Behavioral/Systems/Cognitive

A Temporal Hierarchy for Conspecific Vocalization Discrimination in Humans

Marzia De Lucia,[1] Stephanie Clarke,[2] and Micah M. Murray[1,2,3,4]

[1]Electroencephalography Brain Mapping Core, Center for Biomedical Imaging, [2]Neuropsychology and Neurorehabilitation Service, Department of Clinical Neurosciences, and [3]Radiology Department, Vaudois University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland, and [4]Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, Tennessee 37235

The ability to discriminate conspecific vocalizations is observed across species and early during development. However, its neurophysiologic mechanism remains controversial, particularly regarding whether it involves specialized processes with dedicated neural machinery. We identified spatiotemporal brain mechanisms for conspecific vocalization discrimination in humans by applying electrical neuroimaging analyses to auditory evoked potentials (AEPs) in response to acoustically and psychophysically controlled nonverbal human and animal vocalizations as well as sounds of man-made objects. AEP strength modulations in the absence of topographic modulations are suggestive of statistically indistinguishable brain networks. First, responses were significantly stronger, but topographically indistinguishable to human versus animal vocalizations starting at 169–219 ms after stimulus onset and within regions of the right superior temporal sulcus and superior temporal gyrus. This effect correlated with another AEP strength modulation occurring at 291–357 ms that was localized within the left inferior prefrontal and precentral gyri. Temporally segregated and spatially distributed stages of vocalization discrimination are thus functionally coupled and demonstrate how conventional views of functional specialization must incorporate network dynamics. Second, vocalization discrimination is not subject to facilitated processing in time, but instead lags more general categorization by ~100 ms, indicative of hierarchical processing during object discrimination. Third, although differences between human and animal vocalizations persisted when analyses were performed at a single-object level or extended to include additional (man-made) sound categories, at no latency were responses to human vocalizations stronger than those to all other categories. Vocalization discrimination transpires at times synchronous with that of face discrimination but is not functionally specialized.

Introduction

Vocalizations are essential in communication and social interactions, conveying the speaker's identity, gender, intentions, and emotional state. Whether processing conspecific vocalizations recruits dedicated brain resources remains highly controversial. Studies in nonhuman primates demonstrated response sensitivity to conspecific vocalizations within temporal regions. Some argue for selectivity within circumscribed rostral regions (Tian et al., 2001). Others emphasize distributed mechanisms (Poremba et al., 2004; Cohen et al., 2006; Petkov et al., 2008; Recanzone, 2008; Russ et al., 2008; Staeren et al., 2009). In humans, voice recognition deficits (phonagnosia) after (right) temporo-parietal brain lesions can dissociate from aphasia and agnosia (Assal et al., 1981; Van Lancker and Canter, 1982), but frequently cooccur with amusia (Peretz et al., 1994) or can even be observed in the absence of gross structural damage (Garrido et al., 2009). Hemodynamic imaging has documented selective responsiveness to human vocalizations within the middle and anterior superior temporal sulcus (STS) (Belin et al., 2000). Interpreting these data in terms of functional selectivity is not straightforward. The speech content of the stimuli may strongly contribute to selective effects (Belin et al., 2000; Fecteau et al., 2004), as can the harmonic structure of sounds, which is greater in vocalizations (Lewis et al., 2005, 2009) (for data showing attention-driving modulations with identical acoustic stimuli, see also von Kriegstein et al., 2003). Another consideration is that, as in monkeys, effects can extend to regions beyond the STS (von Kriegstein et al., 2003, 2007; Fecteau et al., 2005), highlighting the importance of high spatial and temporal resolution for ascertaining when/where functional selectivity originates within distributed brain networks.

Despite the suitability of auditory evoked potentials (AEPs) for addressing brain dynamics, extant studies have produced discordant results with limited interpretational power. Levy et al. (2001, 2003) documented an attention-dependent "voice-specific response" peaking at 320 ms after stimulus onset. But this effect may instead reflect living versus man-made categorization (Murray et al., 2006) because voices were only contrasted with musical instruments. Charest et al. (2009) compared responses to human vocalizations (speech and nonspeech) with those to environmental

Received May 2, 2010; revised July 5, 2010; accepted July 8, 2010.

This work was supported by Swiss National Science Foundation Grants 3100A0-118419 (M.M.M.), 3100A0-103895 (S.C.), and K-33K1_122518/1 (M.D.L.), and the Leenaards Foundation 2005 Prize for the Promotion of Scientific Research (M.M.M.). Cartool software was programmed by Denis Brunet (Functional Brain Mapping Laboratory, Geneva, Switzerland) and is supported by the EEG Brain Mapping Core of the Center for Biomedical Imaging. Christoph Michel and Jean-François Knebel provided additional analysis tools. Christian Camen assisted with data collection.

Correspondence should be addressed to Micah M. Murray, Electroencephalography Brain Mapping Core, Center for Biomedical Imaging, Radiology, BH08.078, Vaudois University Hospital Center, rue du Bugnon 46, 1011 Lausanne, Switzerland. E-mail: micah.murray@chuv.ch.

DOI:10.1523/JNEUROSCI.2239-10.2010

 0270-6474/10/3011210-12$15.00/0

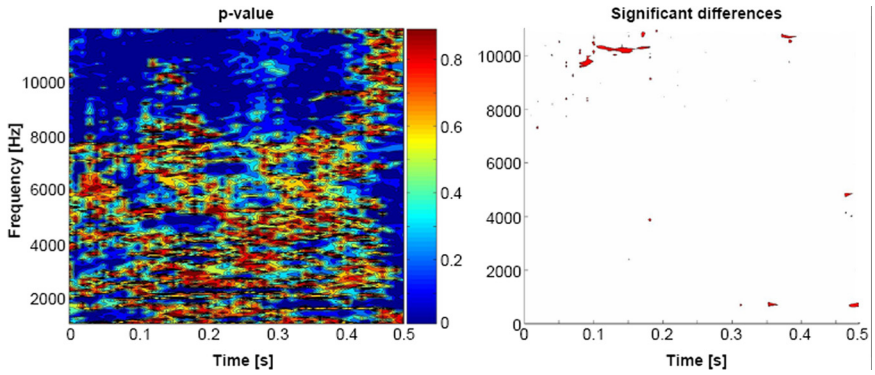

Figure 1. Statistical comparison of stimuli. Left, The spectrogram of each stimulus was generated and comparisons (nonparametric *t* tests) were performed across groups of sounds for each ~5 ms and ~80 Hz time–frequency bin. Right, Bins meeting the following statistical criteria are displayed as red: eight spatially contiguous bins (equivalent to a cluster-level value of $p < 0.00625$).

sounds or bird songs. Voice-related AEP waveform modulations began 164 ms after stimulus onset, but additional analyses revealed their effect was mostly (if not wholly) driven by the speech content of the stimuli and/or acoustic differences. Moreover, because these AEP studies analyzed voltage waveforms, the latency and spatial distribution of statistical effects are valid only for the chosen reference (i.e., variance changes with the reference) and therefore have no unequivocal neurophysiologic validity (Murray et al., 2008). Consequently, the spatiotemporal brain mechanisms mediating conspecific vocalization discrimination in humans remain unclear. We applied electrical neuroimaging analyses to AEPs from psychophysically and acoustically controlled sounds to disambiguate whether vocalization discrimination relies on dedicated neural mechanisms. Such would be predicted to recruit distinct brain regions and to therefore result in topographic AEP differences. More generally, by including analyses of a wide range of object categories and also by performing analyses on responses to individual auditory objects in a manner akin to that typically performed in nonhuman primates, we disambiguated categorical processes from low-level acoustic analyses.

Materials and Methods

Participants. Ten healthy, right-handed individuals (seven females), aged 21–34 years, participated. All subjects provided written, informed consent to participate in the study, the procedures of which were approved by the Ethics Committee of the University of Geneva. None had a history of neurological or psychiatric illnesses, and all reported normal hearing. Data from these individuals have been previously published in an investigation of living versus man-made categorical discrimination (Murray et al., 2006) as well as in a study examining responses to subclasses of man-made sounds (De Lucia et al., 2009). The primary analyses in the present study are thus a more extensive analysis of these data (i.e., the AEPs to specific subclasses of living stimuli; with additional analyses including AEPs to subsets of sounds of man-made objects, detailed below). Plus, AEPs were calculated in response to single vocalizations.

Stimuli. Auditory stimuli were complex, meaningful sounds (16 bit stereo; 22,500 Hz digitization) [for a full listing, including details on the acoustic attributes as well as psychometrics concerning these stimuli, see Murray et al. (2006), their Table 1]. There were 120 different sound files in total, 60 of which represented sounds of living objects (3 exemplars of 20 different referent objects) and 60 of which represented sounds of man-made objects (3 exemplars of 20 different reference objects). Each sound was 500 ms in duration, which included an envelope of 50 ms decay time that was applied to the end of the sound file to minimize clicks at sound offset. All sounds were further normalized according to the root mean square of their amplitude. Our previous work has demonstrated that the sounds used in this study were all highly familiar as well as reliably identified with a high level of confidence (see also supplemental table, available at www.jneurosci.org as supplemental material) (Murray et al., 2009a; De Lucia et al., 2010b).

The 60 sound files that were the focus of the present investigation were restricted to those of living objects, which were further sorted between human nonverbal vocalizations and animal vocalizations (hereafter, human and animal sounds, respectively). The 8 human sounds included 3 exemplars each of the following (i.e., a total of 24 unique sound files): whistling, sneezing, screaming, laughing, gargling, coughing, clearing one's throat, and crying. The 12 animal sounds included 3 exemplars each of the following animals' stereotypical vocalizations (i.e., a total of 36 unique sound files): sheep, rooster, pig, owl, frog, donkey, dog, crow, cow, chicken, cat, and birds.

To assess whether these groups of human and animal vocalizations differed acoustically, we statistically compared the spectrograms (defined with Matlab's spectrogram function with no overlapping and zero padding), using a time–frequency bin width of ~5 ms and ~74 Hz. Statistical contrasts entailed a series of nonparametric *t* tests based on a bootstrapping procedure with 5000 iterations per time–frequency bin to derive an empirical distribution against which to compare the actual difference between the mean spectrograms from each sound category (Aeschlimann et al., 2008; Knebel et al., 2008; De Lucia et al., 2009, 2010b). Note that there was no grouping or averaging of the spectrograms either for a given object or for a given category. Also, it should be noted that this analysis provides complementary (and in some regards more comprehensive) information to an analysis of formants, the latter of which would lack the temporal information provided in the spectrogram analysis. A significant difference at a given time–frequency bin was only considered reliable if all eight of its immediately adjacent bins also yielded values of $p \leq 0.05$ (i.e., a 3×3 bin spatial threshold was applied). This constitutes a minimal level of correction for multiple contrasts and time–frequency autocorrelation, as we were particularly interested in this analysis being overly sensitive to acoustic differences. Nonetheless, there were no statistically reliable differences between the spectrograms from each group of sounds (Fig. 1). Individual sound files of course differed one from the other to render the sound referent identifiable and unique.

The groups of sounds were likewise compared in terms of their mean harmonics-to-noise ratio (HNR), which was calculated using PRAAT software (http://www.fon.hum.uva.nl/praat/). HNR provides an index of the ratio of the energy contained in the harmonics versus nonharmonics of a sound. The mean (±SEM) HNR for the 24 human sounds was 9.7 ± 1.6 (range, −0.1 to 27.1), and for the 36 animal sounds was 9.1 ± 1.2 (range, 0.0 to 29.1). These values did not significantly differ ($p > 0.75$). Thus, although HNR may contribute to the general processing of vocalizations (Lewis et al., 2005, 2009), it should not differentially contribute to processing our set of human and animal vocalizations.

Finally, the groups of sounds were compared in terms of their power spectrum, which quantifies the energy of a sound at a given frequency. The power spectrum was calculated using a sliding Hamming window of 256 data points (11.38 ms) and 50% of overlap between consecutive windows. The spectral peaks of the power spectrum indicate the formants of the sounds, which are known to be related to the size of the creature generating the vocalization and are also a characteristic feature of vocalizations (Ghazanfar et al., 2007). These peaks are displayed in Figure 2 for all of the stimuli (i.e., each of the three exemplars from each vocalization). As expected, human nonverbal vocalizations had formants that clustered around lower frequencies (i.e., below ~2 kHz). This was also generally the case for the animal vocalizations, with some exceptions. Statistical comparison (unpaired *t* test with equal variances not assumed) was performed on the fundamental frequency for each vocalization (f_0;

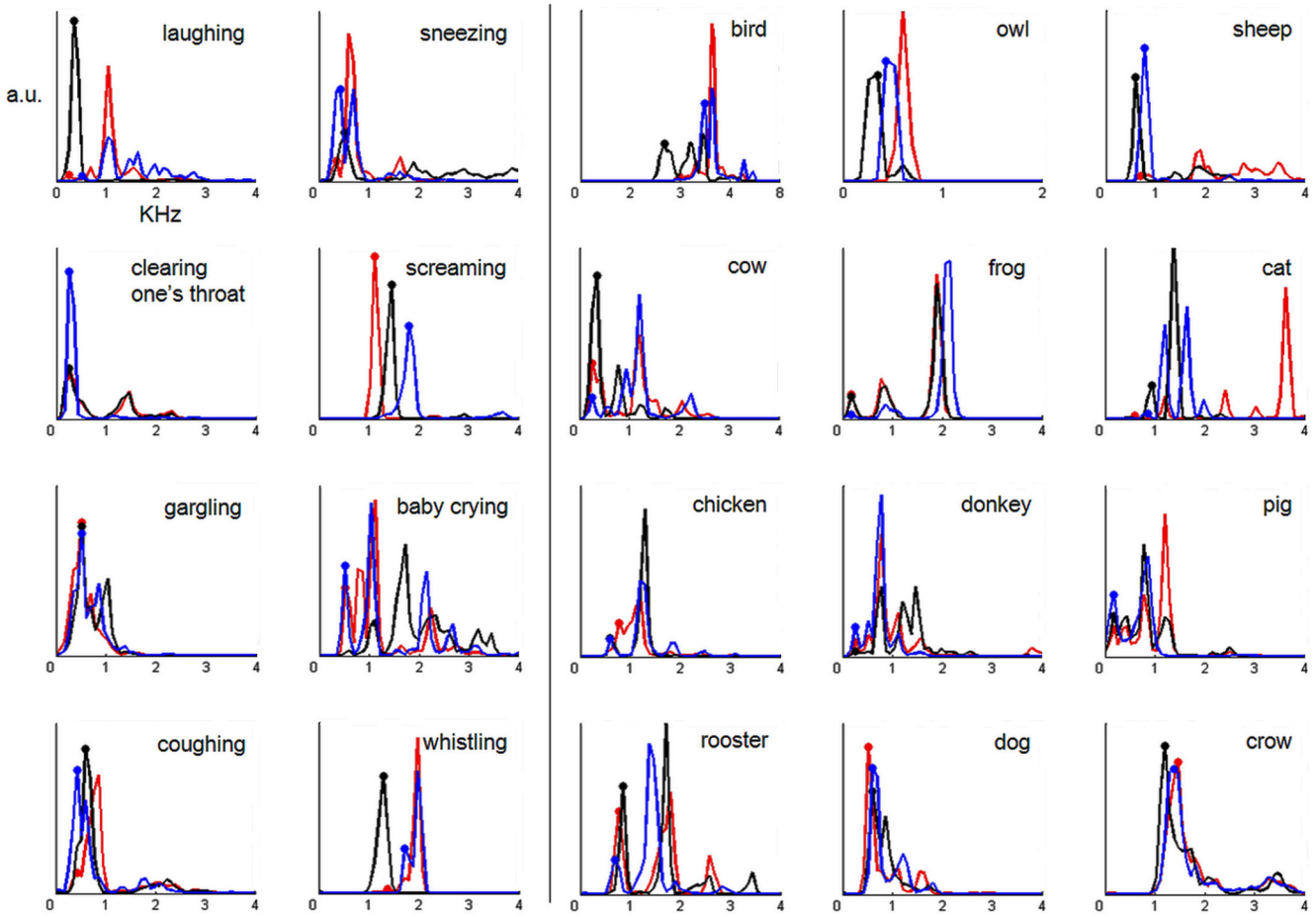

Figure 2. Power spectra of each vocalization. Each line is the power spectrum for a single exemplar of a given vocalization. The two leftmost columns display the spectra for human vocalizations, and the three rightmost columns, the spectra for animal vocalizations. The *x*-axis is frequency in kilohertz, and the *y*-axis is in arbitrary units. The dots indicate the lowest frequency peak in each power spectrum for each of the sounds (i.e., f_0). These f_0 values were not significantly different between the two groups of vocalizations either when considered separately ($t_{(52.6)} = 0.71$; $p = 0.48$) or when first averaged across the exemplars of a given object ($t_{(16.3)} = 0.41$; $p = 0.69$).

indicated by dots in Fig. 2). First, this was done without averaging the f_0 values for the three exemplars of a given object (i.e., so that there were 60 f_0 values; 24 for human vocalizations and 36 for animal vocalizations). There was no evidence that the f_0 values differed (0.71 vs 0.86 kHz; $t_{(52.6)} = 0.71$; $p = 0.48$). Next, we repeated this analysis after first averaging the f_0 values across the three exemplars of a given object (i.e., so that there were 20 f_0 values; 8 for human vocalizations and 12 for animal vocalizations). There was no evidence that the f_0 values differed ($t_{(16.3)} = 0.41$; $p = 0.69$). This was done because the single-object AEPs were calculated by averaging epochs from different exemplars of the same object (to obtain sufficient signal quality). Thus, we could in turn evaluate whether there was a systematic relationship between these single-object AEPs and their corresponding f_0 values.

The other 60 sound files were those of man-made objects. AEPs in response to these sounds were included in an analysis targeted at the issue of whether sounds of human vocalizations yielded significantly stronger responses not only with respect to animal vocalizations, but also more generally with respect to other categories of environmental sounds. These sounds of man-made objects were subdivided between musical instruments and objects associated with a specific sociofunctional context (hereafter "music" and "nonmusic," respectively). The 10 music sounds included exemplars of notes being played on the following musical instruments (three exemplars per object): accordion, flute, guitar, harmonica, harp, organ, piano, saxophone, trumpet, and violin (i.e., both string and brass instruments involving mouth and hand actions). We would emphasize that these stimuli were neither rhythmic nor melodic in character and were not perceived as music, but rather in terms of the instrument generating the sound. We would also note that none of the participants were musicians or had extensive musical training. The 10 nonmusic sounds included exemplars of the following objects (three per object): bicycle bell, car horn, cash register, cuckoo clock, doorbell, closing door, glass shattering, police siren, church bell, and telephone [i.e., sounds that typically trigger a responsive action on being heard, as supported by our previously published psychophysical experiment appearing in the study by De Lucia et al. (2009)]. Likewise, these two subcategories of sounds of man-made objects were likewise controlled at the group level in terms of their acoustic features as assessed with methods akin to those described above for the evaluation of human and animal vocalizations [cf. De Lucia et al. (2009, 2010b), their supplemental Fig. 1].

Procedure and task. Participants performed a living versus man-made "oddball" detection paradigm, such that on a given block of trials "target" stimuli to which subjects pressed a response button occurred 10% of the time. The use of sounds of living and man-made objects as target stimuli was counterbalanced across blocks. The remaining 90% of stimuli ("distracters") were comprised of the other sound category. The living and man-made stimuli were blocked into series of 300 trials (~18 min) with an interstimulus interval of 3.4 s. Each participant completed four blocks of trials (two in which man-made sounds were targets and two in which living sounds were targets) and took a 5–10 min break between blocks to minimize fatigue. The order of blocks was varied across participants. For all the AEP analyses in this study, only blocks of trials when the sounds served as distracters were analyzed. This removes any contamination of motor-related activity from the AEPs. For example, to generate AEPs to the music and nonmusic subcategories of man-made objects, we used the two blocks of trials when living sounds were the targets and man-made sounds were the distracters. To generate AEPs to

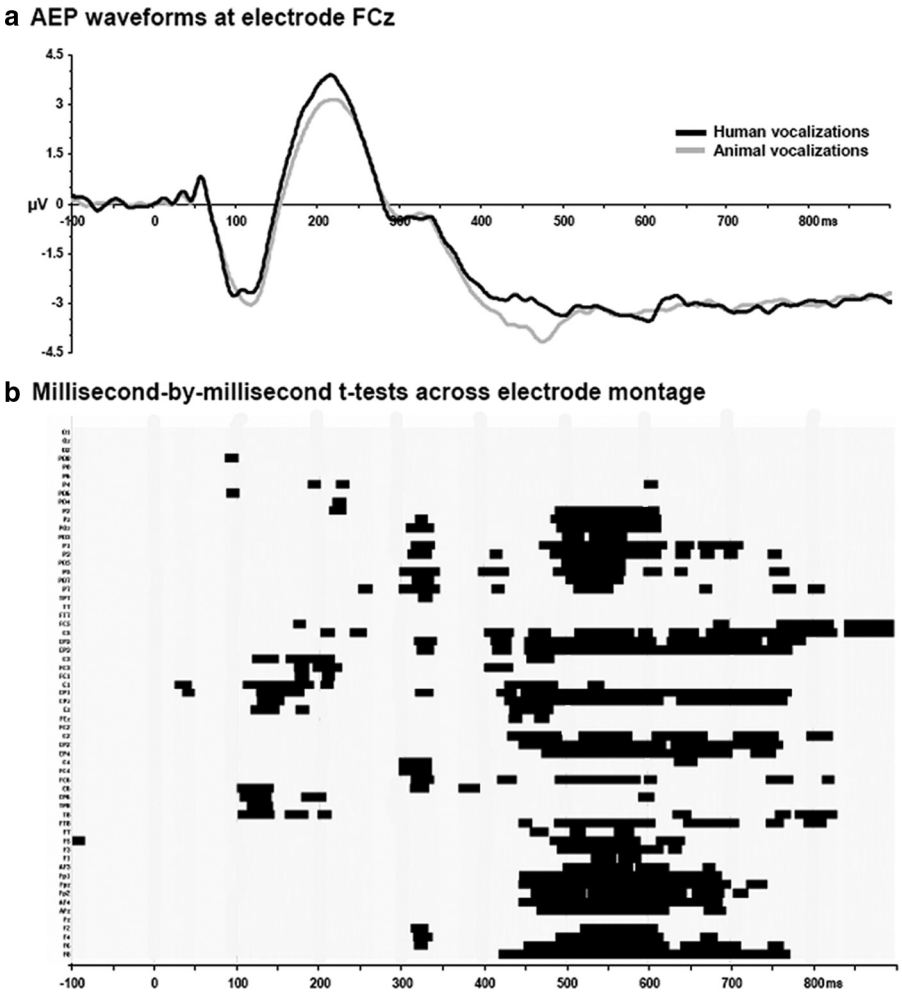

Figure 3. ***a***, Exemplar waveforms from a frontocentral midline electrode (FCz). These group-averaged waveforms exhibit prototypical AEP peaks. Response modulations are visually apparent from 160 ms after stimulus onset. ***b***, The results of millisecond-by-millisecond paired *t* tests at each of the scalp electrodes from the group-averaged AEP waveforms are shown (only $p < 0.05$ with a 25 ms temporal criterion are shown).

human and animal vocalizations, we used the two blocks of trials when man-made sounds were the targets and living sounds were the distracters.

Behavioral as well as EEG data were collected from all conditions throughout the length of the experiment, and STIM (Neuroscan) was used to control stimulus delivery and to record behavioral responses. Audiometric quality insert earphones (supplied by Neuroscan) were used for stimulus delivery. This paradigm is in many regards similar to what has recently been used in studies of nonhuman primates (Petkov et al., 2008; Remedios et al., 2009). In these studies, the participants were awake and centrally fixating, but did not perform any discrimination of the acoustic stimuli. Similarly, in the study by Recanzone (2008), the participants released a lever when the location of the stimulus changed and thus were arguably attending to spatial features of the sounds. In the present study, participants were attending to the auditory modality and also to the general categories of the stimuli (i.e., whether it was living vs man-made) but did not perform any overt discrimination of human versus animal vocalizations. In this regard, any discrimination observable in the AEPs can be considered as implicit. On the one hand, this aspect of the design was intended for allowing a closer comparison with results in animal models. Likewise, any paradigm demonstrating implicit discrimination would also be of relevance as a clinical examination tool. Finally, we opted for this design because previous studies have generated conflicting evidence as to whether AEP correlates of vocalization discrimination rely on overt attention to the "voice-ness" of the stimuli (Levy et al., 2001, 2003; Charest et al., 2009).

EEG acquisition. Continuous 64-channel EEG was acquired through Neuroscan Synamps (impedances, <5 kΩ), referenced to the nose, bandpass filtered 0.05–200 Hz, and digitized at 1000 Hz. In what follows, we first describe the preprocessing and analysis procedures for AEPs calculated across objects (hereafter across-object AEPs). We then detail our procedures of AEPs calculated for individual objects (hereafter, single-object AEPs).

Across-object AEP preprocessing. For across-object AEPs, peristimulus epochs of continuous EEG (-100 to 900 ms) from distracter trials were averaged from each subject separately to compute AEPs. As mentioned above, EEG from target trials was not analyzed, although the behavioral results reported below refer to these trials. Trials with blinks or eye movements were rejected off-line, using horizontal and vertical electro-oculograms. An artifact criterion of ± 100 μV was applied at all other electrodes, and each EEG epoch was also visually evaluated. Data from artifact electrodes from each subject and condition were interpolated using three-dimensional splines (Perrin et al., 1987). There were at least 105 acceptable EEG epochs per condition (human vocalizations, animal vocalizations, music and nonmusic AEPs) for each participant. After this procedure and before group averaging, each subject's data were 40 Hz low-pass filtered, baseline corrected using the -100 ms prestimulus period, downsampled to a common 61-channel montage, and recalculated against the common average reference.

Across-object AEP analyses and source estimations. The first set of across-object analyses focused on identifying differences in AEPs in response to human and animal vocalizations. This was accomplished with a multistep analysis procedure that we refer to as electrical neuroimaging, examining both local and global measures of the electric field at the scalp. These analyses have been extensively detailed previously (Michel et al., 2004; Murray et al., 2008, 2009b). Briefly, they entail analyses of response strength and response topography to differentiate effects attributable to modulation in the strength of responses of statistically indistinguishable brain generators from alterations in the configuration of these generators (viz. the topography of the electric field at the scalp). That is, electrical neuroimaging analyses examine two orthogonal features of the electric field at the scalp—its strength and topography—that have different underlying neurophysiologic bases. In addition, we used the local autoregressive average distributed linear inverse solution (LAURA) (Grave de Peralta Menendez et al., 2001) to visualize and statistically contrast the likely underlying sources of effects identified in the preceding analysis steps.

Electrical neuroimaging analyses, being reference independent, have several advantages over canonical waveform analyses. The statistical outcome with voltage waveform analyses will change with the choice of the reference electrode (Murray et al., 2008). This is because the intersubject (or intermeasurement) variance at the chosen reference will forcibly be zero and in turn vary elsewhere over the electrode montage. Consequently, changing the reference will change the spatial distribution of the variance and in turn the latency and distribution of statistical effects. Nonetheless, a visual impression of effects within the dataset was obtained by analyzing average-reference waveform data from all electrodes as a function of time poststimulus onset in a series of pairwise *t* tests (thresholded at $p < 0.05$) with correction for temporal autocorrelation at

individual electrodes through the application of a 25 consecutive data point criterion for the persistence of differential effects (i.e., 25 ms duration). Note, however, that our conclusions are based solely on reference-independent measures of the electric field at the scalp.

Changes in the strength of the electric field at the scalp were assessed using global field power (GFP) (Lehmann and Skrandies, 1980; Koenig and Melie-Garcia, 2010) from each subject and experimental condition. Values at each time point were compared with a paired *t* test, as above. This measure indicates the global strength of the response, regardless of its topographic distribution. To statistically identify periods of topographic modulation, we calculated the global dissimilarity (Lehmann and Skrandies, 1980) between responses for each time point and applied a Monte Carlo bootstrapping analysis procedure that is colloquially referred to as topographic ANOVA (TANOVA) (Murray et al., 2008). Because electric field changes are indicative of changes in the underlying generator configuration (Murray et al., 2008), this analysis provides a statistical means of determining whether and when brain networks mediating responses to human and animal vocalizations differ.

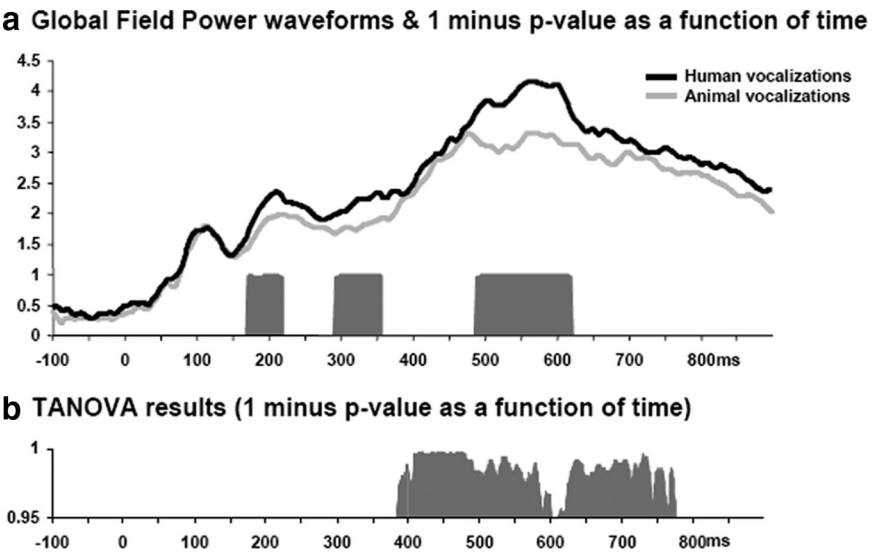

Figure 4. ***a***, Modulations in response strength were identified using GFP. Group-averaged GFP waveforms are displayed along with the results of millisecond-by-millisecond paired *t* tests. ***b***, Topographic modulations between conditions were assessed using global dissimilarity. The results of the TANOVA procedure are illustrated as a function of time (in both panels 1 minus *p* value is shown after applying a $p < 0.05$ and 25 ms temporal criterion, as in Fig. 3).

An agglomerative hierarchical clustering analysis of the AEP topography at the scalp identified time periods of stable topography, which is a data-driven means for defining AEP components (Murray et al., 2008, 2009b; De Lucia et al., 2010a). The optimal number of topographies or "template maps" that accounted for the group-averaged data set (i.e., the poststimulus periods of both conditions, collectively) was determined by a modified Krzanowski–Lai criterion (Murray et al., 2008, 2009b). The pattern of template maps identified in the group-averaged data was then statistically tested in the data of each individual subject, using spatial correlation. The output is a measure of relative map presence for each subject that is in turn submitted to a repeated-measure ANOVA with factors of condition and map. In conjunction with the aforementioned TANOVA, this procedure reveals whether AEPs from a given condition are more often described by one map versus another, and therefore whether different intracranial generator configurations better account for AEPs from each condition.

Intracranial sources were estimated using a distributed linear inverse solution and LAURA regularization approach (Grave de Peralta Menendez et al., 2001). LAURA uses a realistic head model, and the solution space included 4024 nodes, selected from a $6 \times 6 \times 6$ mm grid equally distributed within the gray matter of the Montreal Neurological Institute (MNI) average brain (courtesy of R. Grave de Peralta Menendez and S. Gonzalez Andino, both at the University Hospital of Geneva, Geneva, Switzerland). The above AEP analyses defined the time periods over which sources were estimated. Statistical analyses of source estimations were performed by first averaging the AEP data across time to generate a single data point for each participant and condition. This procedure increases the signal-to-noise ratio of the data from each participant. The inverse solution (10 participants $\times$ 2 conditions) was then estimated for each of the 4024 nodes in the solution space. Paired *t* tests were calculated at each node using the variance across participants. Only nodes with values of $p \leq 0.005$ ($t_{(9)} \geq 3.68$) and clusters of at least 12 contiguous nodes were considered significant. This spatial criterion was determined using the AlphaSim program (available from the Analysis of Functional NeuroImages website), which entailed performing 10,000 Monte Carlo permutations on the 4024 nodes of our lead field matrix to determine the false discover rate for clusters of different sizes. In our case, there was a false-positive probability of 0.0192 for observing a cluster of minimally 12 contiguous nodes. The results of the source estimations were rendered on the MNI brain with the Talairach and Tournoux (1988) coordinates of the largest statistical differences indicated. Functional coupling between regions identified during statistical analysis of source estimations was evaluated using nonparametric correlation (Spearman's ρ).

The second set of across-object AEP analyses focused on determining whether or not there are selectively enhanced responses to human vocalizations relative not only to animal vocalizations but also to the music and nonmusic conditions described above. For this, we used the GFP in response to each condition from each subject. Area measures were taken over time periods either defined based on our previous work (Murray et al., 2006) or based on the above analyses. These were submitted to a one-way ANOVA using the within-subject factor of sound variety.

Single-object AEP preprocessing and analyses. For single-object AEPs, peristimulus epochs of continuous EEG (−100 to 500 ms) from distracter trials were averaged from each subject separately to compute AEPs. A shorter time interval than above was selected in part because these analyses were conducted as a follow-up to the above analyses. Consequently, we could focus our analyses on time intervals identified from the across-object AEPs. Likewise, a shorter epoch length improved the acceptance rate and the consequent signal quality of the single-object AEPs. Trials with blinks or eye movements were rejected off-line, using horizontal and vertical electro-oculograms. An artifact criterion of ± 100 μV was applied at all other electrodes, and each EEG epoch was also visually evaluated. Data from artifact electrodes from each subject and condition were interpolated using three-dimensional splines (Perrin et al., 1987). There was a minimum of 15 acceptable EEG epochs per object for any given participant. After this procedure and before group averaging, each subject's data in response to each object were 40 Hz low-pass filtered, baseline corrected using the −100 ms prestimulus period, downsampled to a common 61-channel montage, and recalculated against the common average reference.

Analyses of single-object AEPs were limited to GFP waveforms and area measures. As is detailed below in Results, the across-object AEP analyses identified robust GFP differences between responses to human and animal vocalizations in the absence of any modulations in AEP topography. The earliest of these effects was over the 169–219 ms poststimulus interval. Consequently, the single-object AEP analyses were limited to GFP area measures over this same time interval, although for completion we include displays of the full time series. These GFP area measures were submitted to a univariate ANCOVA using vocalization type as the fixed factor, subject as the random factor, and f_0 for each object as the covariate. In addition, we used a nonparametric linear regression analysis (Spearman's ρ) both at the single-subject and group

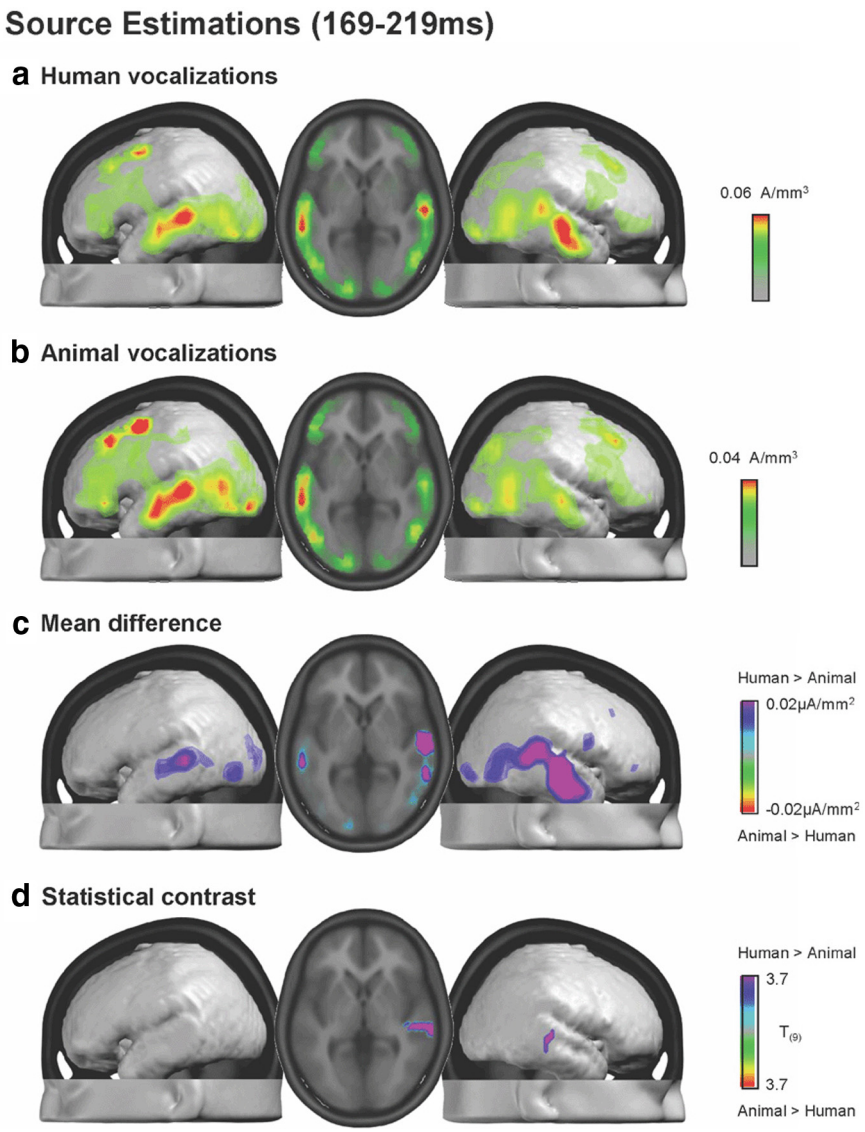

Figure 5. ***a***, ***b***, Group-averaged distributed linear source estimations were calculated over the 169–219 ms poststimulus period for each experimental condition. Results are rendered on the average MNI brain. Axial slice shows the activations for each of the two conditions in correspondence to the maximal t value at 47, −22, 6 mm. ***c***, Mean difference in source estimations included a distributed set of regions. The scaling for this difference is one-half that of the maximum for the estimations in response to animal vocalizations. ***d***, Results of the statistical contrast of the source estimations between AEPs to human and animal vocalizations are displayed in the same manner as in ***a*** and ***b***.

level to assess whether GFP area over the 169–219 ms period was linked to the f_0 of the vocalization.

Results

Behavioral results

Participants accurately performed the target detection task (Murray et al., 2006). The mean (±SEM) percentage of correct responses to human and animal sounds when they served as targets were 85.8 ± 4.1 and 90.2 ± 2.7%, respectively, and did not significantly differ ($t_{(9)} = 1.50$; $p > 0.15$). Likewise, reaction times to human and animal sounds were 895 ± 36 and 901 ± 44 ms, respectively, and did not significantly differ ($t_{(9)} = 0.32$; $p > 0.75$). Thus, behavioral differences cannot readily account for the AEP modulations described below. Plus, because the main AEP analyses were based on data from distracter trials, any response-related activity in the effects we obtained were minimized (if not eliminated).

Vocalization discrimination: across-object AEPs

The first level of analysis focused on determining the onset of response differences (based on average-referenced voltage waveforms) between across-object AEPs in response to sounds of human and animal vocalizations. Figure 3 displays the group-average AEPs from a fronto-central midline electrode (FCz) where the magnitude of the earliest difference was largest, as well as the results of the millisecond-by-millisecond paired t test across the 61-channel electrode montage. Temporally sustained and statistically reliable differences were observed across several electrodes of the montage beginning ~100–200 ms after stimulus onset.

The remainder of analyses with these across-object AEPs was therefore based on reference-independent measures of the electric field at the scalp: one examining response strength independent of topography and the other examining response topography independent of response strength (i.e., GFP and dissimilarity, respectively). The first of these, a millisecond-by-millisecond analysis of the group-averaged GFP waveforms revealed sustained differences between responses over the 169–219, 291–357, and 487–621 ms poststimulus periods (Fig. 4*a*). Responses were stronger in response to human vocalizations over all of these time periods. Second, global dissimilarity between conditions tested on a millisecond-by-millisecond basis whether the topographies of the AEPs differed between conditions. Sustained topographic differences were observed over the 389–667 ms poststimulus periods (Fig. 4*b*), but not over the earlier time periods when GFP modulations were observed. In fact, over the 169–219 ms period, the p value of the TANOVA never dropped below 0.32 (the average p value over this time period was 0.67). Thus, there was no evidence of either short-lived significant periods of topographic difference or trends of such. It is perhaps worthwhile to mention that these features (i.e., response strength and response topography) are ordinarily overlooked in canonical analyses of AEPs. The above analyses therefore allow for differentiating, as a function of time, when AEPs to human versus animal vocalizations differ in either/both of these features that in turn have distinct underlying neurophysiologic bases for their appearance. This pattern would suggest that the earliest differentiation between across-object AEPs is attributable to modulations in the strength of statistically indistinguishable configurations of intracranial brain networks. In other words, the earliest differentiation of human and animal vocalizations appears to rely on the same (or at least statistically indistinguishable) brain networks. Distinct, specialized regions do not appear to be implicated during these early stages.

A topographic hierarchical cluster analysis was then conducted to identify time periods of stable electric field topography both within and between experimental conditions. This analysis, first performed at the group-averaged across-object AEP level, is a means of identifying AEP components and for determining whether the above topographic modulation follows from a singular and stable topographic difference or rather from multiple configuration changes (Murray et al., 2008). The global explained variance of this clustering for the concatenated group-averaged dataset from both experimental conditions was 97.38%. This analysis indicated similar maps were observed for both conditions until ~400 ms after stimulus onset, mirroring the effects obtained when measuring global dissimilarity. Over the 389–667 ms poststimulus period, four different maps were observed at the group average level; two of which predominated in the responses to human vocalizations (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). This was statistically evaluated using a measure of map presence that is based on the spatial correlation between the template maps identified in the group-averaged AEPs and single-subject data. Over the 389–667 ms period, there was a significant main effect of map ($F_{(2,8)} = 4.502$; $p = 0.049$) and a significant interaction between factors of experimental condition and template map ($F_{(2,8)} = 6.429$; $p = 0.022$). Follow-up contrasts revealed that one template map (map HV) was more often spatially correlated with responses to human vocalizations ($t_{(9)} = 4.074$; $p < 0.003$), whereas another was more often spatially correlated with responses to animal vocalizations (map AV; $t_{(9)} = 2.821$; $p = 0.020$). There was no reliable difference between conditions for either of the other two template maps (map X and map Y) (see supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

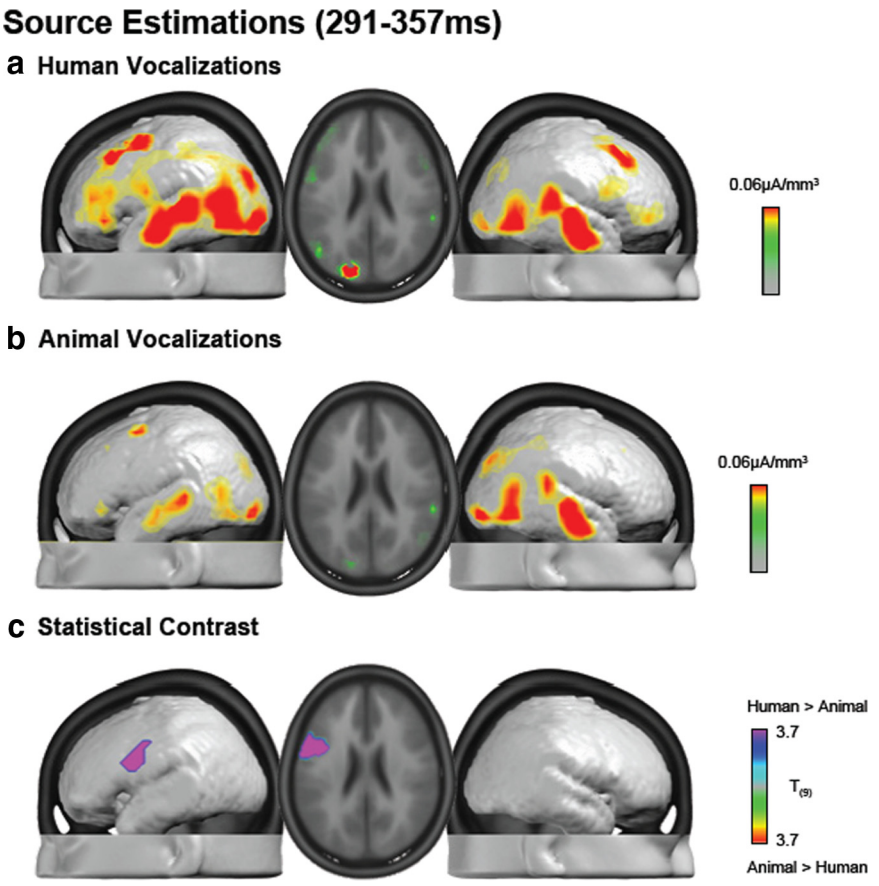

Figure 6. ***a***, ***b***, Group-averaged distributed linear source estimations were calculated over the 291–357 ms poststimulus period for each experimental condition (scale indicated). Results are rendered on the average MNI brain. Axial slice shows the activation for each of the two conditions in correspondence to the maximal *t* value at −53, −3, 40 mm. ***c***, Results of the statistical contrast of the source estimations between AEPs to human and animal vocalization are displayed in the same manner as in ***a*** and ***b***.

Vocalization discrimination: across-object source estimations

Analyses to this point indicate that AEP responses to sounds of human vocalizations and animal vocalizations first differed both in their strength, but not topography, over the 169–219 ms period and that a single and common topography was identified over this time period for both conditions. By extension, such a pattern of effects suggests that human and animal vocalization processing initially involves a statistically indistinguishable brain network that varies in its response strength. This is highly consistent with findings emerging from recent studies in nonhuman primates in which recordings across five auditory regions all exhibited similar selectivity in their responses to conspecific vocalizations (Recanzone, 2008) (see also Petkov et al., 2008).

Intracranial sources were estimated with a distributed inverse solution (for details, see Materials and Methods) for each across-object AEP and participant over the 169–219 ms period and then group-averaged (Fig. 5*a*,*b*). Responses to both human and animal vocalizations included prominent sources along the superior temporal lobes with additional sources evident posteriorly at the temporo-parietal junction and also within the occipital lobe. The mean difference in source estimations revealed a widespread network of brain regions exhibiting stronger activation in response to human than animal vocalizations (Fig. 5*c*). This network principally included bilateral superior temporal and temporo-parietal cortices. It is noteworthy that in these regions group average responses to human vocalizations were ~1.5 times those to animal vocalizations (note difference in scales across Fig. 5*a*,*b*). Figure 5*d* displays the statistical difference between these source estimations, which after applying our threshold criteria yielded one cluster of 13 voxels that was located in BA22/41 in the right hemisphere [maximal *t* value at 47, −22, 6 mm, using the coordinate system of Talairach and Tournoux (1988)]. This distributed difference in absolute source strength is likely the basis for our observation of a GFP modulation in the absence of topographic effects, even though statistical differences between source estimations were spatially restricted with the threshold we applied.

Source estimations were also performed over the 291–357 ms period. Both conditions included prominent sources along the superior temporal lobes bilaterally with additional sources evident posteriorly at the temporo-parietal junction and also within the occipital lobe (Fig. 6*a*,*b*). Responses to human vocalizations also included prominent sources within the prefrontal and inferior frontal cortices bilaterally, although somewhat more strongly within the left hemisphere. Statistical contrast of these

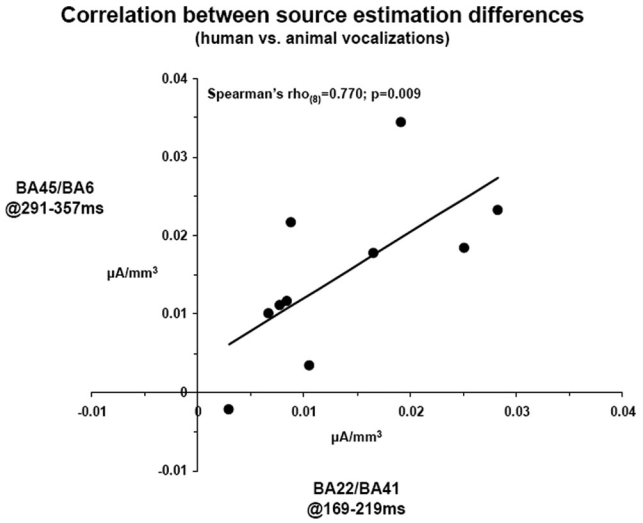

Figure 7. Linear correlation across the activation difference between responses to human and animal vocalizations within the two clusters shown in Figures 5*d* and 6*c*, *x*-axis and *y*-axis, respectively.

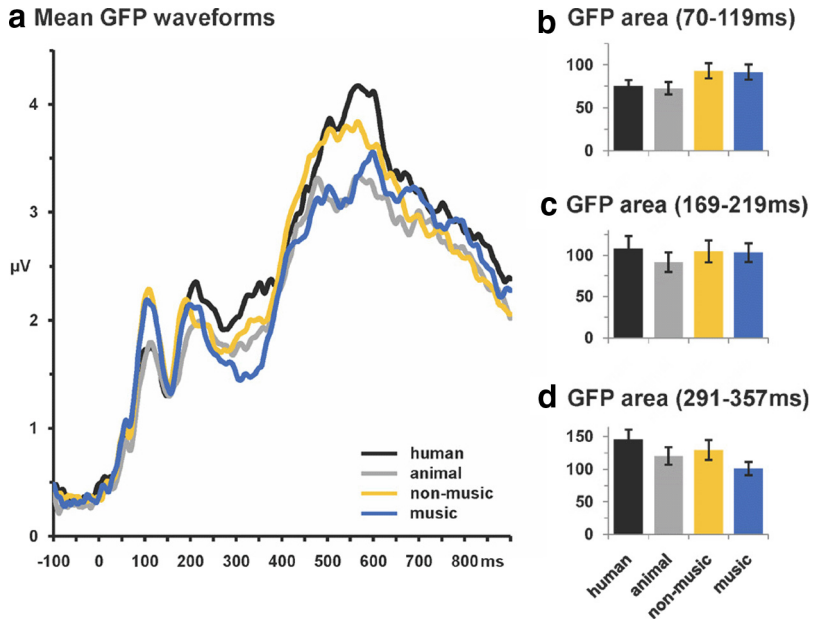

Figure 8. ***a***, Group-averaged GFP waveforms in response to vocalizations as well as two classes of sounds of man-made objects. ***b–d***, GFP area measures (SEM indicated) over selected poststimulus intervals.

source estimations identified a single cluster of 30 significant voxels that extended across BA45 and BA6 in the left hemisphere (maximal *t* value at −53, −3, 40 mm) (Fig. 6*c*). Thus, distinct time periods of differential processing of human vocalizations modulate responses in spatially disparate brain regions (i.e., BA22/41 at 169–219 ms vs BA45/6 at 291–357 ms). This pattern of results suggests that conspecific vocalization discrimination in humans likely involves a wide network of brain regions, each of which is potentially performing a distinct computation at a specific poststimulus latency.

A final set of source estimations was conducted over the 389–667 ms period and revealed sources similar to those observed during the preceding time periods (supplemental Fig. 2, available at www.jneurosci.org as supplemental material). Statistical comparisons revealed a single cluster of nine voxels within BA10 in the left hemisphere (maximal *t* value at −23, 53, 15 mm). As this cluster did not meet our spatial extent threshold, we only discuss it as a basis for generating hypotheses for future research.

We next examined functional coupling of differential responses to human versus animal vocalizations not only across brain regions but also across time intervals of differential processing. Differences in scalar values of source estimations within the right BA22/41 at 169–219 ms were positively correlated with differences in scalar values of source estimations within the left BA45/6 at 291–357 ms (Spearman's $\rho_{(8)} = 0.770$; $p = 0.009$) (Fig. 7).

Vocalization selectivity: across-object AEPs

To further situate effects of vocalization discrimination with respect to other object categories, we compared the GFP of responses to human and animal vocalizations (i.e., the data described and analyzed above) with that of the music and non-music conditions recorded during the same experiment [detailed in the study by Murray et al. (2006)]. The group average GFP waveforms are shown in Figure 8*a*. Based on the above analyses as well as our previous evidence showing living versus man-made categorization effects over the 70–119 ms period (Murray et al., 2006), we calculated GFP area over the 70–119, 169–219, and 291–357 ms poststimulus intervals and subjected them to separate ANOVAs using sound variety (human, animal, non-music, and music) as the within-subject factor (Fig. 8*b–d*). Over the 70–119 ms period, there was a main effect of sound variety ($F_{(3,7)} = 4.43$; $p < 0.05$) that was attributable to stronger responses to either man-made variety than to either human or animal vocalizations. This (unsurprisingly) replicates our previous work (Murray et al., 2006), even though the AEPs here were calculated for each condition with a lower number of trials (and therefore lower signal-to-noise ratio). Over the 169–219 ms period, there was also a main effect of sound variety ($F_{(3,7)} = 8.17$; $p = 0.01$). In this case, responses to human and animal vocalizations significantly differed from each other ($p < 0.005$), which is not surprising in view of all of the above analyses with these data. By contrast, however, responses to neither type of vocalization significantly differed from responses to either subtype of man-made sounds (all values of $p > 0.45$). Over the 291–357 ms period, there was again a main effect of sound variety ($F_{(3,7)} = 13.87$; $p = 0.002$). Responses to human and animal vocalizations significantly differed from each other ($p < 0.0001$), and responses to non-music and music significantly differed from each other ($p < 0.008$). Thus, although there is evidence for the discrimination of vocalizations from each other there is no evidence for selectively stronger responses to human vocalizations over other varieties of sounds (i.e., sounds of subcategories of man-made objects). These collective findings across multiple time periods thus sharply contrast with the observation that responses to human vocalizations are stronger than those to various classes of man-made sounds (Belin et al., 2000) [but see Lewis et al. (2005) and Engel et al. (2009) for evidence for distinct networks for different categories of sounds].

More generally, these analyses involving a larger set of sound categories allow us to establish a temporal hierarchy of auditory

object discrimination, with living versus man-made sounds discriminated first at 70–119 ms, followed by human versus animal nonverbal vocalizations at 169–219 ms, and later still by continued discrimination of vocalizations as well as the discrimination of subtypes of man-made sounds at 291–357 ms (Murray and Spierer, 2009; Spierer et al., 2010). These temporal considerations provide a complementary argument against a model of facilitated and/or selective discrimination of conspecific vocalizations, because at no latency were responses to human vocalizations reliably distinct from those to all other sound categories tested here. Moreover, there was no evidence to indicate that human vocalizations are subject to earlier discrimination than other categories of sound objects. Rather, the high temporal resolution of our data provide evidence for the contrary (i.e., that a general-level of living/man-made categorization precedes discrimination of human vocalizations).

Vocalization discrimination: single-object AEPs

Despite the above acoustic analyses, it could be argued that the differences between across-object AEPs is the result of undetected acoustic differences. To address this possibility, we calculated within-object AEPs (for details, see Materials and Methods). This generated a matrix of 20 single-object AEPs × 10 subjects. The GFP waveforms for each object (averaged across subjects) are shown in Figure 9*a*. GFP area measures were calculated over the 169–219 ms poststimulus interval (i.e., the time period when vocalization discrimination was identified using across-object AEPs) (Fig. 9*b*). These were submitted to a univariate ANCOVA using vocalization type as the fixed factor, subject as the random factor, and f_0 of the objects as a covariate. There was a significant effect of vocalization type ($F_{(1,187)} = 9.51$; $p = 0.002$), thereby replicating the observation of human versus animal vocalization discrimination using a more fine-grained analysis of AEPs. There was no evidence that vocalization type covaried with f_0 ($F_{(2,187)} = 1.43$; $p = 0.242$), providing no indication that vocalization discrimination was (directly) linked to this low-level acoustic feature of the stimuli. We also assessed whether single-object GFP area measures over the 169–219 ms period from individual subjects correlated with f_0. Nonparametric correlations were calculated. None of the 10 subjects exhibited a significant correlation between these measures (Spearman's $\rho_{(18)}$ ranged from 0.392 to 0.020; p values ranging from 0.09 to 0.91). Similarly, there was no evidence for a significant correlation when using the group average GFP area measures (Spearman's $\rho_{(18)} = 0.139$; $p = 0.56$) (Fig. 9*c*).

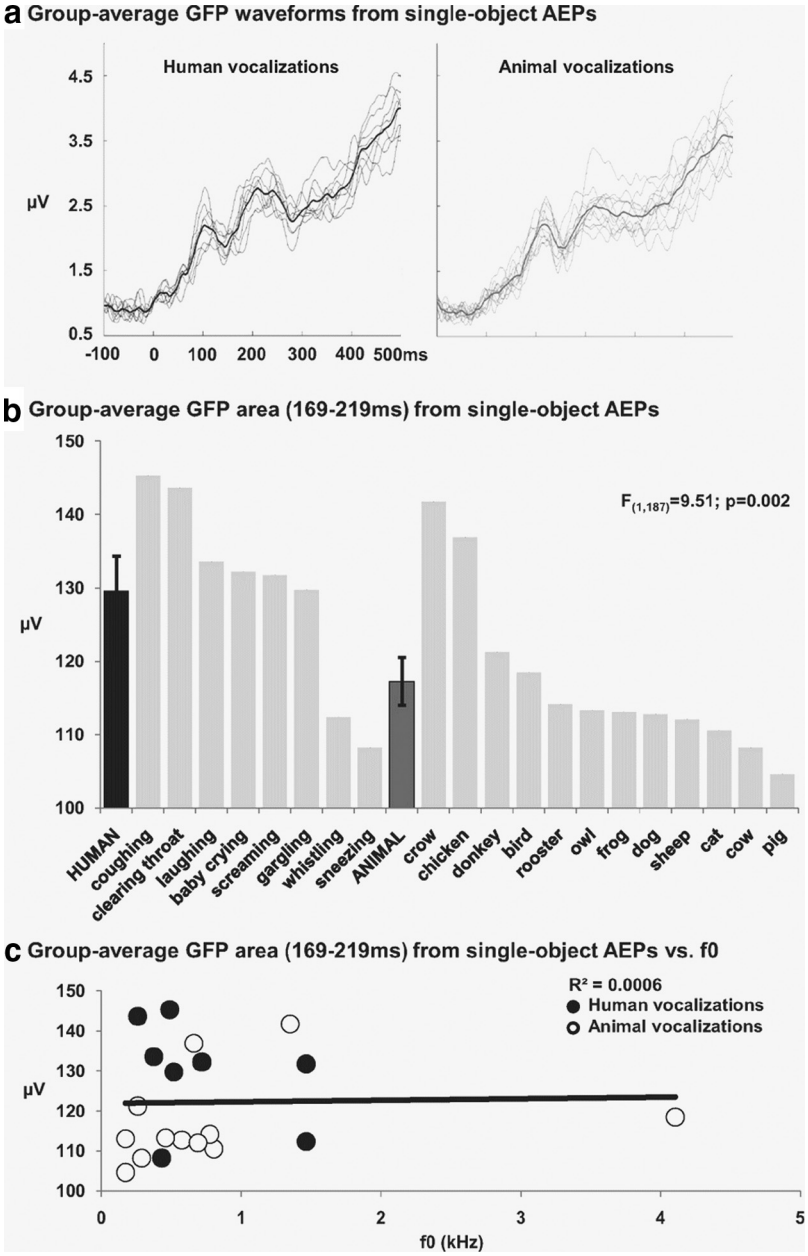

Figure 9. Results of single-object AEP analyses. ***a***, Group-averaged GFP waveforms are displayed for each vocalization (left panel, human vocalizations; right panel, animal vocalizations) along with the mean across vocalizations (thicker lines). ***b***, GFP area taken over the 169–219 ms poststimulus interval for each single-object AEP (light gray bars) as well as the average for each category of vocalizations (black bar, human vocalizations; dark gray bar, animal vocalizations; SEM indicated). The inset displays the main effect of object category after conducting a univariate ANCOVA (see Results). ***c***, Scatterplot comparing GFP area over the 169–219 ms period and the corresponding f_0 value for each object (black dots refer to human vocalizations and white dots to animal vocalizations). There was no evidence for a systematic relationship between these measures ($R^2 = 0.0006$).

Discussion

Electrical neuroimaging analyses identified the spatiotemporal dynamics of conspecific vocalization discrimination in humans. Responses were significantly stronger to conspecific vocalizations over three poststimulus periods. The first (169–219 ms) followed from strength modulations of a common network within the right STS and extending into the superior temporal gyrus (STG) and was functionally coupled with a subsequent difference at 291–357 ms within the left inferior prefrontal gyrus and precentral gyrus. The third effect (389–667 ms) followed from strength as well as topographic modulations and was lo-

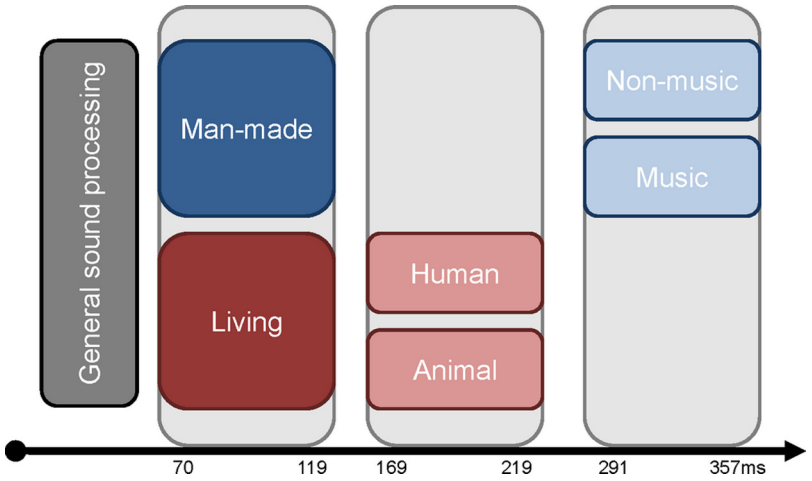

Figure 10. Schematic representation of a temporal hierarchy in auditory object discrimination summarizing the results of this study. Categorical effects on GFP are shown as a function of time relative to when they are first observed (subsequent effects not shown for simplicity). In a hierarchical fashion over time, general sound processing (initial 70 ms) is followed by living versus man-made discrimination (70–119 ms), then by human versus animal vocalization discrimination (169–219 ms), and finally by the discrimination of musical instruments versus other man-made objects (291–357 ms).

calized to the left superior frontal gyrus. These results support several conclusions regarding the mechanisms subserving vocalization discrimination in humans. First, the initial stages of vocalization discrimination are based on modulations in response strength within a statistically indistinguishable network of brain regions. Additional control analyses with single-object AEPs ruled out explanations of the earliest differentiation in terms of low-level acoustics (Fig. 9). Second, at no latency were responses to human vocalizations stronger than those to a wider set of object categories, even though responses reliably differed from animal vocalizations (Fig. 8). Third, the latency of our effects allowed us to situate voice discrimination along a more general timeline of auditory object discrimination (Fig. 10). Vocalization discrimination lags general living/man-made categorization by ~100 ms (Fig. 8). There is no evidence that voices are subject to facilitated processing over other types of objects either in terms of recruiting a voice-selective module/network or in terms of the speed with which the brain performs its discrimination. Such notwithstanding, it is noteworthy that the latency of the earliest voice discrimination is nearly synchronous with effects of face discrimination (Bentin et al., 2007), supporting the possibility that voice and face processes unfold in parallel, mutually informing one another (Schroeder et al., 2008; Ghazanfar, 2009).

A principal outcome is that there was no evidence for the selectivity of responses to human vocalizations. Rather, the earliest effects were the consequence of modulations in response strength in the absence of reliable topographic differences. Parsimony argues for common (or at least a statistically indistinguishable) networks of brain regions varying in strength as a function of vocalization type. In line with these findings at the level of the surface-recorded AEPs, our source estimations identified highly similar distributions of active brain regions in response to both human and animal vocalizations over both of the initial time periods (Figs. 5, 6). Statistical differences were limited to focal brain regions, although absolute differences were more widely distributed (Fig. 5*c*). Additionally, because responses were always stronger for human than for animal vocalizations, conspecific vocalizations may represent a more salient stimulus (for corresponding findings in the monkey, see Petkov et al., 2008; for a discussion of auditory saliency maps, see Kayser et al., 2005). By performing additional analyses comparing GFP responses to a wider set of categories of sound sources, we showed that human vocalizations were at some latencies less salient (i.e., had significantly weaker responses) than sounds of man-made sources and were never significantly stronger than all other sound categories (although nonetheless stronger than animal vocalizations) (Fig. 8). Stronger responses would have been expected had human vocalizations been subject to selective processing (Belin et al., 2000).

Several aspects of our study allowed us to evaluate the intrinsic "tuning" of the auditory system to human vocalizations, which can be viewed as another approach to addressing the topic of functional selectivity. Previous AEP research has suggested that correlates of conspecific vocalization discrimination may depend on selective attention to voices (Levy et al., 2003), although studies in nonhuman primates have repeatedly demonstrated vocalization sensitivity without task requirements (Recanzone, 2008; Russ et al., 2008) or in anesthetized subjects (Tian et al., 2001). Here, participants performed a living/man-made discrimination with no requirement to discriminate vocalizations. Performance did not differ across vocalization types. Finally, the temporal information afforded by AEPs allowed us to situate the earliest vocalization-related difference in the AEPs (170 ms) both with respect to mean reaction times on this task (~900 ms) and also with respect to target-distracter AEP differences (100 ms) (Murray et al., 2006), the latter of which provides an upper temporal limit on the speed by which categorical and decision-related brain processes initiate. Thus, vocalization discrimination transpires subsequently to these processes and is therefore unlikely to be driving decision-related effects.

The timing of these effects is also highly consistent with predictions based on recordings in monkeys. Multisensory integration of specific face and voice signals peaks at ~85–95 ms within core and lateral belt cortices (Ghazanfar et al., 2005, 2008). The selectivity of these integration effects suggests that categorization of voices occurred within this latency. However, the temporal dynamics of vocalization discrimination has to our knowledge not been specifically assessed in this or other studies in monkeys. Nonetheless, applying a "3:5" conversion ratio between latencies in macaques and humans (Schroeder et al., 2008) would suggest that vocalization discrimination in humans should manifest around 150–160 ms after stimulus. Although near-synchronous timing of face and vocalization discrimination has been previously hypothesized (Belin et al., 2004), previous AEP studies have hitherto produced discordant results that moreover cannot be unequivocally attributed to vocalization discrimination. Effects recently reported at 164 ms appeared to be driven by the speech content of the stimuli [cf. Charest et al. (2009), their Fig. 4]. Others reported effects at ~320 ms, although these depended on participants' attention to the voices (Levy et al., 2001, 2003; Gunji et al., 2003) and might also be explained as resulting from more general discrimination of living versus man-made objects because a musical instrument was used for the main contrast. The present study circumvented these caveats not only in the para-

digm but also in the use of electrical neuroimaging analyses with across-object and single-object AEPs. These analyses firmly situate the timing of conspecific vocalization discrimination at latencies consistent with observations in nonhuman primates and contemporaneous with face discrimination.

In addition to their timing and likely mechanism, we localized differential processing of conspecific vocalizations first to BA22/BA41 in the right hemisphere (169–219 ms) and subsequently to BA45/6 in the left hemisphere (291–357 ms), although we would note that a wider network of regions was also observed to be equally responsive to both types of vocalizations (Figs. 5, 6). These loci are in general agreement with previous hemodynamic imaging evidence in humans (Belin et al., 2000, 2002, 2004; von Kriegstein et al., 2003; Fecteau et al., 2005) and monkeys (Poremba et al., 2004; Petkov et al., 2008), as well as microelectrode recordings in monkeys (Cohen et al., 2007; Romanski, 2007; Recanzone, 2008; Russ et al., 2008).

The temporal information provided in the present study allows us to situate effects of vocalization discrimination with respect to general semantic analyses and task-related effects. Our previous research has shown that object discrimination processes already onset at 70 ms with task-related effects at 100 ms after stimulus (Murray et al., 2006). Aside from their consistency with human imaging, our findings are also highly consistent with recent imaging findings in awake monkeys showing a set of auditory fields whose activity was enhanced in response to conspecific vocalizations versus vocalizations from other animals and primates as well as phase-scrambled counterparts (Petkov et al., 2008). In particular, right-lateralized primary and posterior parabelt fields as well as bilateral anterior fields exhibited response enhancements. It should be noted, however, that imaging studies in nonhuman primates either limited their field of view (Petkov et al., 2008) or selected regions of interest (Poremba et al., 2004), leaving unknown the full spatial distribution of differential responses to vocalizations. More germane, differential activity during the 169–219 ms period observed in the present study extended across what are undoubtedly multiple distinct functional regions from the STG to the STS and middle temporal cortex. Together, the results of Petkov et al. (2008) and our own highlight the role of several distributed auditory regions in conspecific vocalization discrimination (Recanzone, 2008).

The distributed nature of these processes is all the more evident in the fact that several distinct time periods of differential responsiveness were observed. In particular, stronger source strengths within the left inferior prefrontal cortex in response to human versus animal vocalizations were observed over the 291–357 ms poststimulus period. Studies in both humans (Fecteau et al., 2005) and monkeys (Cohen et al., 2006, 2007; Romanski, 2007; Russ et al., 2008) have shown that prefrontal neurons respond differentially to conspecific vocalizations. One possibility is that the initial differentiation of human vocalizations within the right STS/STG is causally related to effects at 291–357 ms within left prefrontal cortices, particularly given the known connectivity between the temporal and frontal cortices (Romanski et al., 1999; Petrides and Pandya, 2007) (for the role of interhemispheric fibers, see also Poremba et al., 2004). This proposition receives some support from our analysis showing a significant positive correlation between response modulations at 169–219 ms within right BA22/41 and those at 291–357 ms within left BA45/6 (Fig. 7).

In conclusion, the present electrical neuroimaging findings reveal that voice discrimination transpires substantially earlier than conventionally held and occurs over multiple, functionally coupled stages in a wide network of brain regions. Such findings highlight that models of functional specialization must incorporate network dynamics.

References

Aeschlimann M, Knebel JF, Murray MM, Clarke S (2008) Emotional preeminence of human vocalizations. Brain Topogr 20:239–248.

Assal G, Aubert C, Buttet J (1981) Asymétrie cérébrale et reconnaissance de la voix. Revue Neurolog 137:255–268.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res 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.

Bentin S, Taylor MJ, Rousselet GA, Itier RJ, Caldara R, Schyns PG, Jacques C, Rossion B (2007) Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. Nat Neurosci 10:801–802; author reply 802–803.

Charest I, Pernet CR, Rousselet GA, Quiñones I, Latinus M, Fillion-Bilodeau S, Chartrand JP, Belin P (2009) Electrophysiological evidence for an early processing of human voices. BMC Neurosci 10:127.

Cohen YE, Hauser MD, Russ BE (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. Biol Lett 2:261–265.

Cohen YE, Theunissen F, Russ BE, Gill P (2007) Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. J Neurophysiol 97:1470–1484.

De Lucia M, Camen C, Clarke S, Murray MM (2009) The role of actions in auditory object discrimination. Neuroimage 48:475–485.

De Lucia M, Michel CM, Murray MM (2010a) Comparing ICA-based and single-trial topographic ERP analyses. Brain Topogr 23:119–127.

De Lucia M, Cocchi L, Martuzzi R, Meuli RA, Clarke S, Murray MM (2010b) Perceptual and semantic contributions to repetition priming of environmental sounds. Cereb Cortex 20:1676–1684.

Fecteau S, Armony JL, Joanette Y, Belin P (2004) Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage 23:840–848.

Fecteau S, Armony JL, Joanette Y, Belin P (2005) Sensitivity to voice in human prefrontal cortex. J Neurophysiol 94:2251–2254.

Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, Schweinberger SR, Warren JD, Duchaine B (2009) Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 47:123–131.

Ghazanfar AA (2009) The multisensory roles for auditory cortex in primate vocal communication. Hear Res 258:113–120.

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 25:5004–5012.

Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. Curr Biol 17:425–430.

Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci 28:4457–4469.

Grave de Peralta Menendez R, Gonzalez Andino S, Lantz G, Michel CM, Landis T (2001) Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain Topogr 14:131–137.

Gunji A, Koyama S, Ishii R, Levy D, Okamoto H, Kakigi R, Pantev C (2003) Magnetoencephalographic study of the cortical activity elicited by human voice. Neurosci Lett 348:13–16.

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48:373–384.

Knebel JF, Toepel U, Hudry J, le Coutre J, Murray MM (2008) Generating controlled image sets in cognitive neuroscience research. Brain Topogr 20:284–289.

Koenig T, Melie-Garcia L (2010) A method to determine the presence of averaged event-related fields using randomization tests. Brain Topogr 23: 233–242.

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 48:609–621.

Levy DA, Granot R, Bentin S (2001) Processing specificity for human voice stimuli: electrophysiological evidence. Neuroreport 12:2653–2657.

Levy DA, Granot R, Bentin S (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40:291–305.

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.

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging. Clin Neurophysiol 115:2195–2222.

Murray MM, Spierer L (2009) Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. Hear Res 258:121–133.

Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006) Rapid brain discrimination of sounds of objects. J Neurosci 26:1293–1302.

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20:249–264.

Murray MM, De Santis L, Thut G, Wylie GR (2009a) The costs of crossing paths and switching tasks between audition and vision. Brain Cogn 69:47–55.

Murray MM, De Lucia M, Brunet D, Michel CM (2009b) Principles of topographic analyses for electrical neuroimaging. In: Brain signal analysis: advances in neuroelectric and neuromagnetic methods (Handy TC, ed), pp 21–54. Cambridge, MA: MIT.

Perrin F, Pernier J, Bertrand O, Giard MH, Echallier JF (1987) Mapping of scalp potentials by surface spline interpolation. Electroencephalogr Clin Neurophysiol 66:75–81.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11:367–374.

Petrides M, Pandya DN (2007) Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. J Neurosci 27:11573–11586.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427:448–451.

Recanzone GH (2008) Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. J Neurosci 28:13184–13193.

Remedios R, Logothetis NK, Kayser C (2009) An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J Neurosci 29:1034–1045.

Romanski LM (2007) Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. Cereb Cortex 17 [Suppl 1]:i61–i69.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131–1136.

Russ BE, Ackelson AL, Baker AE, Cohen YE (2008) Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99:87–95.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113.

Spierer L, De Lucia M, Bernasconi F, Grivel J, Bourquin NMP, Clarke S, Murray MM (2010) Learning-induced plasticity in human audition: objects, time, and space. Hear Res. Advance online publication. Retrieved July 29, 2010. doi:10.1016/j.heares.2010.03.086.

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.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

Van Lancker DR, Canter GJ (1982) Impairment of voice and face recognition in patients with hemispheric damage. Brain Cogn 1:185–195.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48–55.

von Kriegstein K, Smith DR, Patterson RD, Ives DT, Griffiths TD (2007) Neural representation of auditory size in the human voice and in sounds from other resonant sources. Curr Biol 17:1123–1128.