 250-ms ECDs differed from the N100m/N400m ECDs in orientation, in both hemispheres (mean difference 130° in the left and 70° in the RH, $P < 0.05$), and their locations showed considerable interindividual variability (see Fig. 3A).

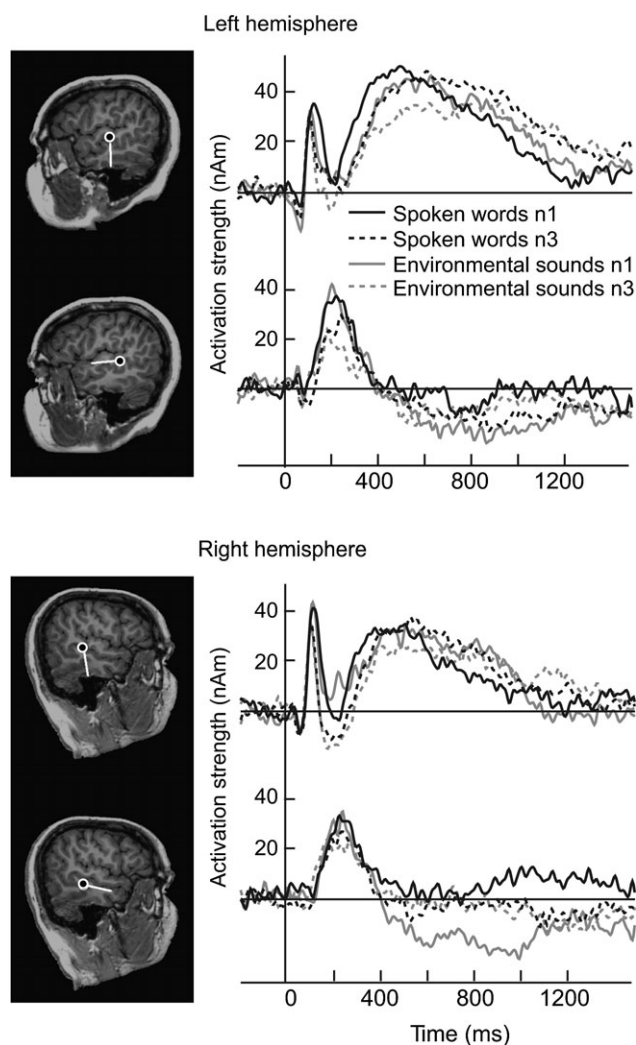


Figure 2. MEG source analysis in one subject. ECD locations (dots) and orientations (tails) superimposed on subject S4's MR images, with 2 source areas in each hemisphere. Neural activation strength in each area as a function time to spoken words (black traces) and environmental sounds (gray traces) at the noise levels $n1$ and $n3$.

MEG: Spoken Words versus Environmental Sounds

The main focus of this study was the effect of parametrically added noise, separately on spoken words and environmental sounds. Although the stimuli were not matched in all acoustic and behavioral dimensions (see Materials and Methods), the MEG responses to the 2 stimulus categories showed fairly similar activation strengths and latencies (Fig. 3), with a few exceptions. The left-hemispheric 200-ms deflections were overall more pronounced for environmental sounds than spoken words

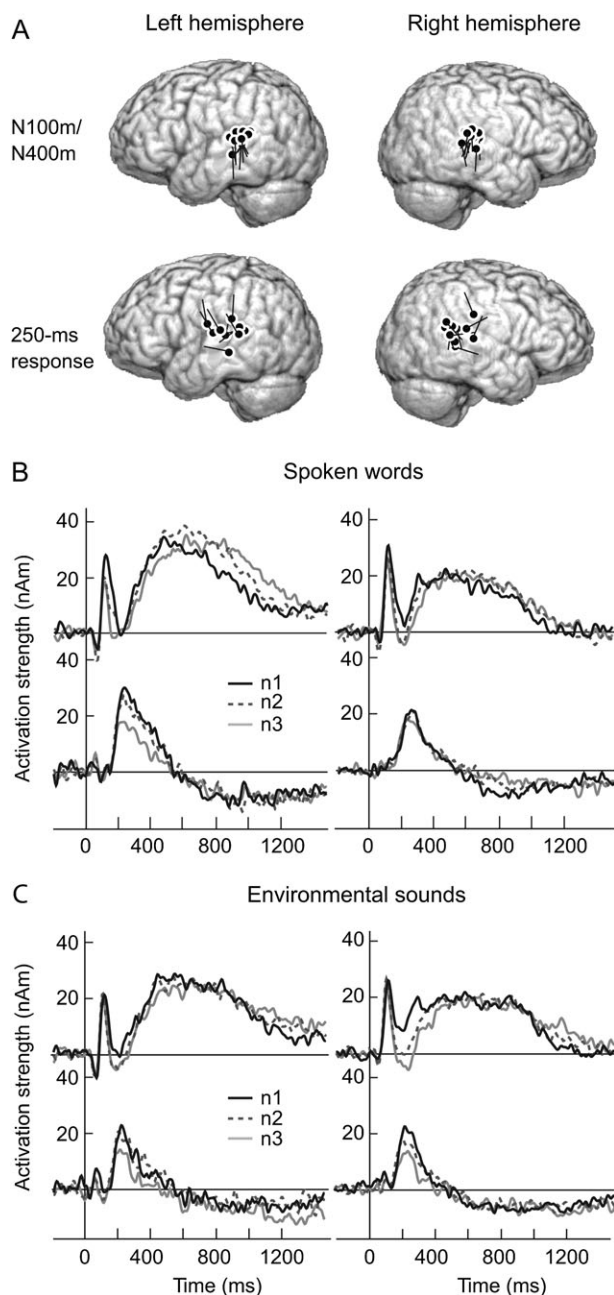


Figure 3. MEG group-level data. (A) Distribution of the N100m/N400m sources (top) and the 250-ms sources (bottom), displayed on an average brain. The individual centers of activation and directions of current flow are indicated as dots and tails, respectively. (B) Mean time courses of the N100m/N400m (top) and 250-ms sources (bottom) across all subjects for spoken words at all 3 noise levels. (C) Mean time courses of the N100m/N400m (top) and 250-ms sources (bottom) across all subjects for environmental sounds at all 3 noise levels.

(Stimulus type \times Hemisphere interaction, $F_{1,9} = 8.3$, $P < 0.02$). The left-hemispheric N400m responses, on the other hand, were stronger for spoken words than environmental sounds (Stimulus type \times Hemisphere interaction, $F_{1,9} = 19.1$, $P < 0.002$).

MEG: Effect of Noise Level on Responses to Spoken Words

The N100m response amplitudes decreased with the increasing noise level in both hemispheres (LH: $F_{1,9} = 16.7$, $P < 0.004$ for linear trend in contrasts; RH: $F_{1,9} = 17$, $P < 0.004$): the change was, on average, -50% in the LH (from $n1$ to $n3$) and -30% in the RH.

Shortening of peak latencies at higher noise levels was observed for the N100m response in the RH ($F_{1,9} = 26.6$, $P < 0.002$ for linear trend in contrasts) and for the 200-ms response in the LH ($F_{1,9} = 8.1$, $P < 0.02$), whereas the N400m peak latency and duration increased with the increasing noise level in the LH (peak: $F_{1,9} = 30.4$, $P < 0.001$; duration: $F_{1,9} = 15.3$, $P < 0.005$). The mean (\pm SEM) source strengths and response latencies are listed in Supplementary Table S2.

MEG: Effect of Noise Level on Responses to Environmental Sounds

The amplitude of the 200-ms response in the RH was reduced with the increasing noise level ($F_{1,9} = 44.6$, $P < 0.001$ for linear

trend in contrasts), whereas the N100m, 250-ms and N400m response amplitudes were not statistically significantly affected by the noise level in either hemisphere.

The N100m peak latency decreased in the LH ($F_{1,9} = 15.6$, $P < 0.004$ for linear trend in contrasts) with the increasing noise level.

fMRI: Stimulation versus Baseline

The 2 types of stimuli (across all noise levels) evoked spatially extended and generally overlapping BOLD activity in the bilateral temporal cortices relative to the baseline, extending from the Heschl's gyrus to planum polare anteriorly and to PT and STS posteriorly (overlap between word and environmental sound conditions 745 voxels in the RH and 743 voxels in the LH; see Fig. 4). Spoken words also activated the left premotor cortex. In the PT and STS of the LH, spoken words evoked stronger activation than the environmental sounds (contrast "spoken words $>$ environmental sounds" in 281 voxels, $P < 0.005$, $\alpha = 0.05$).

fMRI: Effect of Noise Using Stimulus-Based versus MEG-Based Predictors

The MEG results presented above suggested that the strength of the cortical response was not directly proportional to the stimulus noise level. Furthermore, the activation strengths showed considerable individual variation, especially for the left-

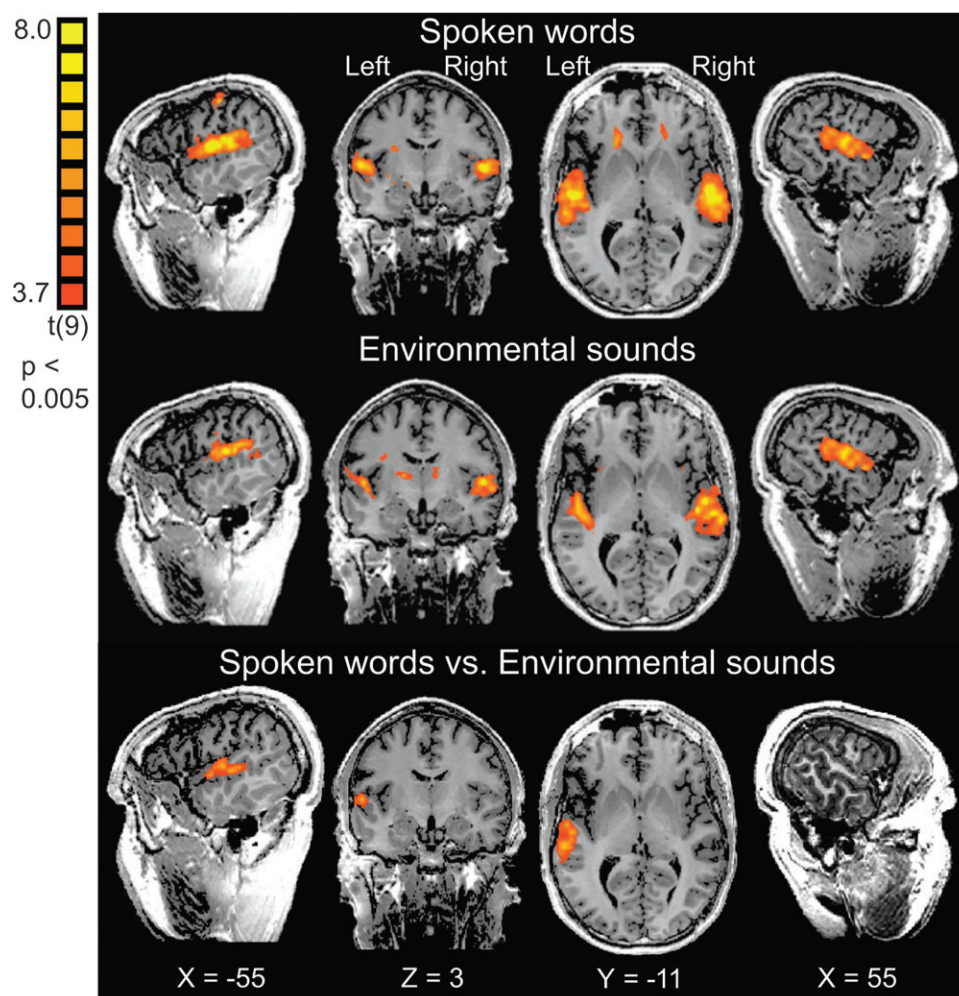


Figure 4. fMRI: task versus baseline. Brain regions activated in fMRI by spoken words (top) and environmental sounds (middle) relative to the baseline condition and spoken words versus environmental sounds (bottom; t maps, $P < 0.005$, cluster size corrected at $\alpha = 0.05$).

hemispheric responses to spoken words (see Fig. 5D). Thus, it may well be that relevant fMRI activations are not captured using the conventional stimulus-dependent predictors (Fig. 5A). An alternative, neurally based approach focused on the statistically significant MEG results on activation strengths (spoken words: LH and RH N100m responses; environmental sounds: RH 200-ms response). The MEG response strengths at different noise levels were transformed into response ratios according to, first, the group average of the MEG responses and, second, the individual MEG responses. After normalization and

orthogonalization, these ratios were used as predictors in 2 separate GLM models of the fMRI data (see Materials and Methods).

For the spoken words, the individual MEG data-based predictor calculated on the basis of the left-hemispheric N100m response strength had the highest T values and it revealed the largest activated areas (see Fig. 5 and Table 1). In the voxels unveiled by the stimulus-based predictors, use of the individual MEG-based predictors resulted in a further, significant increase of the mean T values (LH: 4.0 vs. 4.8, RH: 4.1 vs.

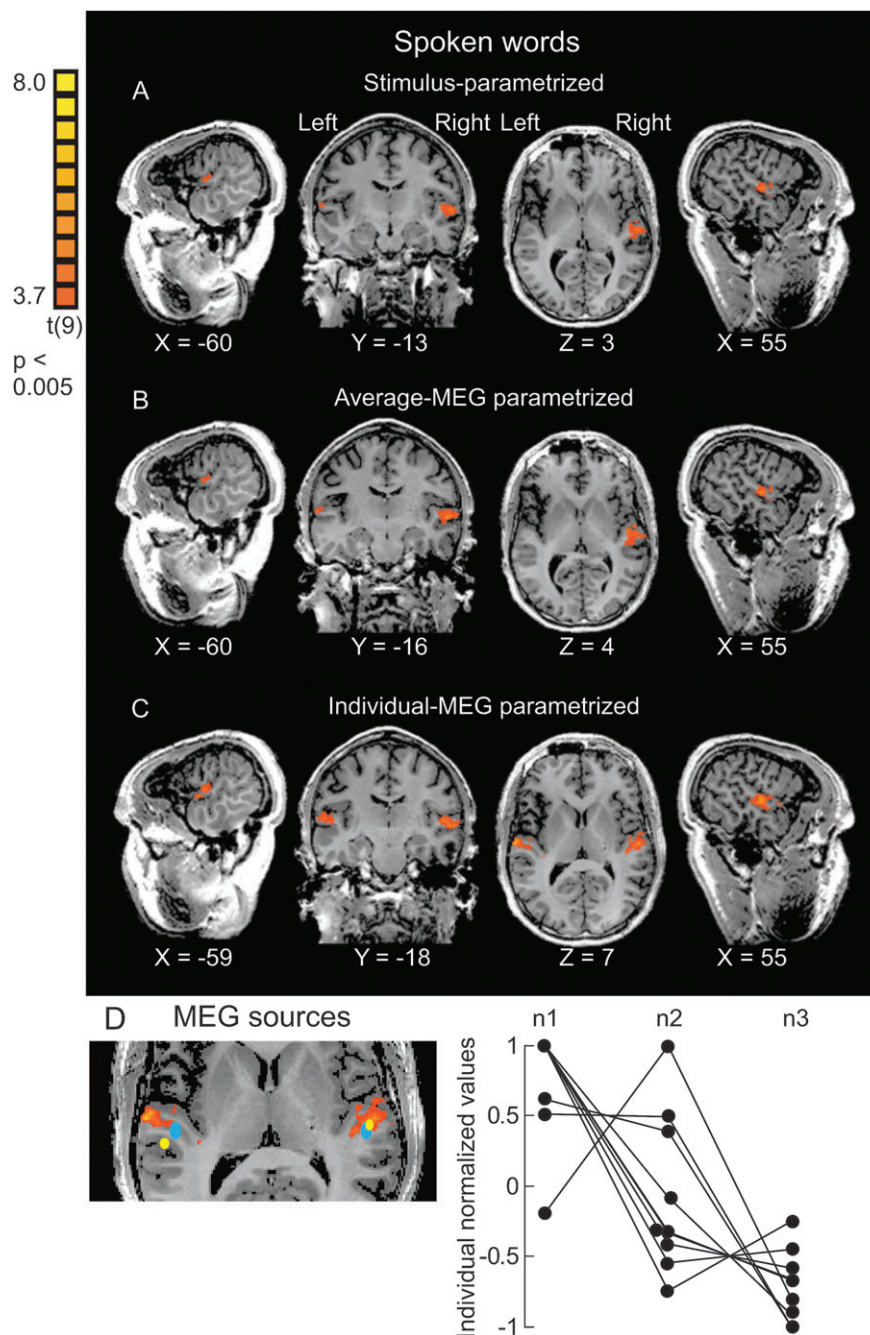


Figure 5. fMRI: noise on spoken words. The group volume maps of the responses to spoken words with predictors corresponding (A) to the stimulus noise levels, (B) to the averaged MEG data, and (C) to the individual MEG data, projected on to anatomical slices of one subject in Talairach space ($P < 0.005$, cluster size corrected at $\alpha = 0.05$). (D) The mean (\pm SEM) location of the sources of N100m/N400m (yellow) and 250-ms (blue) MEG responses represented by ECDs, superimposed on the group volume map (left), and the individual normalized MEG responses (LH N100m strength as a function of noise level) used in the analysis.

4.6 for stimulus-based vs. individual MEG-based predictors; Wilcoxon signed-rank test $P < 0.0005$).

The other predictor based on significant MEG effects, reflecting the right-hemispheric N100m response strengths, revealed practically the same activated areas but only in the RH. We additionally tested individual predictors that were based on the LH N400m response duration, but those did not reveal significant fMRI activation.

Using the individual MEG-based predictors, activation was detected in the left-hemispheric STG including PT and in the right-hemispheric PT and STS. The stimulus noise-level predictors revealed fMRI activity mainly in the right-hemispheric PT. The mean source locations of the N100m/N400m and 250-ms MEG responses were located within the active fMRI area in the RH, and within 6 mm of an active fMRI voxel in the LH (Fig. 5D and Supplementary Table S3).

For the environmental sounds, the MEG responses as a function of noise were more comparable between subjects than in the spoken word condition (see Figs 5D and 6D). In agreement with this observation, the results obtained with the stimulus-based and MEG-based predictors were fairly comparable to each other (Fig. 6), unlike for the spoken word condition (see Fig. 5 and Table 1).

For the environmental sound stimuli, the stimulus-based and average MEG-based predictors revealed fMRI activity bilaterally in the STG including Heschl's gyrus and PT and in the left STS. The individual MEG-based predictors detected basically the same active areas but failed to reveal activity bilaterally in the medial PT. The stimulus-based and average MEG-based predictors resulted in higher mean T values in the voxels unveiled by the individual MEG-based predictors (LH: 4.4 vs. 4.1, RH: 4.6 vs. 4.4 for stimulus-based/average MEG-based vs. individual MEG-based predictors; Wilcoxon signed-rank test $P < 0.0005$).

The maxima of group-level fMRI foci of noise-sensitive processing differed for spoken words and environmental sounds, suggesting the involvement of at least partly disparate areas in processing of the 2 sound types (see Fig. 7 and Supplementary Table S3).

Discussion

We studied the effect of parametrically varying noise on neural responses to spoken words and environmental sounds by measuring MEG and fMRI data in the same subjects. Both words and environmental sounds evoked a typical pattern of MEG responses ~100 to 600 ms after the stimulus onset. The added noise modulated both the recognizability of the stimuli and the timing and location of the cortical responses but differently for spoken words and environmental sounds. MEG

responses to spoken words were diminished at ~100 ms bilaterally in the supratemporal cortex and the later (>600 ms) sustained left-hemispheric MEG responses were prolonged with the increasing noise level. MEG responses to environmental sounds were affected by the noise mainly in the RH at ~200 ms after the stimulus onset. Conventional stimulus-based fMRI analysis revealed noise sensitivity in the superior temporal cortex, right lateralized for speech and bilaterally for environmental sounds.

The parametric design and the observed differences in electrophysiological cortical reactivity between stimulus categories were ideally suited for exploring new means of merging MEG and fMRI data. For spoken words, fMRI analysis based on individual MEG responses provided a spatially more extended but also more detailed substructure and suggested that bilateral noise-sensitive processing of spoken words and environmental sounds takes place in partly nonoverlapping areas of the temporal cortex.

Combining MEG/fMRI Analyses

Based on MEG, the N100m responses to spoken words displayed a markedly nonlinear response to variation of stimulus noise levels and high interindividual variability, particularly in the left temporal cortex. This was probably the reason why MEG-parametrized fMRI analysis was far more successful (as measured by high T values and number of active voxels) than the usual stimulus-based parametrization. The other significant MEG effect, sustained left-hemispheric long-latency (>600 ms) response that was prolonged as a function of noise level, did not find a correspondence in fMRI, possibly because of the long temporal integration window of fMRI BOLD responses.

The MEG responses used in the fMRI analysis of the environmental sounds (RH 200-ms response) displayed less interindividual variability and, thus, generally followed the stimulus noise levels more faithfully than the responses to spoken words. In this case, the exact choice of predictor had little influence. The stimulus-based and average MEG-based predictors were most successful, possibly due to their higher resistance to small variations that are inevitably present in defining individual predictors. In contrast to the right-lateralized effect for the environmental sounds in MEG, fMRI demonstrated statistically significant noise-sensitive processing also in the LH, possibly reflecting activity that is nonphase-locked to the stimuli and thus remains undetected by the evoked MEG responses.

Here, we compared measures of neural activation that are most frequently used in neurophysiological and hemodynamic brain mapping, namely, MEG evoked responses and fMRI BOLD signal. Earlier studies have suggested that oscillatory MEG activity, especially at high frequencies, may correlate more closely with BOLD responses but such findings have, so far, mostly been limited to the study of the visual cortex (e.g., Foucher et al. 2003; Brookes et al. 2005). However, as evoked responses are the most prominent measure of activation in most experimental designs and provide typically a good spatial coverage, they continue to be widely used in the MEG (and EEG) settings. Thus, the knowledge on their comparability to BOLD responses is highly relevant.

Stimulus parametrization and the use of MEG responses as a measure of cortical reactivity seem to provide a meaningful

Table 1

Number of activated fMRI voxels ($T \geq 3.7$; cluster level corrected) when using stimulus-based predictors and MEG-based predictors (average and individual predictors)

	LH	RH
Spoken words		
Stimulus-based predictors	16	125
Average MEG-based predictors	17	134
Individual MEG-based predictors	131	246
Environmental sounds		
Stimulus-based predictors	188	263
Average MEG-based predictors	181	289
Individual MEG-based predictors	153	188

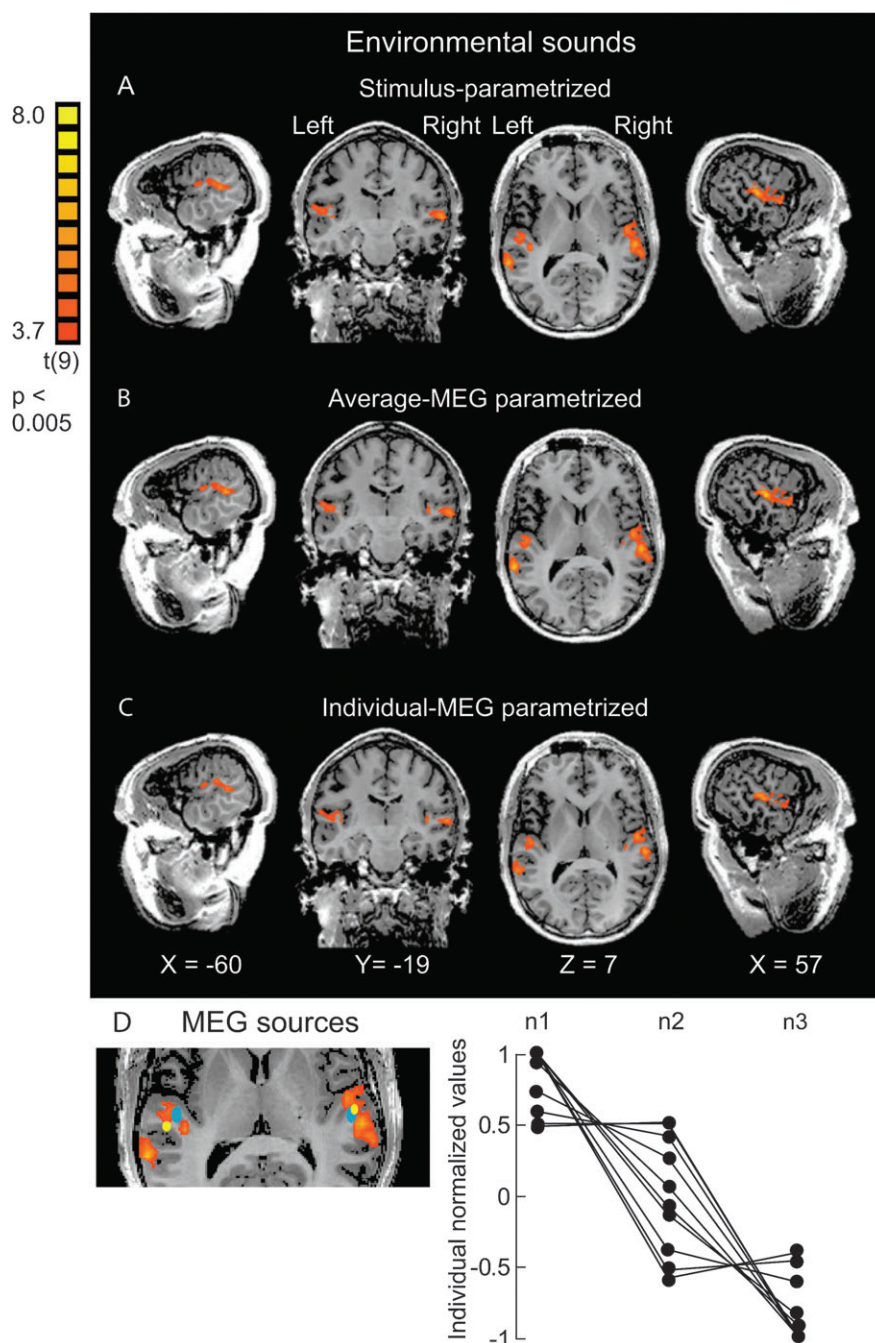


Figure 6. fMRI: noise on environmental sounds. The group volume maps of the responses to environmental sounds with predictors corresponding (A) to the stimulus noise levels, (B) to the averaged MEG data, and (C) to the individual MEG data, projected on to anatomical slices of one subject in Talairach space ($P < 0.005$, cluster size corrected at $\alpha = 0.05$). (D) The mean (\pm SEM) location of the sources of N100m/N400m (yellow) and 250-ms (blue) MEG responses represented by ECDs, superimposed on the group volume map (left), and the individual normalized MEG responses (RH 200-ms response strength as a function of noise level) used in the analysis.

way of combining the information obtained with electrophysiological and hemodynamic measures. Essentially, we utilized an actual neurophysiological filter in building the statistical model for the analysis of fMRI data.

Cortical Reactivity to Spoken Words versus Environmental Sounds in Noise

Based on fMRI, noise-sensitive processing of spoken words was concentrated mainly to the bilateral PT and the right-hemispheric STS. These areas play a role in processing of both verbal

and nonverbal sounds in brain-damaged patients (Schnider et al. 1994; Saygin et al. 2003). Processing of sentences embedded in noise has been reported to activate the left premotor cortex in a noise level-dependent manner, possibly reflecting the recruitment of additional articulatory processes (Scott et al. 2004). Here, activity in the left premotor cortex was detected in response to spoken word stimuli but without significant changes as a function of noise level. This apparent discrepancy may be related to the different task demands in processing sentences versus isolated spoken word stimuli.

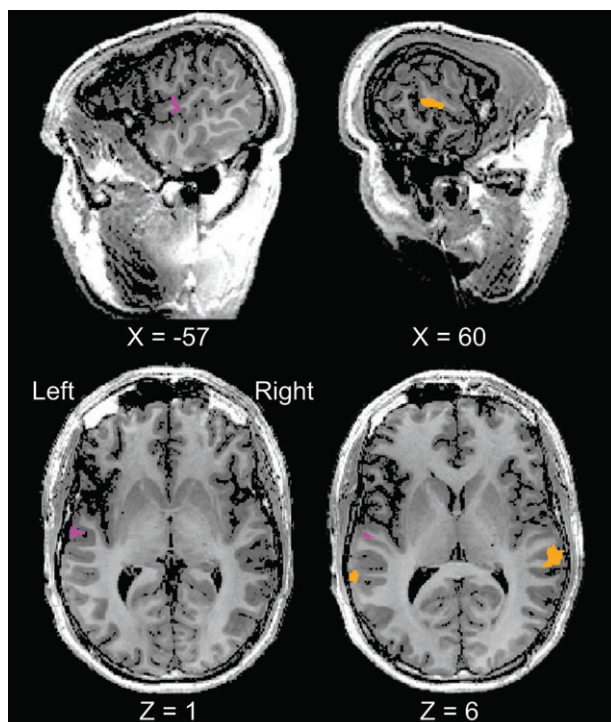


Figure 7. fMRI: spoken words versus environmental sounds. The group volume maps of the fMRI foci for noise-sensitive processing of spoken words (purple) and environmental sounds (orange) ($P < 0.01$, cluster size corrected at $\alpha = 0.05$).

Noise-sensitive processing of environmental sounds activated the left STS and the Heschl's gyrus and PT bilaterally. The same and adjacent regions within the left and right STG have earlier been implicated in processing of environmental sounds after unilateral brain damage (Schnider et al. 1994; Clarke et al. 2000) and, in healthy adults, close-by regions in the bilateral posterior STG and MTG are activated during processing of different categories of environmental sounds (Lewis et al. 2005).

Our MEG results demonstrated that the increasing level of noise modulated the cortical activation to both spoken words and environmental sounds in the RH, whereas LH effects were only seen for spoken words. The prolonged MEG responses to spoken words > 600 ms after the stimulus onset at increased stimulus noise levels, as well as the greater interindividual variability of the MEG responses to spoken words than environmental sounds, speak for stronger top-down modulation of speech stimuli in the LH, for example, in the form of attention needed to allocate when processing words embedded in noise.

The stimuli used in the present study differed in their spectral structure, as tends to be the case for natural speech and nonspeech stimuli. The environmental sounds displayed a rather evenly distributed spectrogram at all noise levels, whereas the speech sounds contained more energy in the frequency range 0–700 Hz than at higher frequencies, and this difference diminished gradually with increasing noise. Thus, the observed differences between stimulus categories may be partly explained by the differential acoustical effects of the applied wideband noise. It is possible that specific subsets of environmental sounds, such as animal vocalizations, might show noise dependence more similar to that of speech

sounds. Earlier literature suggests that the effect of stimulus bandwidth on the cortical responses is highly stimulus specific. Electric N100 responses to band-passed noise decreased (Soeta et al. 2005) and magnetic N100m responses to complex tones increased (Seither-Preisler et al. 2003) together with the stimulus bandwidth. Furthermore, increasing spectral complexity of musical sounds did not affect the N100/N100m responses but increased activation at about 200 ms after the stimulus onset (Shahin et al. 2005). Clearly, future studies are needed to clarify the effect of stimulus bandwidth and also the contribution of central versus peripheral mechanisms (Seither-Preisler et al. 2003) on the observed results.

Several neuroimaging studies have pointed to functional asymmetries in the temporal cortices, and the left and right auditory cortices have been suggested to be predominantly sensitive to temporal and spectral changes, respectively (e.g., Zatorre and Belin 2001; Obleser et al. 2008). Recent MEG studies on processing of tonal sounds in the presence of noise have pointed to left-hemisphere dominance and right-hemisphere “assistance” in noisy environments (Okamoto et al. 2007; Stracke et al. 2009). Our MEG results suggest that added noise affects the analysis of spoken words bilaterally in the early time window, with later pronounced attentional emphasis on the left, whereas processing of environmental sounds—for which rapid online analysis of spectral information is crucial—is modulated more strongly in the RH than LH. In contrast, our fMRI results point to rather bilateral effects for both sound types, probably reflecting also nonphase-locked neuronal activity and, due to the long time integration window, summation across multiple levels of processing.

In the present study, we were not primarily interested in the differences between the stimulus categories, as such, but in the effects of noise, applied similarly to both speech and environmental sounds, and in utilizing the observed differences for optimizing the combined MEG/fMRI analysis. To conclude, our results indicate that while auditory cortical responses to spoken words and environmental sounds share many features, they are modulated in a different manner in the presence of noise. These differences may be partly related to the different acoustic properties of the sound categories, but top-down cortical mechanisms are also likely to be involved, as suggested by the prolonged left-hemispheric responses to speech sounds > 600 ms after the sound onset. Importantly, our stimulus parametrization and the use of MEG evoked responses as neurally filtered measures of stimulus properties provide a valuable and conceptually relevant approach to combining the information obtained from electrophysiological and hemodynamic measures, especially if the neural responses do not linearly follow the stimulus parameters.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>

Funding

Academy of Finland (National Centers of Excellence Programme 2006–2011 and grant numbers #213828 and 127401 to

H.R. and #129160 to R.S.), Sigrid Jusélius Foundation, Netherlands Organisation for Scientific Research, Helsingin Sanomat Centennial Foundation, Finnish Cultural Foundation.

Notes

We thank Antti Puurula for help with the stimulus preparation and in collecting the pilot data, Marita Kattelus for help with the fMRI experiments, and Niclas Kilian-Hütten and Jan Kujala for help with the fMRI data analysis. *Conflict of Interest:* None declared.

References

- Ballas JA, Howard JH, Jr. 1987. Interpreting the language of environmental sounds. *Environ Behav.* 19:91–114.
- Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. 1993. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med.* 30:161–173.
- Benson RR, Richardson M, Whalen DH, Lai S. 2006. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage.* 31:342–353.
- Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, Liberman AM. 2001. Parametrically dissociating speech and non-speech perception in the brain using fMRI. *Brain Lang.* 78:364–396.
- Billings CJ, Tremblay KL, Stecker GC, Tolin WM. 2009. Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear Res.* 254:15–24.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex.* 10:512–528.
- 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.
- 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.
- Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci.* 16:4207–4221.
- Bradlow AR, Alexander JA. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am.* 121:2339–2349.
- Brookes M, Gibson A, Hall S, Furlong P, Barnes G, Hillebrand A, Singh K, Holliday I, Francis S, Morris P. 2005. GLM-beamformer method demonstrates stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. *Neuroimage.* 26:302–308.
- Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ. 2000. Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia.* 38:797–807.
- Cummings A, Čeponiene R, Dick F, Saygin AP, Townsend J. 2008. A developmental ERP study of verbal and non-verbal semantic processing. *Brain Res.* 1208:137–149.
- Cummings A, Čeponiene R, Koyama A, Saygin AP, Townsend J, Dick F. 2006. Auditory semantic networks for words and natural sounds. *Brain Res.* 1115:92–107.
- Cummings A, Saygin AP, Bates E, Dick F. 2009. Infants' recognition of meaningful verbal and nonverbal sounds. *Lang Learn Dev.* 5:172–190.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol.* 112:758–767.
- Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E. 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron.* 26:55–67.
- Davis MH, Johnsrude I. 2003. Hierarchical processing in spoken language comprehension. *J Neurosci.* 23:3423–3431.
- Dick F, Saygin AP, Galati G, Pitzalis S, Bentrovato S, D'Amico S, Wilson S, Bates E, Pizzamiglio L. 2007. What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *J Cogn Neurosci.* 19:799–816.
- 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.
- Foucher JR, Otzenberger H, Gounot D. 2003. The BOLD response and the gamma oscillations respond differently than evoked potentials: an interleaved EEG-fMRI study. *BMC Neurosci.* 4:22.
- Goebel R, Esposito F, Formisano E. 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp.* 27:392–401.
- Gygi B, Kidd GR, Watson CS. 2004. Spectral-temporal factors in the identification of environmental sounds. *J Acoust Soc Am.* 115:1252–1265.
- Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys.* 65:413–497.
- Hari R. 1990. The neuromagnetic method in the study of the human auditory cortex. In: Grandori F, Hoke M, Romani GL, editors. *Auditory Evoked Magnetic Fields and Electric Potentials.* Basel (Switzerland): Karger. p. 222–282.
- Helenius P, Salmelin R, Service E, Connolly J, Leinonen S, Lyytinen H. 2002. Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J Neurosci.* 22:2936–2944.
- Jäncke L, Wüstenberg T, Scheich H, Heinze HJ. 2002. Phonetic perception and the temporal cortex. *Neuroimage.* 15:733–746.
- Kaplan-Neeman R, Kishon-Rabin L, Henkin Y, Muchnik C. 2006. Identification of syllables in noise: electrophysiological and behavioral correlates. *J Acoust Soc Am.* 120:926–933.
- Korvenoja A, Huttunen J, Salli E, Pohjonen H, Martinkauppi S, Palva JM, Lauronen L, Virtanen J, Ilmoniemi RJ, Aronen HJ. 1999. Activation of multiple cortical areas in response to somatosensory stimulation: combined magnetoencephalographic and functional magnetic resonance imaging. *Hum Brain Mapp.* 8:13–27.
- Kozou H, Kujala T, Shtyrov Y, Toppila E, Starck J, Alku P, Näätänen R. 2005. The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hear Res.* 199:31–39.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. 2005. Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci.* 25:5148–5158.
- Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA. 2009. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J Neurosci.* 29:2283–2296.
- Liebsenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. *Cereb Cortex.* 15:1621–1631.
- Martin BA, Kurtzberg D, Stapells D. 1999. The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *J Speech Lang Hear Res.* 42:271–286.
- Martin BA, Sigal A, Kurtzberg D, Stapells DR. 1997. The effect of decreasing audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *J Acoust Soc Am.* 101:1585–1599.
- Martin BA, Stapells DR. 2005. Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear.* 26:195–213.
- Müller-Gass A, Marcoux A, Logan J, Campbell K. 2001. The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. *Neurosci Lett.* 299:197–200.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, et al. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature.* 385:432–434.
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. 2006. Vowel sound

- extraction in anterior superior temporal cortex. *Hum Brain Mapp.* 27:562-571.
- Obleser J, Eisner F, Kotz S. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci.* 28:8116-8124.
- Obleser J, Lahiri A, Eulitz C. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J Cogn Neurosci.* 16:31-39.
- Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C. 2007. Attention improves population-level frequency tuning in human auditory cortex. *J Neurosci.* 27:10383-10390.
- Parbery-Clark A, Skoe E, Kraus N. 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci.* 29:14100-14107.
- Poeppel D, Yellin E, Phillips C, Roberts TP, Rowley HA, Wexler K, Marantz A. 1996. Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Res Cogn Brain Res.* 4:231-242.
- Price C, Thierry G, Griffiths T. 2005. Speech-specific auditory processing: where is it? *Trends Cogn Sci.* 9:271-276.
- Raij T, Ahveninen J, Lin F, Witzel T, Jääskeläinen I, Letham B, Israeli E, Sahyoun C, Vasios C, Stufflebeam S, et al. 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci.* 31:1772-1782.
- Riecke L, van Opstal AJ, Goebel R, Formisano E. 2007. Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci.* 27:12684-12689.
- Riede T, Herzel H, Hammerschmidt K, Brunnberg L, Tembrock G. 2001. The harmonic-to-noise ratio applied to dog barks. *J Acoust Soc Am.* 110:2191-2197.
- Salmelin R. 2007. Clinical neurophysiology of language: the MEG approach. *Clin Neurophysiol.* 118:237-254.
- Saygin A, Dick F, Wilson S, Dronkers N, Bates E. 2003. Neural resources for processing language and environmental sounds. Evidence from aphasia. *Brain.* 126:928-945.
- Schnider A, Benson DF, Alexander DN, Schnider-Klaus A. 1994. Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain.* 117:281-287.
- Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain.* 123:2400-2406.
- Scott SK, Rosen S, Wickham L, Wise RJS. 2004. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am.* 115:813-821.
- Seither-Preisler A, Krumbholz K, Lütkenhöner B. 2003. Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery? *Audiol Neurotol.* 8:322-337.
- Shah NJ, Steinhoff S, Mirzazade S, Zafiris O, Grosse-Ruyken ML, Jäncke L, Zilles K. 2000. The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. *Neuroimage.* 12:100-108.
- Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B. 2005. Modulation of P2 auditory evoked responses by the spectral complexity of musical sounds. *Neuroreport.* 16:1781-1785.
- Sharon D, Hämäläinen M, Tootell R, Halgren E, Belliveau J. 2007. The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. *Neuroimage.* 36:1225-1235.
- 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.
- Stracke H, Okamoto H, Pantev C. 2009. Interhemispheric support during demanding auditory signal-in-noise processing. *Cereb Cortex.* 19:1440-1447.
- Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme Medical Publishers.
- 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.
- van Atteveldt N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. *Neuron.* 43:271-282.
- Van Petten C, Rieffelder H. 1995. Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia.* 33:485-508.
- Vartiainen J, Liljeström M, Koskinen M, Renvall H, Salmelin R. 2011. Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. *J Neurosci.* 31:1048-1058.
- Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci.* 13:994-1005.
- Whiting KA, Martin BA, Stapells DR. 1998. The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear Hear.* 19:218-231.
- Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex.* 11:946-953.