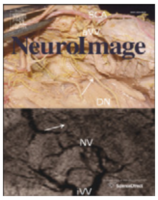
NeuroImage 60 (2012) 1704–1715

Contents lists available at SciVerse ScienceDirect

NeuroImage

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



Full Length Article

Auditory perceptual decision-making based on semantic categorization of environmental sounds

Marzia De Lucia [a,b,*], Athina Tzovara [a,b], Fosco Bernasconi [c], Lucas Spierer [c,d], Micah M. Murray [a,b,c,e]

[a] Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, Switzerland
[b] Department of Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), CH-1011 Lausanne, Switzerland
[c] Neuropsychology and Neurorehabilitation Service, Department of Clinical Neurosciences, CHUV, CH-1011 Lausanne, Switzerland
[d] Neurology Unit, Department of Medicine, Faculty of Science, University of Fribourg, Fribourg, Switzerland
[e] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

ARTICLE INFO

Article history:
Received 29 April 2011
Revised 24 January 2012
Accepted 25 January 2012
Available online 8 February 2012

Keywords:
Perceptual decision-making
Single-trial electroencephalography
Decoding

ABSTRACT

Discriminating complex sounds relies on multiple stages of differential brain activity. The specific roles of these stages and their links to perception were the focus of the present study. We presented 250 ms duration sounds of living and man-made objects while recording 160-channel electroencephalography (EEG). Subjects categorized each sound as that of a living, man-made or unknown item. We tested whether/when the brain discriminates between sound categories even when not transpiring behaviorally. We applied a single-trial classifier that identified voltage topographies and latencies at which brain responses are most discriminative. For sounds that the subjects could not categorize, we could successfully decode the semantic category based on differences in voltage topographies during the 116–174 ms post-stimulus period. Sounds that were correctly categorized as that of a living or man-made item by the same subjects exhibited two periods of differences in voltage topographies at the single-trial level. Subjects exhibited differential activity before the sound ended (starting at 112 ms) and on a separate period at ~270 ms post-stimulus onset. Because each of these periods could be used to reliably decode semantic categories, we interpreted the first as being related to an implicit tuning for sound representations and the second as being linked to perceptual decision-making processes. Collectively, our results show that the brain discriminates environmental sounds during early stages and independently of behavioral proficiency and that explicit sound categorization requires a subsequent processing stage.

Introduction

We are constantly immersed in environments where accurate and quick sound identification is essential for daily activities such as communicating and navigating. We recognize, for example, whether the sound was emitted by a machine (i.e. telephone ringing) or produced by a living source (i.e. a dog barking) even in the absence of visual cues and in noisy contexts. Correct sound identification depends on the amount of available sensory evidence, on the ability to compare different options within a range of possibilities and often implies making a choice, i.e. a perceptual decision.

Investigating how sensory information is gathered and used to form a decision variable represents an active field of research in both monkeys and humans (reviewed in Gold and Shadlen, 2007; Heekeren et al., 2008). In both visual and somatosensory domains, several studies emphasize a causal link between sensory evidence accumulated in low-level cortices and perceptual decision-making. The idea is that even if the actual decision is made by higher-level cortices by integrating accumulated sensory evidence, stimulus identification can already be inferred from activity in low-level sensory areas. During a vibrotactile discrimination experiment, single-unit recordings in somatosensory cortex predicted whether monkeys could perceive a difference in the stimulation frequency (Salinas et al., 2000). In the same kind of tasks, Romo et al. (Romo et al., 2003) demonstrated the role of medial prefrontal and premotor cortices in forming the decision initially driven by evidence in somatosensory areas (see also de Lafuente and Romo, 2005; Hernandez et al., 2002). The same pattern of sequential stages has been repeatedly shown in the visual domain. Here, a popular paradigm involves discriminating the direction of motion of a field of moving dots (Bennur and Gold, 2011; Britten et al., 1992; Kim and Shadlen, 1999; Newsome et al., 1989; Shadlen and Newsome, 1996) during which activity in area MT could predict motion discrimination, whereas activity in the lateral intraparietal area (LIP), the frontal eye-field (FEF) and the dorsolateral prefrontal cortex (DLPFC) was directly related to forming the decision. More recently, perceptual decision-making in the visual domain has been investigated

---

* Corresponding author at: Electroencephalography Brain Mapping Core, Lemanic Center for Biomedical Imaging, Centre Hospitalier Universitaire Vaudois, BH 07 081 1, Department of Biology and Medicine, University of Lausanne, 1011 Lausanne, Switzerland. Fax: +41 21 314 46 45.
E-mail address: marzia.de-lucia@chuv.ch (M. De Lucia).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.01.131

with complex stimuli and tasks involving perceptual and semantic categorization (Murphy et al. 2011; Philiastides and Sajda, 2006; Philiastides et al., 2006; Ratcliff et al., 2009; Simanova et al., 2010).

To date, few studies have examined perceptual decision-making in the auditory domain and were mainly based on the discrimination of phonemes (Binder et al., 2004; Kaiser et al., 2006). Binder et al. carried out a functional magnetic resonance imaging (fMRI) experiment where participants were asked to identify speech sounds masked by varying levels of noise (Binder et al. 2004). Activity within a cortical region anterior to primary auditory cortex predicted performance accuracy, whereas activity in the inferior frontal cortex was related to response time. With more emphasis on the temporal dynamics of perceptual decision-making in the auditory domain, Kaiser et al. (2006) showed differential activity in the gamma frequency band starting at 120 ms post-stimulus onset when comparing different sounds patterns, and differential activity between easy and difficult decisions at later stages of sound processing. However, localizing or at least distinguishing the cortical regions underlying these sequential stages in humans remains still unexplored in speech identification and even more in the context of complex sound identification.

Recently, the analysis of auditory event-related potentials (AEPs) in response to complex stimuli revealed that sound processing starts unfolding at 70 ms post-stimulus onset for living/man-made categorizations (Murray et al. 2006) and that finer semantic analysis takes place along later discontinuous temporal periods including the discrimination between human and animal voices (De Lucia et al. 2010a), and musical and other man-made sounds (De Lucia et al. 2009). These results show that latencies of sound processing and discrimination can depend critically on the level of semantic categorization. Given that the participants were not explicitly asked to identify sound subcategories in these studies, the specific functional role of each of these differential stages and their link to perception remains unresolved.

Here, we apply a multivariate analysis to AEPs in response to sounds of living and man-made objects while participants were explicitly instructed to categorize (and identify) the sound. The ability to categorize sounds was challenged by manipulating the sound duration. In a preliminary psychophysics experiment we evaluated the duration for which the number of recognized and unrecognized sounds was roughly counterbalanced. This sound duration was then used for the main experiment. In the analysis, we separately considered those trials corresponding to sounds that subjects could correctly categorize and those for which subjects were not able to provide an answer. Our main goal was to evaluate whether differential activity to the living and man-made semantic categories manifests irrespective of behavioral outcome and to uncover the electrophysiological response linked to subjects' ability to perceive the sound's category. We hypothesized that living and man-made categorization is still detectable in the EEG response even when subjects do not categorize the sounds and that a second (later) modulation is present only for correctly categorized sounds. This hypothesis stems from the abovementioned works on perceptual decision-making showing the existence of at least two stages in stimulus processing; the first one related to a coarse level stimulus representation, the second (later in time) to integration of available sensory information. We based our analyses on a multivariate decoding approach revealing at which latency it is possible to accurately decode the semantic category of the sounds in presence or not of correct categorization. Importantly, our method is based on modulation in voltage topographies (Bernasconi et al., 2011; De Lucia et al., 2007, 2010b; Murray et al., 2009; Tzovara et al., 2012; Tzovara et al., in press) which takes advantage of the overall voltage configuration at the scalp without *a priori* selection of specific electrode locations.

Materials and methods

In the following we first present a pilot study whose aim was to select the sound duration at which the number of sounds identified was roughly equal to that unidentified. We then present the main experiment of this study that used the same sounds from the pilot study (and the duration estimated based on the pilot study results) in combination with the recording of EEG.

Pilot study

Subjects

Fifteen healthy subjects (all men aged 20–26 years old) participated in a psychophysical test. All subjects were right-handed and provided written, informed consent to participate in the study, the procedures of which were approved by the Ethics Committee of the *Centre Hospitalier Universitaire Vaudois* and University of Lausanne. No subject had a history of neurological or psychiatric illness, and all reported normal hearing and vision (or corrected-to-normal).

Stimuli

Auditory stimuli were complex, meaningful sounds (16 bit mono; 44,100 Hz digitalization) of living and man-made objects, including 108 sounds (53 sounds of living objects) (Supplemental Material 1). Sounds of living items included human non-verbal vocalizations and animal vocalizations (15 and 38, respectively). Within the man-made category there were 12 sounds of musical instruments. We prepared three sound sets each including 108 stimuli, which differed in the sound duration: 100 ms, 250 ms, and 500 ms.

Each subject was tested using one fixed duration of the sounds to avoid any priming effects. Initially, we planned to test 5 subjects for each duration. However, as detailed below, it was immediately evident from the first 3 subjects that the 100 ms duration was too short for accurate performance. Consequently, 3 subjects were tested with 100 ms duration sounds, 7 subjects were tested with 250 ms duration sounds, and 5 subjects were tested with 500 ms duration sounds. Sounds were presented via insert earphones (model ER-4P; Etymotic Research) at an individually adjusted volume, and subjects were asked to categorize the sound as living or as man-made.

Results

With the 100 ms duration the first three subjects were unable to categorize or identify the majority of the presented sounds. We decided therefore to exclude the 100 ms sound duration from further testing. With the 500 ms duration sound bank, the five subjects categorized the overwhelming majority of the sounds. Finally, we observed that the best sound length was 250 ms. In this case we tested seven subjects (five subjects initially planned plus those that we did not test with 100 ms sound duration). With 250 ms sound duration, five of the seven subjects could correctly categorize 75 sounds (within which 38 were also correctly identified) and could not categorize 33 sounds. We concluded that 250 ms duration was optimal within the three tested durations for obtaining a balanced number of identified and unidentified sounds.

EEG experiment

Subjects

Nine healthy subjects (4 women), aged 20–30 years participated in the EEG study. All subjects provided written, informed consent to participate in the study, the procedures of which were approved by the Ethics Committee of the *Centre Hospitalier Universitaire Vaudois* and University of Lausanne. All subjects were right-handed (Oldfield, 1971). No subject had a history of neurological or psychiatric illness, and all reported normal hearing and vision (or corrected-to-normal).

Procedure and task

Subjects listened to each sound and, after a 700 ms intervening period, they were first asked to indicate via a 3-alternative-forced-choice button press its category (living, man-made, or unknown) and second indicated via a 2-alternative-forced-choice button press if they were able to identify the object (yes vs. no). After one second, an image was presented on the screen for 1000 ms. This image represented or not the same object as the initially presented sound. After 800 ms subjects listened to the same sound and were asked to answer to the same questions as after the first presentation. Subjects were asked to answer each question with the maximum of confidence (the speed of the answer was not emphasized, but we limited the time to answer to 5 s). During the EEG recording, the stimuli were divided into 14 blocks, each block was about 6 min long and included 50 sequences of sound, image and sound presentation. Sounds were randomly selected from the complete list of sounds and therefore some sounds were sometimes repeated within each block and across blocks. Between each block a break was proposed. In the following, we consider only the first part of the experiment; that is to say the initial sound presentation and the behavioral outcome of the categorization task only.

EEG acquisition and preprocessing

Continuous 160-channel EEG was acquired through a Biosemi system (1024 Hz sampling rate, offline band-pass filtered 0.1–40 Hz). Peri-stimulus epochs of continuous EEG (−100 to 400 ms) were extracted for each subject and for each living/man-made condition. Trials with blinks or eye movements were rejected off-line. An automated artifact rejection criterion of ±100 μV was applied at all electrodes, and EEG epochs were also visually evaluated. Data from artifact electrodes from each subject and condition were interpolated using three-dimensional splines (Perrin et al., 1987). On the basis of the behavioral outcome for each subject, we separately considered three possible subgroups of trials: those in response to sounds that subjects correctly categorized as living or man-made, trials for which subjects could not guess (i.e. answered ‘unknown’) and those in response to sounds for which subjects gave a wrong answer (i.e. categorize as living a man-made sound or vice versa). Because this trial selection was performed for each subject and depending on his/her specific behavioral output, epochs in response to different sounds where included in the same subcategories across subjects. Artifact-free trials in response to man-made sounds that subjects misjudged as living were relatively fewer than those in response to sounds of living items that subjects categorized as man-made. We therefore decided to ignore this subgroup of trials in the following analysis. In summary, we considered only two subgroups of responses: those in response to sounds whose category was correctly categorized and those for which subjects could not say whether the sound was living or man-made (hereafter termed *uncategorized*).

Datasets definition

For both the uncategorized and correctly categorized stimuli, we first considered two sets of single trials, one in response to sounds of living items and one in response to man-made sounds. We used part of these trials for training and testing the classifier as explained in **Classification procedure** section. In addition, for both the uncategorized and correctly categorized sounds, we considered an independent dataset (validation dataset) that we used to have a final evaluation of the classification performance (see next two sections for details on the number of trials for training/testing and validation). Keeping separate the datasets used for optimizing the classifier's parameters and the one used for estimating the classification performance is essential for assessing how well the classifier generalizes to previously unseen data (Kriegeskorte et al., 2009). For both categorized and uncategorized cases, we randomly selected the trials included in the training/testing dataset and the validation dataset within the whole set of artifact-free single trials available. This random selection is fixed once before all the analyses.

As will be clear in the following, we ran the analysis at the group level, finding a common spatio-temporal pattern across all subjects (for similar approaches see Shinkareva et al., 2011). This analysis is performed separately for uncategorized and categorized sounds.

Datasets in response to uncategorized sounds

Only five of the subjects (one woman) had at least 10 artifact-free trials in response to uncategorized sounds for each of the two semantic categories (living and man-made). All the analyses described in the following refer therefore to these 5 subjects only. The whole list of uncategorized sounds is shown in the Supplemental Material 2 (see also behavioral results for the EEG experiment). The dataset in response to sounds that subjects could not guess and that we used for training and testing the classifier included 50 trials from the living category and 50 trials from the man-made category overall. Collapsing data across subjects was necessary to include a sufficient number of trials for the classification analysis. The remaining single-trials were used to validate the classification performance. Specifically we consider as validation dataset a total of 77 single-trials in response to sounds of living items (that were not categorized). These data were extracted from all the 5 subjects ranging from 1 to 29 trials from each of them. By contrast the single trials in response to man-made sounds that we could use for the validation were too few (14 in total across the 5 subjects). We therefore decided to validate the classifier performance on responses to sounds of living items.

Datasets in response to correctly categorized sounds

For the same cohort of five subjects, we considered a total of 40 single trials for each subject and each semantic category for training and testing the classifier. This number of single trials was chosen so as to limit the computational time required for training and testing the classification performance at the group level (that is to say when collapsing single-trial data across subjects and therefore including 200 single-trials per condition).

In addition, we considered a validation dataset that included a total of 40 trials across subjects for each category, living and man-made.

Sound acoustic analysis

We analyzed the sets of living and man-made sounds whose single-trial EEG responses were included in the datasets as described above. The goal was to assess whether the sets of living and man-made sounds (the uncategorized and correctly categorized were analyzed separately) differed acoustically.

For both uncategorized and correctly categorized sounds, we first considered living and man-made sounds for each subject separately including only those sounds that were used in the training and testing phase. In addition, for the correctly categorized sounds, we considered the set of sounds collapsed across subjects whose single-trial EEG responses were included in the validation datasets. This analysis was not possible in the case of the uncategorized sounds because the validation datasets included only EEG responses to living sounds.

The acoustic analysis between living and man-made sounds consisted in statistically comparing 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 (De Lucia et al., 2009, 2010a; Knebel et al., 2008). 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 $p<0.05$ (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 a result of these analyses, we found minor statistically reliable differences at the single subject level for uncategorized sounds (Fig. 1, panels S1–S5). It is worth noting that each of the ten test datasets included single trials belonging to one single subject only. Therefore in the training/testing phase, the classifier discriminated between trials in response to living and man-made sounds that exhibited minor acoustical differences.

By contrast we found statistical differences at the single subject level in the case of correctly categorized sounds (not shown). Finally there were only minor statistically reliable differences between the sets of correctly categorized living and man-made sounds of the validation dataset (Fig. 1, bottom left panel).

Classification procedure

The aim of the study was decoding the semantic category of each sound presented to the subjects both when the sound was correctly categorized and when subjects could not indicate the sound's category. Decoding was here based on a classification algorithm that identified at a single-trial level the sound's semantic category. The classification analysis was run separately for correctly categorized and uncategorized sounds.

For each subject and semantic category, we partitioned the training/testing dataset into ten splits. Training was performed based on nine splits; the remaining (independent) split was used to test the classifier. Training and testing were repeated ten times using different splits and in a way that testing was always run on an independent partition of the data. Classification performance was assessed based on average classification accuracy (ratio of correctly classified single-trials) across the ten test datasets and on the validation dataset. However, only the classification performance obtained on the validation dataset is a reliable estimation the classification accuracy. Indeed, this set of trials had not been used at any point for training and for selecting the parameters of the classifier.

Method for single-trial classification

Preprocessing

For each experimental condition (living and man-made), we considered the concatenation of peri-stimulus EEG epochs (trials). At each time point, we represent the data across the electrode montage as a vector of voltage measurements $\mathbf{m} = \{m_1, m_2, \ldots, m_N\}$, where N is the total number of electrodes (Fig. 2a). We refer to this vector as topography. These trials were pre-processed by normalizing each topography $\mathbf{m}$ by its total EEG power time point by time point. The aim of this normalization is to minimize the difference between voltage topographies due to the strength of the field evoked on the scalp (and to emphasize instead differences due to the overall shape of the topography). In the following, we refer to $\mathbf{m}$ as a generic normalized

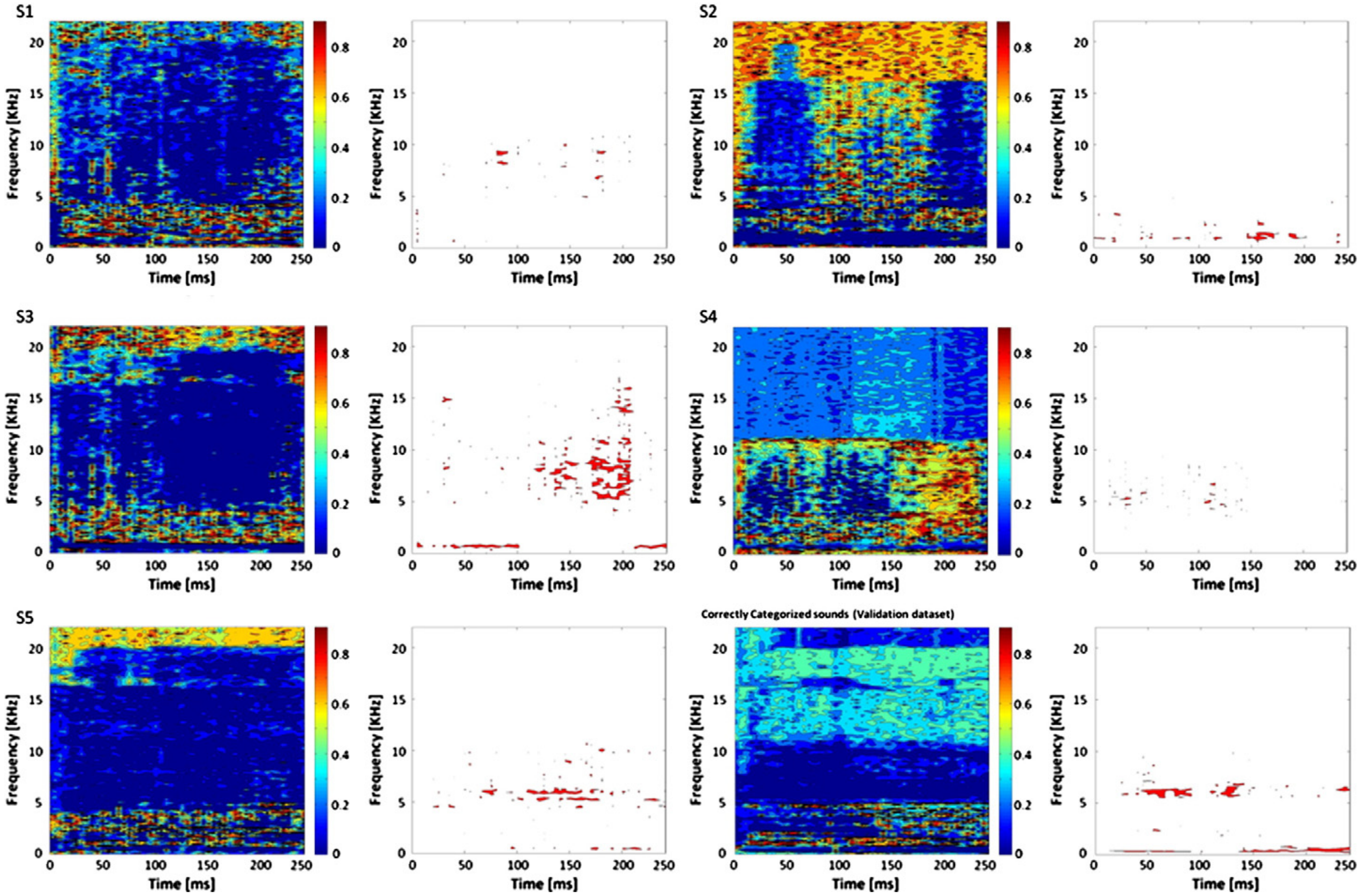


Fig. 1. Statistical comparison between spectrograms of living and man-made sounds. For each panel we show on the left the *p* values of the results of this comparison at each time–frequency bin (~5 ms and ~80 Hz); on the right we show in red the bins meeting the following statistical criterion: eight spatially contiguous bins exhibit statistically significant results (non-parametric *t*-test; $p<0.05$ equivalent to a cluster-level value of $p<0.00625$). (Black-framed panels) Results of the analyses conducted on the sounds included in the training and testing datasets of uncategorized sounds for each subject separately (numbered from 1 to 5). (Blue-framed panel) Results of the comparison between the spectrograms of living and man-made sounds that were correctly categorized and whose single-trial EEG responses were included in the validation dataset.

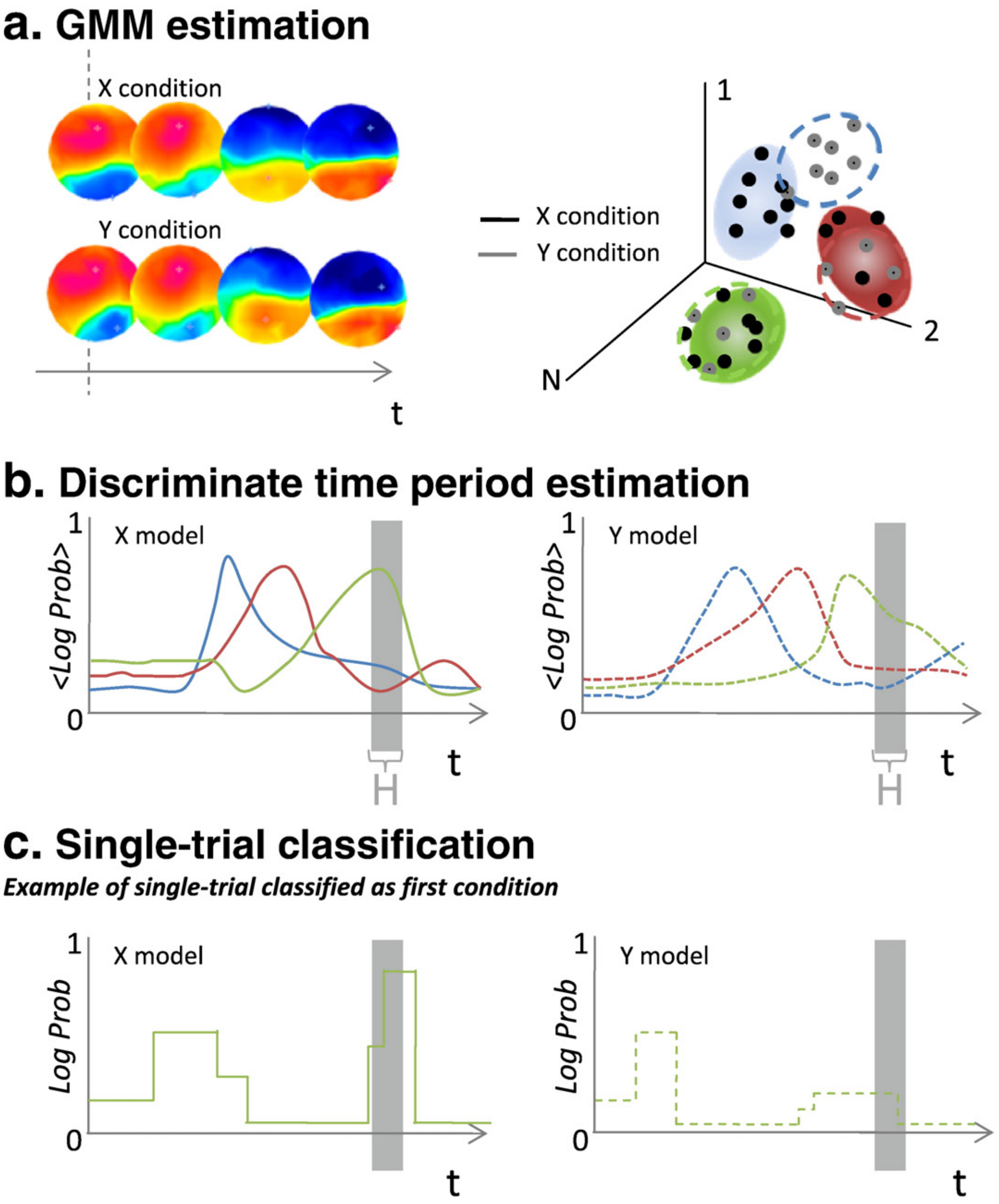


Fig. 2. Schematic representation of the single-trial classification. a. (left panel) At each time-point, the topography is represented as a vector in N-dimensional space (where N = #electrodes). (right panel) At this stage we do not take into account the latency and the trial to which it belongs and all the voltage topographies are pooled together in an N-dimensional space. Only the semantic category information is kept in this representation. Each dataset is modeled as a GMM; in the figure we show two GMMs with three Gaussians in the mixture. b. Based on GMM models' parameters, each topography is assigned a set of posterior probabilities, the number of which equals that of the Gaussians in the mixture. The set of posterior probabilities values are re-arranged time-point by time-point and re-assigned to each of the original trials. Left and right panels show the average posterior probabilities across trials for each of the two models and datasets, together with the discriminative time-period H where the two conditions are most different. c. Example of posterior probabilities of a test single-trial. Only those posterior probabilities that appeared during the discriminative time-period are relevant for the classification. This single-trial is classified as belonging to the X condition because the green posterior probability estimated from the X model (left panel) is higher than estimated from the Y model (right panel), during the relevant time-period.

topography at one time point. T indicates the total number of trials in the training dataset for each condition.

Training phase

Gaussian mixture model (GMM) estimation. The first step of our analysis consisted of modeling the statistical distribution of the ensemble of topographies in the training dataset (see Tzovara et al., 2012 for more details). At this stage of the analysis, each experimental condition (living/man-made) was processed separately and all the available topographies were pooled together disregarding the latencies at which they were observed and the trial to which they belonged (Fig. 2a, right panel).

We considered a GMM probability distribution in an N-dimensional space (Fig. 2a right panel)

$$P(\boldsymbol{m}|\mu,\sigma) = \sum_{k=1}^{Q} p_k P_k(\boldsymbol{m}|\boldsymbol{\mu}_k,\boldsymbol{\sigma}_k) \qquad (1)$$

where P_k is the kth Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\sigma}_k$, p_k is the prior probability of the class label k and Q is the total number of Gaussians. In the following, we refer to the means -$\underline{\boldsymbol{\mu}}_k$- of the Gaussians as template maps, and for simplicity we replace the notation "$\boldsymbol{\mu}_k$, $\boldsymbol{\sigma}_k$" with "c_k" to indicate the k-th Gaussian within the GMM.

The GMM estimation was based on an expectation–maximization procedure (Dempster et al., 1977; Bishop, 1995) that iterates the estimation of the model's parameters (priors, means and covariance matrices) in order to minimize the error function, or equivalently maximize the likelihood $\mathcal{L}$:

$$E = -\ln\mathcal{L} = -\sum_j \ln\left\{\sum_{k=1}^{Q} P(\boldsymbol{m}|c_k)p_k\right\} \tag{2}$$

where the index j spans the total number of topographies in the training dataset, i.e. the total number of topographies in one trial multiplied by the total number of trials in the training dataset for one condition. The initialization of the expectation–maximization procedure was based on a k-means clustering algorithm (Bishop, 1995) that provides a first guess of the means and the covariance matrices. The values for the priors, p_k, were obtained by the relative number of topographies for each cluster. Due to the limited number of training samples, we reduced the number of free parameters by constraining the covariance matrix to be diagonal.

Model estimation based on expectation–maximization requires knowing in advance the total number of Gaussians, Q, in the model (in the example of Fig. 2a right panel, Q = 3). In the following, we explored many possible values of Q for each condition, and we then selected the optimal values based on maximizing classification accuracy (see Optimizing the total number of Gaussians).

Evaluating the posterior probabilities of single-trial ERP data. Once the GMM models had been estimated, we could assign each topography of the original dataset to one of the Gaussians c_k in the model. This could be done by computing the posterior probabilities (Bishop, 1995: p. 23–27) defined as in the following:

$$P(c_k|\boldsymbol{m}) = \frac{P(\boldsymbol{m}|c_k)\cdot p_k}{p(\boldsymbol{m})} \tag{3}$$

where $p(\boldsymbol{m})$ is the unconditional density function, i.e. the density function for m irrespective of the Gaussian c_k.

In order to investigate stimulus-related information, we rearranged the posterior probabilities to their original temporal order in the data, providing a new representation of the single-trial ERPs in time and across trials (see Fig. 1 in Tzovara et al., 2012). For each trial, this rearrangement generated a time series of the posterior probability for each template map in the mixture.

The posterior probabilities across trials typically revealed a pattern of presence of a given template map that was structured in time (Fig. 2b, Figs. 2a–b), where $\boldsymbol{m}_{th}$ is the topography of t-th trial, at time-latency h. This posterior probability was computed at each latency and for each Gaussian *ck* in the GMM model for each experimental condition. At this point we could compare the two conditions by taking advantage of the model estimated on each of the two conditions, separately.

Identifying discriminative time periods between conditions. In the following $\boldsymbol{m1}_{th}$ and $\boldsymbol{m2}_{th}$ indicate the topographies of Condition 1 and Condition 2, respectively, appearing at the trial t and latency h; c_i and c_j refer to two generic Gaussians in the GMMs for the two conditions. Based on the two models and datasets, we computed $P(c_i|\boldsymbol{m1}_{th})$ and $P(c_j|\boldsymbol{m2}_{th})$ as shown in Eq. (3).

At each time-point h, we considered only one Gaussian per model, i.e. the one that yielded the highest posterior probability across trials within all the Q Gaussians in the model:

$$i_h = \underset{i\in 1:Q}{\operatorname{argmax}}\left\{\frac{\Sigma_{t=1}^{T}\log P(c_i|\boldsymbol{m1}_{th})}{T}\right\} \tag{4}$$

$$j_h = \underset{j\in 1:Q}{\operatorname{argmax}}\left\{\frac{\Sigma_{t=1}^{T}\log P(c_j|\boldsymbol{m2}_{th})}{T}\right\} \tag{5}$$

Within all the Gaussians in the two models, we expected that only a subset of them would be informative about the difference between the two conditions and possibly only along a certain time period (discriminative time period).

For identifying discriminative periods between single-trials in response to living and man-made sounds, we compared time-point by time-point whether the posterior probabilities $P(c_{i_h}|\boldsymbol{m}_{th})$ and $P(c_{j_h}|\boldsymbol{m}_{th})$ were different between conditions by means of non-parametric statistical analyses (e.g. Kruskal–Wallis test) in the training dataset (see Tzovara et al., 2012 for a similar approach based on Bayes Factor). This strategy could reveal one or more continuous time periods of differential activity between conditions. We call H the set of time-points over which the two conditions differed (Fig. 2b). This approach estimated discriminative time-periods in a data driven manner without *a priori* selection of the duration of these periods.

Testing phase

Discrimination function and accuracy. During the test phase, we computed the posterior probabilities $P(c_{i_h}|\boldsymbol{m}_{th})$ and $P(c_{j_h}|\boldsymbol{m}_{th})$ on the single trials belonging to the test dataset and along the temporal period H that had been estimated in the training phase (Fig. 2c). The discrimination function for the trial t was defined as:

$$DF_t = \Sigma_{h\in H}\left[\log P\left(c_{i_h}\middle|\boldsymbol{m}_{th}\right) - \log P\left(c_{j_h}\middle|\boldsymbol{m}_{th}\right)\right] \tag{6}$$

The idea was to assign a trial from the test dataset to Condition1 if the posterior probability of the most likely Gaussian for the first model (Eq. (4)) was higher than the posterior probability of the most likely Gaussian for the second model (Eq. (5)), and vice versa.

The classification performance was evaluated by the absolute classification accuracy (percentage of correctly classified single trials as living or man-made). We therefore obtained a set of values of classification accuracy evaluated on ten independent test datasets.

Optimizing the total number of Gaussians and classification performance. In order to find the optimum values for the total number of Gaussians, Q, we generated a set of models for each condition separately, with each of these two parameters ranging from three to nine. Because we had considered ten splits of the data, for each pair

Table 1
Response time in milliseconds in response to living and man-made sounds when subjects could correctly indicate the sounds categories and when they could not guess. Asterisks denote response times to correctly categorized sounds that were significantly faster than those to uncategorized ones (t-tests; $p<0.05$). The difference between these response times was always significant for all subjects and for living and man-made sounds.

Subject number	Response time			
	Living sounds		Man-made sounds	
	Correctly categorized	Uncategorized	Correctly categorized	Uncategorized
S1	518 ms*	811 ms	517 ms*	1070 ms
S2	559 ms*	1302 ms	577 ms*	1467 ms
S3	690 ms*	2775 ms	853 ms*	2576 ms
S4	793 ms*	1867 ms	1117 ms*	1538 ms
S5	660 ms*	1782 ms	786 ms*	1326 ms

Table 2
Average position index of living sounds (second and third columns for correctly categorized and uncategorized ones, respectively) and man-made sounds (fourth and fifth columns for correctly categorized and uncategorized ones, respectively) for each of the five subjects. Asterisks indicate those position indexes of correctly categorized sounds that were significantly different from the uncategorized ones in the same category.

Subject number	Position order			
	Living sounds		Man-made sounds	
	Correctly categorized	Uncategorized	Correctly categorized	Uncategorized
S1	337	294	330^*	246
S2	330	328	329	290
S3	349^*	242	327	290
S4	291	274	275	223
S5	325	326	331	325

of models we obtained ten values of classification accuracy. The best number of Gaussians was determined so as to maximize the mean classification accuracy across the ten datasets.

We report the mean classification accuracy obtained on the test datasets and on the validation dataset separately. The classification accuracy obtained on the validation dataset was compared to the chance classification performance obtained by randomly shuffling the labels 100 times in the training and tests datasets. For each shuffle of the labels, we re-trained the classifier for the previously selected parameters (i.e. number of maps for each of the two models) and we evaluated its performance on the validation dataset. We carried out a paired statistical comparison (Wilcoxon signed-rank test; $p < 0.001$) between performance of the true classifier and the 100 mean classification results obtained by randomly reshuffling the labels (see Pereira et al., 2009 for general guidelines on comparison between classifiers' performance and chance-level). This consists in a single comparison between the true classification performance and the 100 performances obtained on the permuted data. In practice, this test entails computing the difference between the values that the true classifier assigns to each validation trial (1 if correct, 0 otherwise), and those assigned by the 'random classifier'. The mean of these differences across trials in the validation dataset is tested with a Wilcoxon test to reject the null hypothesis that these 100 mean differences have median zero (see also the distribution of these differences across the 100 permutations in the Supplemental Material 3).

Testing the role of specific temporal period. The ten-fold training and testing analyses described above could reveal one or more temporal periods over which the two conditions significantly differ. In order to establish their contribution to the above-chance classification performance we ran a further analysis in the case when more than one continuous period of time points was found by the classification algorithm. For each of the latencies that we found significant in the previous analysis, we tested whether it was possible to decode semantic categories using the posterior probability computed along these periods of time only. As it will be clear in the following, decoding of semantic categories of correctly categorized sounds involved two separated time-periods. We therefore tested whether each of them could be informative enough for obtaining above-chance classification accuracy in the validation dataset.

Results

Behavioral results

Accuracy and response time

Subjects accurately categorized most of the sounds. The average percentage (±SEM is indicated for all the average percentages) of correctly categorized man-made sounds was 88 ± 1%; 5 ± 1% were misjudged as living and 7 ± 1% of the total number of the presented sounds were uncategorized. The average percentage of correctly categorized sounds of living items was 72 ± 3%, whereas miscategorized sounds was 15 ± 2% and uncategorized was 13 ± 2%.

Response times to uncategorized sounds were significantly longer than response times of correctly categorized ones (*t*-test; $t > 1.66$, $p < 0.05$) (Table 1).

Distribution of correctly categorized and uncategorized sounds throughout the experiment

We analyzed the possible influence of fatigue and/or a priming effect on behavior. We first tested whether miscategorization of sounds was due to a fatigue effect by controlling the presentation order of sounds that were correctly categorized with respect to uncategorized ones. Presentation order (indicated as PO in Tables 2 and 3) was analyzed by assigning an index (from 1 to the total number of presented sounds) to each sound along the experiment. We tested statistically whether this index of all the correctly categorized sounds included in the analysis was different from that of the uncategorized sounds. This test was performed for each category and each subject separately (this separation is justified by the same data partition in the single-trial analysis). We found that there was no statistical difference for most of the subjects and the semantic category (Table 2); two significant differences were found (first subject man-made category and third subject living category) always with the uncategorized sounds occurring on average earlier than the correctly categorized one. Therefore, we concluded that we did not have any evidence of a fatigue effect.

For all the subjects, we found a subset of sounds that were sometimes correctly categorized and sometimes uncategorized (indicated as overlapping sounds in Table 3, first and fourth columns). We investigated whether this behavioral result was interpretable in terms of a priming effect. Specifically we checked if uncategorized sounds were not priming those sounds that were, later in time, correctly categorized. Because each sound was presented many times during the course of the experiment, we computed the mean index (i.e. mean

Table 3
Number of living and man-made sounds to which subjects correctly or incorrectly answered depending on presentation order (first and fourth columns for living and man-made sounds, respectively). Average position order of these living sounds (second and third columns for correctly categorized and uncategorized ones, respectively) and man-made sounds (fifth and sixth columns for correctly categorized and uncategorized ones, respectively) for each of the five subjects. All position indexes of correctly categorized sounds were significantly later than the uncategorized ones in the same category.

Subject number	Living sounds			Man-made sounds		
	# overlapping sounds	Position order		# overlapping sounds	Position order	
		Correctly categorized	Uncategorized		Correctly categorized	Uncategorized
S1	13	82	18	13	82	8
S2	8	82	9	10	107	8
S3	9	77	13	12	86	7
S4	16	76	18	11	122	11
S5	14	61	23	8	101	8

presentation order) for each of the sounds when they were correctly categorized and when they were not. We therefore obtained a set of mean presentation order for each of the sounds when they were correctly categorized and when they were not (Table 3, second and third columns for living category, fifth and sixth columns for man-made category), considering here only those sounds to which subjects did not answer consistently. We finally tested statistically whether these mean values differed (paired *t*-test; $t>1.66$, $p<0.05$). We found that this difference was always significant, and we therefore concluded that it was not possible to exclude the presence of a priming effect.

Classification results

Uncategorized sounds

By training and testing across ten splits of the data, we evaluated the discriminative time interval between single trials in response to living and man-made sounds (that were all uncategorized) and the corresponding classification performance computed as the mean ratio of correctly classified single-trials (±SEM is indicated for all the average accuracies). The discriminative time interval between living and man-made single trials was found between 116 and 174 ms post-stimulus onset (thick line in Fig. 3ab) as observed in at least 8 out of ten training dataset (Fig. 4). This differential activity was observed in all the ten training datasets (Fig. 4), therefore demonstrating a persistent and robust effect across subjects.

The average accuracy across the ten test datasets was 0.62 (±0.05). On a validation dataset (one class of single-trials in response to sounds of living items) the classification accuracy was 0.68. This accuracy was significantly above chance-level (Wilcoxon signed-rank test; $z=-3.79$, $p<0.001$).

Fig. 3 shows the average posterior probability for the living and man-made datasets and the corresponding template maps (Figs. 3a, b). We show here only those posterior probabilities that were on average higher than all the others at least at one time point during the discriminative time period. The average posterior probabilities are highly structured in both cases during the post-stimulus period and in comparison to the baseline, with some template maps that remain 'active' (i.e. higher on average than the others) for up to 50 ms. This is the case of the blue-framed template maps around 100 ms and the red-framed template maps starting at around 200 ms.

The discriminative time interval was between 116 and 174 ms post-stimulus onset (thick line in Figs. 3ab and 4) involved mostly four template maps for the living condition and the red-framed map for the man-made condition (Figs. 3 a, b). A visual inspection of the voltage distributions in each of these template maps and the corresponding average posterior probabilities should clarify the role of this time-interval in discriminating between the two categories: the red-framed map of the man-made model shows a frontal positive polarity; in the same interval and for the living model the most represented template maps are characterized by different spatial distributions with the exception of the red-framed one (similar to the red-framed map of the man-made model). Because of the difference in the spatial distributions of most of these maps across models, the difference of the posterior probabilities obtained based by these two models leads to a robust classification between living and man-made sounds. From the same figure and following the same reasoning, we can observe that other periods are less informative for discriminating living and man-made objects. For example we can observe that starting at around 200 ms post-stimulus onset, both datasets are on average best represented by the red-framed map of both models. This time-period is not discriminative because the red-framed maps have a similar spatial distribution on the scalp.

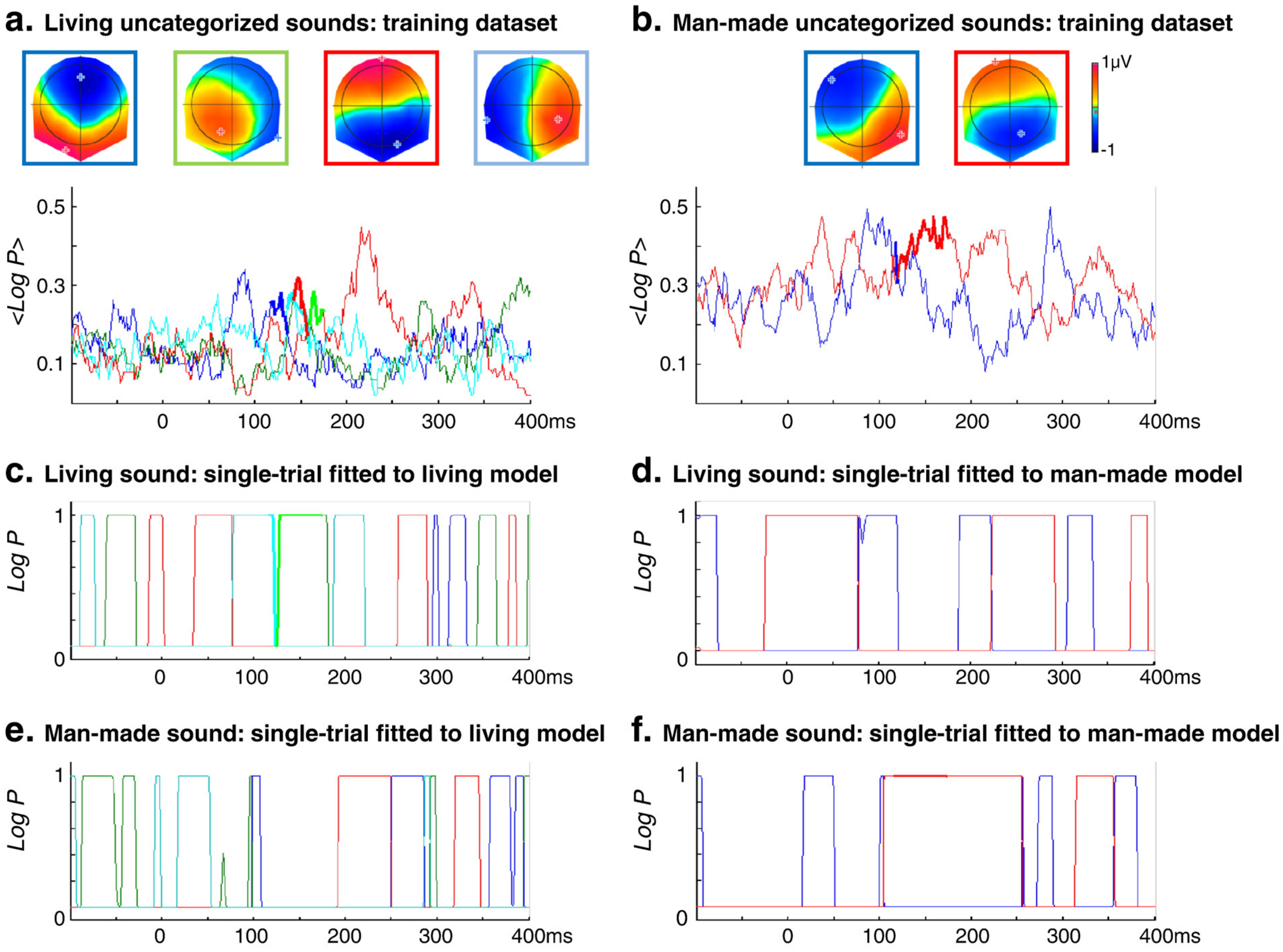
At the test level (Figs. 3c,d), the posterior probability of living and man-made single-trials, respectively, offer an intuitive representation of the classification result: a living sound is represented – during the discriminative period (thick line) – by both the green-framed and cyan template maps as it is evident when fitting the single-trial related to this sound to the living model (Fig. 3c); vice versa the same response when fitted to the man-made model is not providing the same amount of evidence that the red template map of the man-made model was actually active (Fig. 3d). An analogous observation can be derived for a man-made sound, that basically does not activate any of the template maps shown in panel a (Fig. 3e), while it shows a similar pattern of red-framed template maps as in panel b when fitted to the man-made model (Fig. 3f).

Correctly categorized sounds

As we did for the uncategorized sounds dataset, we considered all the single trials collapsed across subjects for each of the category living and man-made (±SEM is indicated for all the average accuracies). We obtained two intervals of significant difference between 112 and 185 ms post-stimulus onset, and between 270 and 342 ms (Fig. 5, time intervals evaluated in at least 8 out of the training datasets). Mean accuracy across the ten test datasets was 0.55 (±0.02). When classifying based on the two time periods separately, mean accuracy —across ten datasets was 0.54 (±0.03) and 0.58 (±0.03), respectively. On the validation dataset we obtained 0.56 when taking advantage of both intervals. This classification accuracy was significantly above chance-level (Wilcoxon signed-rank test; $z=-7.08$, $p<0.001$). We obtained 0.52 and 0.63 when considering separately the first and the second interval, respectively. These classification accuracies obtained on the validation dataset were significantly above chance level (Wilcoxon signed-rank test; $z=-7.84$, $p<0.001$ for the first time period; $z=-8.42$, $p<0.001$ for the second time period). These significant values are also reflected in the distributions shown in the Supplemental Material 3. For all these three accuracy estimations, the fraction of times the true classifier performed better that its permuted version was much higher than the opposite. In particular the classification performance (0.52), obtained by taking into account only the first time interval, was always higher than its permuted versions in all but three cases (see Supplemental Material 3, panel c), demonstrating the high significance of this – albeit relatively low – performance.

Discussion

We decoded semantic categories at a single-trial level by identifying modulations of the voltage topographies, previously normalized by the strength of their activity (Bernasconi et al., 2011; De Lucia et al., 2007, 2010b; Murray et al., 2009; Tzovara et al., 2012). Through multivariate decoding, we could establish whether specific differential activity in the EEG response to living and man-made sounds can be exploited to accurately decode a sound's semantic category. We carried out this analysis when subjects explicitly and correctly categorized sounds as living or as man-made, as well as when they could not provide an answer. We obtained two main findings. First, we showed that the brain can discriminate semantic categories (living and man-made) at early stages (116 ms to 174 ms post-stimulus onset) of sound processing even when it does not transpire behaviorally. Second, perceptual decision-making of environmental sounds typically relies on two stages of differential processing; during the first of these periods – overlapping with that observed for uncategorized sounds – enough evidence was accumulated for decoding sound categories from the EEG signal with an above-chance accuracy; the second stage of differential processing occurred after the sound ended. The first of these sets of results shows that even when subjects could not categorize the sounds, an above-chance decoding accuracy was obtained at early stages of sound processing. We interpret this evidence as the effect of a coarse level representation of the semantic category that is probably not sufficiently accurate for allowing a


a. Living uncategorized sounds: training dataset
b. Man-made uncategorized sounds: training dataset
1µV
-1
<Log P>
0.5
0.3
0.1
0
100
200
300
400ms
c. Living sound: single-trial fitted to living model
d. Living sound: single-trial fitted to man-made model
e. Man-made sound: single-trial fitted to living model
f. Man-made sound: single-trial fitted to man-made model
Log P
1
0

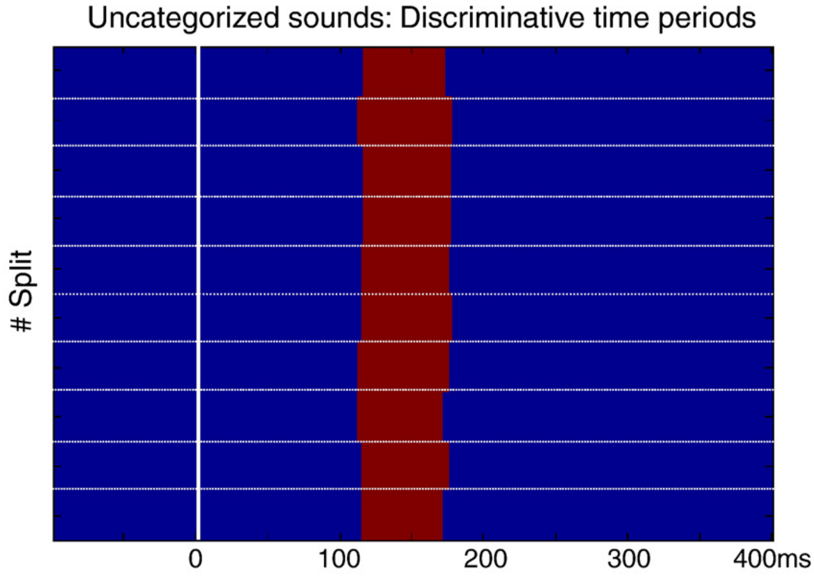


Fig. 4. Time-periods of differences between posterior probabilities assigned to trials of uncategorized living and man-made sounds (trials in response to sounds for all subjects). Red intervals highlight periods of differential activity between living and man-made trials as identified for each training dataset. Differential activity between 116 and 174 ms post-stimulus onset is consistently observed despite subjects' heterogeneity.

confident behavioral response during the categorization task. This result is particularly interesting in view of the fact that this early differential activity cannot readily be explained in terms of low-level acoustic differences between the groups of living and man-made sounds. Indeed comparisons between the spectrograms of the living and man-made sounds at the single subject level highlighted only sparse and minor time-frequency bins exhibiting statistically significant differences (Fig. 1). Moreover, because uncategorized sounds were different across subjects, possibly undetected low-level differences between living and man-made sounds (see Supplemental Material 2) would unlikely occur at the same latency for all the subjects. We conclude that decoding at this stage of sound processing is based on its semantic content (at least at a gross categorical level).

Aside from there being a certain degree of semantic representation that underlies the above-chance decoding accuracy of uncategorized sounds, it is worth noting that this result is based on time-locked differential activity between living and man-made sounds. One would expect that the dynamic nature of the presented stimuli would produce a certain jitter in the timing when accumulated sensory evidence reached a certain level for decoding the semantic content of the sound. If this were the case, our results indicate that for our sets of sounds the limit for the evidence accumulation time interval can jitter between 116 and 174 ms post-stimulus onset.

In the case of correctly categorized sounds, we observed two discriminative intervals. The first can be interpreted as playing the role of a first level of semantic categorization in a similar way as observed for uncategorized sounds. Indeed, the latencies of this first differential activity for categorized and uncategorized sounds were highly overlapping (cf. Figs. 4 and 5), despite the fact that these two analyses were run independently. When considering the first period of differential activity (112–185 ms post-stimulus onset) between living and man-made sounds, we obtained a relatively lower classification accuracy for correctly categorized sounds in comparison to uncategorized sounds. This effect can be explained by a higher degree of heterogeneity of the first of these two groups. Indeed, correctly categorized sounds included both sounds whose related object could be later identified and those for which subjects could only categorize the sound as man-made or as that of a living item. By contrast, the uncategorized sounds were forcibly also unidentified.

The analysis of single trials in response to correctly categorized sounds revealed in addition a discriminative time-period after the sound ended (270–342 ms post-stimulus onset). Because this second component is present only when subjects explicitly categorized the sound as living or man-made, it is likely representing the stage connecting the perceptual level to the process of executing a behavioral response, that is to say directly associated to the decision-making process. The relatively early latencies of this differential activity suggest that the present effects were not an artifact of motor planning and preparation, further supporting its role as a perceptual decision-making component. As for the uncategorized sounds, the comparison between the spectrograms of living and man-made sounds lead to minor statistically significant differences. These results support the conclusion that the above-chance classification accuracy in decoding single trials in response to living and man-made sounds cannot be explained in terms of acoustic differences but rather in relation to their semantic content.

It remains an open question whether it is possible to predict if a specific sound will be correctly categorized. In other words, it is presently unclear which features make a sound ultimately be correctly categorized. Ability to recognize a sound is certainly a function of the

Fig. 3. Panels a and b. Summary of the template maps and average posterior probabilities obtained when training the GMM model separately on the living and man-made datasets (panels a and b, respectively). During the discriminative time-period 116–174 ms, (see Classification results section) average posterior probabilities are plotted in thick lines, highlighting those values of probabilities that were used to classify test data as living or as man-made. Color codes between panels a and b should be read independently. (i.e. red-framed template map in a relates to red curve in the same panel only; the same holds for panel b). Panels c and d. Posterior probabilities of the single-trial response to the sound of a living object. Panels c and d show the posterior probabilities obtained using the living and the man-made model, respectively (color code in c and d refers to the template maps shown in a and b, respectively). The sum of posterior probabilities along the discriminative period is higher in panel c than in panel d, that is to say the living sound is correctly recognized by the classified. Panels e and f. Posterior probabilities of the single-trial response to the sound of a man-made object. Panels c and d show the posterior probabilities obtained using the living and the man-made model respectively (color code in e and f refers to the template maps shown in a and b, respectively). The sum of posterior probabilities along the discriminative period is higher in panel f than in panel e, that is to say the man-made sound is correctly classified.

accumulated evidence over time, i.e. their duration in time (Bigand et al.; 2011; see also Bestelmeyer et al., 2011). This is because all the presented sounds were easily recognized when presented for a longer duration (i.e. 500 ms). However the analysis of subjects' behavior during the EEG experiment has shown that subjects were also influenced by the number of times the same sound was repeated during the experiment, at least for part of the presented sounds (cf. Table 3 and Supplemental Material 2). Further investigations and experiments are required to disentangle effects of low-level features of the presented sounds and possible priming effects in relation to their impact on subjects' behavior (see Murray et al., 2008; De Lucia et al., 2010c for priming effects induced by environmental sounds repetition).

Behavioral analyses have also shown that we could exclude that the difference between correctly categorized and uncategorized sounds was due to an effect of fatigue. Indeed, we did not find evidence of a systematic deterioration of subjects' ability during the course of the experiment as would be expected if fatigue was influencing behavior.

Comparison to existing literature

Perceptual decision-making both in the visual and somatosensory domains has emphasized the existence of at least two stages in stimulus processing leading to stimulus discrimination; mainly a bottom up process followed by higher-level involvement of areas responsible for perceptual decision-making (Philiastides and Sajda, 2006; Romo et al., 2003; Salinas et al., 2000). Our results confirm the existence of at least two stages leading to perceptual decision across sensory modalities. In addition we show for the first time that discrimination at early stages of sound processing is not related to low-level features of auditory stimuli. Rather, semantic categorization can be decoded at these latencies and in the absence of behavioral proficiency.

Most of the studies focusing on perceptual decision-making are based on sets of stimuli that differ with respect to their low-level feature content. Consequently, it was not possible to evaluate whether early and time-locked differential activity was related to the specific physical characteristic of the presented stimuli or more generally to its related semantic category. This is the case of studies looking at perceptual decision-making in the visual domain. Philiastides et al. (Philiastides and Sajda, 2006; Philiastides et al., 2006) found that the N170 component could be used to discriminate face versus car stimuli, irrespective of the task difficulty induced by the noise level in the stimuli. Other common experimental paradigms in the visual domain, such as those involving the discrimination of motion direction also includes differences at the level of physical features of the presented stimuli (i.e. Bennur and Gold, 2011; Kim and Shadlen, 1999). Clearly, in both cases the advantage is that it is possible to control *a priori* the noise level, and looking at the dependency of neural activity on the degree of task difficulty. By contrast, in our study we could not determine which sounds were more difficult to categorize. Our distinction between categorized and uncategorized sounds is based on behavior and therefore evaluated *a posteriori*. Despite these differences, the latencies of our effects are similar to those found in the visual domain with complex stimuli (Philiastides et al., 2006); the first processing component occurring around 170 ms post-stimulus, the second component starting at 300 ms and shifting up to 450 ms post-stimulus onset depending on task difficulty.

Our results are not readily comparable with the limited extant literature on perceptual decision-making in the auditory domain. In previous studies subjects were asked to attribute the presented stimuli to one of two possible alternative phoneme 'categories', independent of the level of noise mask (Binder et al. 2004; Kaiser et al. 2006). The two alternative options were well separated in terms of their frequency content and therefore likely to be directly related to the quality of sensory information. In the current study, we consider environmental sounds that were grouped on the basis of behavioral outcome (correctly categorized or uncategorized) with different uncategorized sounds for each subject and each living/man-made category. Another important difference with respect to previous works is in the nature of the experimental task. In previous studies participants had to choose between two possible alternatives after a pair of sounds had been presented in sequence, therefore implicitly asking the subject to memorize the sequence before giving an answer. In the current study the categorization is based on subjects' past experience not on the direct memory of recent stimuli.

Multivariate decoding

Our multivariate decoding approach is based on the modulation of the presence of voltage topographies that were previously normalized.

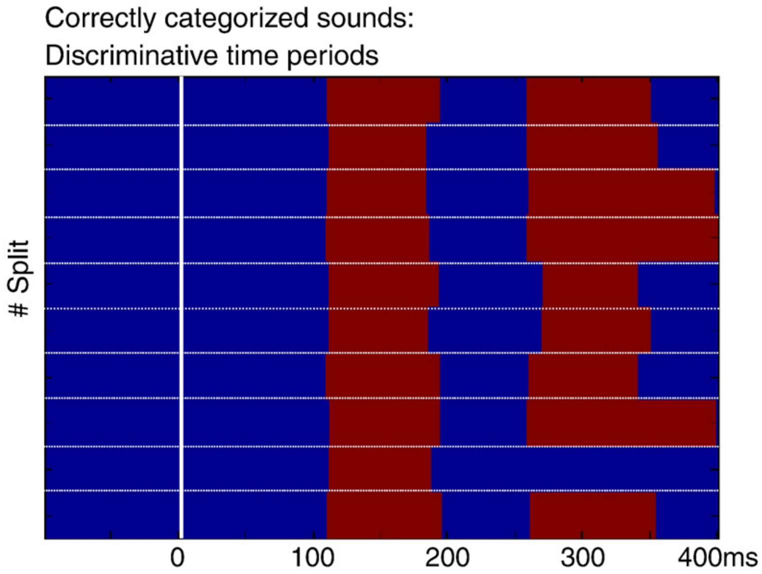


Fig. 5. Time-periods of differences between posterior probabilities assigned to trials of correctly categorized living and man-made sounds (trials in response to sounds for all subjects). Red intervals highlight periods of differential activity between living and man-made trials as identified for each training dataset. Differential activity between 112 and 185 ms post-stimulus onset, and between 270 and 342 ms post-stimulus onset is consistently observed despite subjects' heterogeneity.

This approach is therefore not biased by an *a priori* selection of the electrodes where one would expect to observe an effect and is informative about variations in the spatial configurations of the underlying neural sources. Indeed our multivariate approach is based on differences in the relative presence of specific voltage configurations that reflect specific spatial arrangements of neural generators. Other approaches based on the voltage measurement at multiple electrode sites do not consider the voltage topographies as multidimensional measurement, and therefore the information about the voltage distribution and its neurophysiologic interpretability is not taken into account (i.e. Philiastides et al., 2006). With respect to the values of the classification accuracy obtained, these are in line with what has been shown with environmental sounds and multivariate decoding in fMRI (Staeren et al. 2009). Using electrophysiologic signals, decoding sound categories at the single-trial AEP level has turned out to be more challenging than in the visual domain. Simanova et al. (Simanova et al., 2010) attempted at decoding semantic categories of auditory stimuli at the single-trial level in an experiment in which subjects were asked to perform a task irrelevant to the categorization. A multivariate pattern analysis based on Bayesian logistic regression performed poorly in comparison to a similar analysis applied to visual stimuli; and in about half of the subjects they obtained chance-level classification accuracy. Despite that, in a more challenging experimental paradigm, Bernasconi et al. (2011) could show that the same single-trial analysis as applied in the present study lead to an above-chance prediction of subjects auditory percepts while listening to acoustically identical auditory stimuli that were erroneously perceived as differing in pitch or duration.

Conclusions

We have shown that the brain is able to discriminate the semantic category of environmental sounds even when this does not transpire behaviorally. Two stages of differential activity over the initial 400 ms post-stimulus onset contribute to perceptual decision-making. An early stage of sound processing (at ~116 ms) is observed irrespective of whether or not subjects could explicitly categorize the sound as belonging to the man-made or to the living semantic category. Correct categorization requires a subsequent processing stage (at ~270 ms). To the best of our knowledge we provide for the first time an insight into brain mechanisms of auditory perceptual decision-making of complex environmental stimuli. Moreover, using normalized voltage topographies we could neurophysiologically interpret our classification results in terms of changes in the underlying neural generators activated in response to sounds belonging to different semantic categories.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2012.01.131.

Acknowledgments

The Swiss National Science Foundation provided financial support (grants #K-33K1_122518/1 to MDL and 310030B_133136 to MMM).

References

Bennur, S., Gold, J.I., 2011. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. J. Neurosci. 31 (3), 913–921.

Bernasconi, F., De Lucia, M., Tzovara, A., Manuel, A.L., Murray, M.M., Spierer, L., 2011. Noise in brain activity engenders perception and influences discrimination sensitivity. J. Neurosci. 31 (49), 17971–17981.

Bestelmeyer, P.E., Belin, P., Grosbras, M.H., 2011. Right temporal TMS impairs voice detection. Curr. Biol. 21, R838–R839.

Bigand, E., Delbé, C., Gérard, Y., Tillmann, B., 2011. Categorization of extremely brief auditory stimuli: domain-specific or domain-general processes? PLoS One 6, e27024.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Britten, K.H., Shadlen, M.N., Newsome, W.T., Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12, 4745–4765.

de Lafuente, V., Romo, R., 2005. Neuronal correlates of subjective sensory experience. Nat. Neurosci. 8, 1698–1703.

De Lucia, M., Michel, C.M., Clarke, S., Murray, M.M., 2007. Single-trial topographic analysis of human EEG: a new 'image' of event-related potentials. Proceedings Information Technology Applications in Biomedicine.

De Lucia, M., Camen, C., Clarke, S., Murray, M.M., 2009. The role of actions in auditory object discrimination. Neuroimage 48 (2), 475–485.

De Lucia, M., Clarke, S., Murray, M.M., 2010a. A temporal hierarchy for conspecific vocalization discrimination in humans. J. Neurosci. 30 (33), 11210–11221.

De Lucia, M., Michel, C.M., Murray, M.M., 2010b. Comparing ICA-based and single-trial topographic ERP analyses. Brain Topogr. 23 (2), 119–127.

De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R.A., Clarke, S., Murray, M.M., 2010c. Perceptual and semantic contributions to repetition priming of environmental sounds. Cereb. Cortex 20 (7), 1676–1684.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. R. Stat. Soc. B 39, 1–38.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.

Heekeren, H.R., Marrett, S., Ungerleider, L.G., 2008. The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci. 9 (6), 467–479.

Hernandez, A., Zainos, A., Romo, R., 2002. Temporal evolution of a decision-making process in medial premotor cortex. Neuron 33, 959–972.

Kaiser, J., Lennert, T., Lutzenberger, W., 2006. Dynamics of oscillatory activity during auditory decision making. Cereb. Cortex 17, 2258–2267.

Kim, J.N., Shadlen, M.N., 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. 2, 176–185.

Knebel, J.F., Toepel, U., Hudry, J., le Coutre, J., Murray, M.M., 2008. Generating controlled image sets in cognitive neuroscience research. Brain Topogr. 20, 284–289.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540 Review.

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., Lakany, H., 2011. EEG decoding of semantic category reveals distributed representations for single concepts. Brain Lang. 117 (1), 12–22.

Murray, M.M., Camen, C., Gonzalez Andino, S.L., Bovet, P., Clarke, S., 2006. Rapid brain discrimination of sounds of objects. J. Neurosci. 26 (4), 1293–1302.

Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20, 249–264.

Murray, M.M., De Lucia, M., Brunet, D., Michel, C.M., 2009. Principles of topographic analyses of electrical neuroimaging. In: Handy, T.C. (Ed.), Event-Related Potentials II: Advances in ERP, EEG, & MEG Analysis. MIT Press.

Newsome, W.T., Britten, K.H., Movshon, J.A., 1989. Neuronal correlates of a perceptual decision. Nature 341, 52–54.

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

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45 (1 Suppl), S199–S209.

Perrin, F., Pernier, J., Bertrand, O., Giard, M.H., Echallier, J.F., 1987. Mapping of scalp potentials by surface spline interpolation. Electroencephalogr. Clin. Neurophysiol. 66, 75–81.

Philiastides, M.G., Sajda, P., 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb. Cortex 16 (4), 509–518.

Philiastides, M.G., Ratcliff, R., Sajda, P., 2006. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J. Neurosci. 26 (35), 8965–8975.

Ratcliff, R., Philiastides, M.G., Sajda, P., 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc. Natl. Acad. Sci. U. S. A. 106 (16), 6539–6544.

Romo, R., Hernandez, A., Zainos, A., Salinas, E., 2003. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. Neuron 38, 649–657.

Salinas, E., Hernandez, A., Zainos, A., Romo, R., 2000. Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. J. Neurosci. 20, 5503–5515.

Shadlen, M.N., Newsome, W.T., 1996. Motion perception: seeing and deciding. Proc. Natl. Acad. Sci U. S. A. 93 (2), 628–633.

Shinkareva, S.V., Malave, V.L., Just, M.A., Mitchell, T.M., 2011. Exploring commonalities across participants in the neural representation of objects. Hum. Brain Mapp. doi:10.1002/hbm.21296.

Simanova, I., van Gerven, M., Oostenveld, R., Hagoort, P., 2010. Identifying object categories from event-related EEG: toward decoding of conceptual representations. PLoS One 5 (12), e14465.

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 (6), 498–502.

Tzovara, A., Murray, M.M., Plomp, G., Herzog, M.H., Michel, C.M., De Lucia, M., 2012. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recognition. 45 (6), 2109–2122.

Tzovara, A., Murray, M.M., Bourdaud, N., Chavarriaga, R., del R. Millán, J., De Lucia, M., in press. The timing of exploratory decision-making revealed by single-trial EEG analysis. Neuroimage. doi:10.1016/j.neuroimage.2012.01.136.