



Processing of vocalizations in humans and monkeys: A comparative fMRI study

Olivier Joly ^{a,*}, Christophe Pallier ^{b,1}, Franck Ramus ^c, Daniel Pressnitzer ^d,
Wim Vanduffel ^{a,e,f}, Guy A. Orban ^a

^a Lab Neuro-en Psychofysiologie, K.U. Leuven, Medical School, Herestraat 49, B-3000, Leuven, Belgium

^b INSERM-CEA Cognitive Neuroimaging Unit, Neurospin center, Bat 145, F-91191 Gif-sur-Yvette, France

^c Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, EHESS, ENS, 29 rue d'Ulm, 75005 Paris, France

^d Département d'études cognitives, ENS, 29 rue d'Ulm, 75230 Paris cedex 05, France

^e Athinoula A. Martinos Center for Biomedical Imaging, 13th Street, Charlestown, MA 02129, USA

^f Harvard Medical School, Department of Radiology, Boston, MA 02114, USA

ARTICLE INFO

Article history:

Accepted 25 May 2012

Available online 31 May 2012

Keywords:

Audition

Macaque

Human

Monkey calls

Speech

Functional brain imaging

ABSTRACT

Humans and many other animals use acoustical signals to mediate social interactions with conspecifics. The evolution of sound-based communication is still poorly understood and its neural correlates have only recently begun to be investigated. In the present study, we applied functional MRI to humans and macaque monkeys listening to identical stimuli in order to compare the cortical networks involved in the processing of vocalizations. At the first stages of auditory processing, both species showed similar fMRI activity maps within and around the lateral sulcus (the Sylvian fissure in humans). Monkeys showed remarkably similar responses to monkey calls and to human vocal sounds (speech or otherwise), mainly in the lateral sulcus and the adjacent superior temporal gyrus (STG). In contrast, a preference for human vocalizations and especially for speech was observed in the human STG and superior temporal sulcus (STS). The STS and Broca's region were especially responsive to intelligible utterances. The evolution of the language faculty in humans appears to have recruited most of the STS. It may be that in monkeys, a much simpler repertoire of vocalizations requires less involvement of this temporal territory.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The evolutionary origins of human language remain largely mysterious. To mediate social interactions, many non-human species use vocalizations which might constitute precursors of human speech. These vocalizations can convey meaning (e.g. alarm calls in vervet monkeys; Seyfarth et al., 1980), as well as the identity and the emotional state of the speaker (Banse and Scherer, 1996; Ghazanfar et al., 2007). An important question is to what extent the processing of vocalizations in humans relies on mechanisms shared with our close relatives, non-human primates. Did the human auditory system become very different from that of other primates because of special demands of speech perception (Liberman and Mattingly, 1985)? The data available do not provide a clear-cut answer. Cross-species studies have shown that non-human animals can be as sensitive as humans to differences between human speech sounds (Brown and Sinnott,

2006). Like humans, monkeys spontaneously perceive changes in formant frequencies (Fitch and Fritz, 2006) and recognize other individuals by their voices (Ghazanfar et al., 2007). From such behavioral evidence, Hauser et al. (2002) argued that, with respect to speech perception, "The available data suggest a much stronger continuity between animals and humans than previously believed".

Monkeys' and humans' responses to vocalizations have already been explored in several single-unit recording and brain imaging experiments, although never within a single comparative study. Studies on monkeys have reported neurons selective for monkey calls in different regions such as the anterior lateral belt of the auditory cortex (Rauschecker et al., 1995), the ventrolateral prefrontal cortex (Romanski, 2004) and the insular cortex (Remedios et al., 2009; for a review see Romanski and Averbach, 2009). Positron Emission Tomography (PET) in monkeys (Gil-da-Costa et al., 2004, 2006) reported that conspecific vocalizations elicited greater activity than non-biological sounds in higher-order visual areas of the temporal lobe (TEO, TE and superior temporal sulcus, STS), as well as in the temporo-parietal area (Tpt) and the ventral premotor cortex, considered by these authors as homologous to Wernicke and Broca's areas, respectively. In contrast, a recent monkey fMRI study (Petkov et al., 2008) found no strong preference for species-specific vocalizations in these areas but instead reported a "voice" region in the anterior superior temporal plane. This region showed response suppression when several vocalizations from the same individuals were played, similar to a

* Corresponding author at: Institute of Neuroscience, Newcastle University Medical School, Newcastle upon Tyne, NE2 4HH, UK. Fax: +44 191 222 5227.

E-mail addresses: olivier.jjoly@gmail.com (O. Joly), christophe@pallier.org (C. Pallier), franck.ramus@ens.fr (F. Ramus), daniel.pressnitzer@ens.fr (D. Pressnitzer), wim.vanduffel@med.kuleuven.be (W. Vanduffel), guy.orban@med.kuleuven.be (G.A. Orban).

¹ The two first authors contributed equally.

previous finding in humans that reported adaptation to the speaker's voice in the right anterior superior temporal gyrus (STG, [Belin and Zatorre, 2003](#)).

In humans, listening to speech or to vocal non-speech sounds, compared to environmental or musical sounds ([Belin et al., 2000](#)) or to other animals' vocalizations ([Fecteau et al., 2004](#)) activates several bilateral regions along the STS. When compared to acoustic controls, responses to speech in passive listening conditions can be bilateral ([Hickok and Poeppel, 2004](#)) or left-dominant ([Narain et al., 2003](#)), while responses to non-linguistic vocal sounds (laughs, cries, moans, sighs) involve mainly the right anterior STS ([Belin et al., 2002](#); [Meyer et al., 2005](#)).

Here, we compared the neural substrates involved in processing vocalizations in rhesus monkeys and humans, using whole-brain functional magnetic resonance imaging (fMRI). Rhesus monkeys and humans were scanned while listening to monkey calls, human speech, human emotional non-speech vocalizations, bird songs (for humans only) and acoustic controls matched in spectral content (see [Fig. 1](#)). For humans there was an additional distinction in that speech stimuli could either be intelligible (mother language) or not (foreign language). The goal of the study was to compare cortical activations in humans and monkeys tested under experimental conditions, as similar as possible, in order to determine to what extent monkey cortical activations associated to processing of vocalizations resemble that of humans. There were two main questions of interest: first, would we observe species-specific responses, that is, regions responding more strongly to monkey calls than to human vocalizations in monkeys, and vice versa in humans? Second, would the pattern of areas activated by monkey calls in monkeys be similar to that involved in low-level speech processing (unintelligible speech), high-level speech processing (intelligible speech), or emotional vocalizations in humans?

In a previous report ([Joly et al., 2012](#)), the data acquired with monkeys were studied with a focus on detailed analyses of the acoustic properties of the signal associated with the activations observed. In the present paper, we perform a systematic comparison of activations obtained using very similar protocols in monkeys and in humans.

Materials and methods

Subjects

Monkeys

Three adult rhesus monkeys (*Macaca mulatta*), one female (M13) and two males (M14, M18), 5 to 6 years of age and weighing between

4 and 5 kg, participated in the experiment. These animals were born in captivity and had social experience limited to interaction with conspecifics in group housing and with humans during experiments. Before the scanning sessions, monkeys were trained daily to perform a visual fixation task with the head rigidly fixed to a primate chair. The fixation task was used to equalize attention across conditions and minimize body movement during scanning. The monkeys had little or no prior exposure to the French language, as this is not the primary language spoken in the laboratory. However, they were exposed daily to human voices in the animal facilities from both the radio and communication between monkey handlers. Details concerning head-post surgery and behavioral procedures are described in [Vanduffel et al. \(2001\)](#). Animal care and experimental procedures met the National and European guidelines and were approved by the local ethical committee.

Humans

Twenty right-handed native French speakers (9 men; 11 women; average age = 23.7 years (range, 20 to 28 years) with no history of neurological or psychiatric disease, participated in the experiment. None understood Arabic, the language of the stimuli in the unintelligible-speech condition. All participants gave written informed consent and were paid for their participation.

Stimuli

Five classes of sounds were used to construct the stimuli used in the experiment ([Fig. 1](#)): monkey vocalizations, human emotional (non-linguistic) vocalizations, intelligible speech, non-intelligible speech and bird songs. One hundred monkey vocalizations (Mvoc), uttered by several individuals of both sexes and drawn from the Rhesus Monkey Repertoire recorded in Cayo Santiago, Puerto Rico over a period of several years were provided by Marc Hauser. We selected recordings of five types of social calls which were described as having either positive (coos, girneys and harmonic arches) or negative valence (screams and shrill barks) ([Gouzoules et al., 1984](#); [Hauser and Marler, 1993](#)). Human vocalizations, including intelligible speech (French), unintelligible speech (Arabic) and emotional sounds (Hemo), were recorded from eight speakers (4 females and 4 males) while others were extracted from movie soundtracks. In an attempt to match the typical brevity of the monkey calls, speech utterances were very short sentences ("It is raining", "It is not possible"...) or interjections ("Hi!", "Excuse me!") averaging 1 s in duration. The human emotional (Hemo)

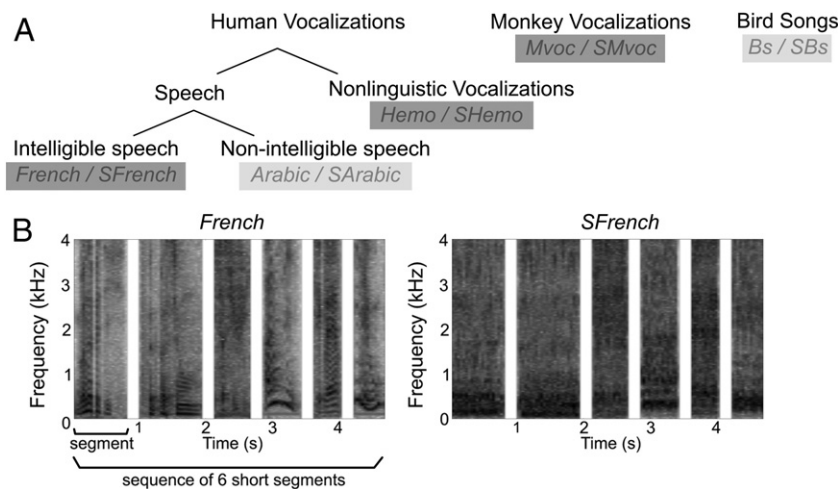


Fig. 1. Different categories of vocalizations. (A) The five sound categories are: intelligible speech, unintelligible speech, non-linguistic vocalizations (emotional sounds), monkey vocalizations and bird songs. Light gray represents conditions presented to human subjects only. For each category, the abbreviated condition label is indicated, the prefix S stands for the scrambled control. (B) Spectrograms of an example stimulus from the French category and the corresponding scrambled control (SFrench). A stimulus is defined as a sequence of short segments.

Download English Version:

<https://daneshyari.com/en/article/6030314>

Download Persian Version:

<https://daneshyari.com/article/6030314>

[Daneshyari.com](https://daneshyari.com)