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-