Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

Research Report

Early neurophysiological correlates of vocal versus non-vocal sound processing in adults

Nicole Bruneau[a,b,*], Sylvie Roux[a,b], Helen Cléry[a,b], Ophélie Rogier[a,b], Aurélie Bidet-Caulet[d], Catherine Barthélémy[a,b,c]

[a]INSERM, U930, Imaging and Brain Tours, France
[b]Université François-Rabelais de Tours, France
[c]CHRU de Tours, France
[d]INSERM, U1028, CNRS, UMR5292, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon F-69500, France

ARTICLE INFO

Article history:
Accepted 7 June 2013
Available online 29 June 2013

Keywords:
Voice processing
Event-related potentials
Topographical analyses
Right fronto-temporal response to voice
Occipito–temporo–parietal response to voice

ABSTRACT

Electrophysiological correlates of voice processing were studied in twenty adults by comparing auditory evoked potentials in response to voice and environmental sounds in passive condition. Both categories of stimuli elicited similar cortical auditory responses (i.e. N1, P2, N2 peaks); however these peaks were overlapped by two components specifically elicited by voice. The first component was evidenced as a positive deflection recorded over the fronto-temporal sites, and lateralized on the right hemiscalp. This fronto-temporal positivity to voice (FTPV) may constitute the electrophysiological counterpart of the activation of the temporal voice areas previously described in neuroimaging studies. The second component was recorded at occipito-temporo-parietal sites. This occipito–temporo–parietal negativity to voice might correspond to visual mental imagery of the vocal sounds or to some form of mental simulation of the action sounds (e.g. coughing). Both components began as early as 70 ms post-stimulus onset indicating a rapid discrimination of voice in our auditory environment, which might be the basis of communication functions in humans.

1. Introduction

The voice is one of the most significant sources of information for communication. Cortical areas specifically tuned to its processing have been extensively described. Functional magnetic resonance imaging (fMRI) studies performed in adults have identified “voice-selective areas” i.e. brain structures that are more activated by vocal than by non-vocal stimuli. These areas are located bilaterally along the upper bank of the superior temporal sulcus (STS) but show greater sensitivity to voice on the right than on the left hemisphere (Belin et al., 2000, 2002, 2004 for review).

Abbreviations: Voc, Vocal; NVoc, Non-vocal; FTPV, Fronto-temporal positivity to voice; SCD, Scalp current density; SPL, Sound pressure level

*Corresponding author at: CHU Bretonneau, Child Psychiatry Centre, INSERM U930, 2 Boulevard Tonnellé, Cedex 9, 37044 Tours, France. Fax: +33 2 47 47 38 46.
E-mail address: nicole.bruneau@univ-tours.fr (N. Bruneau).

0006-8993/$ - see front matter
http://dx.doi.org/10.1016/j.brainres.2013.06.008

The aim of the current study was to further explore spatio-temporal processing of voices, given the contradictory results observed in electroencephalography (EEG) and magnetoencephalography (MEG) studies. A late Voice-Sensitive Response (VSR) was first described in response to sung voices compared to musical instruments sounds, culminating at around 320 ms after stimulus onset at fronto-central sites (Levy et al., 2001). However, this response was afterward demonstrated to be attention-dependent rather than voice specific (Levy et al., 2003; Gunji et al., 2003). Charest et al. (2009) compared event-related potentials (ERPs) responses to human vocalizations (speech and non-speech) with those to environmental sounds or bird songs in adults while they were performing a pure tone detection task. Significant differences between vocal and non-vocal responses were described as early as 120 ms after stimulus onset at bilateral fronto-temporal sites, and culminating at around 200 ms. This "fronto-temporal positivity to voice" (FTPV) has also been observed in 4- to 5-year-old children (Rogier et al., 2010) in passive condition, emerging significantly around 60 ms after stimulus onset, elicited by non-speech vocal stimuli. In a recent MEG study, Capilla et al. (2012) recorded responses to vocalizations involving both speech and non-speech stimuli and the difference with non-vocal response was described emerging as early as 147 ms and peaking in the 200–250 ms time range. Moreover this study confirmed that the differential brain response to human voice is observed across different sub-categories but also across different tasks as well as during passive listening.

Importantly, the specific response to vocal sounds has been located in the mid anterior part of the STS bilaterally and in the right planum temporale using both speech and non-speech vocal sounds and MEG (Capilla et al., 2012), and in the superior temporal sulcus/gyri with a right-hemisphere prominence using only non-speech vocal sounds and EEG (De Lucia et al., 2010). All of these studies have focused on the fronto-temporal activities in response to voice and how they are influenced by different stimulus subcategories or tasks. However results of previous ERP studies indicated, in addition to the FTPV, a component recorded over the occipito–parietal sites in response to vocalizations, occurring in the same latency range as the FTPV (Charest et al., 2009; De Lucia et al., 2010).

The aim of the present study is to further investigate the spatio–temporal dynamics of the diverse brain responses specifically elicited by non-speech vocal stimuli (Voc), in comparison to non-vocal environmental sounds (NVoc), during passive listening. Using scalp current density distributions, which have the properties of being reference-free and of showing sharper peaks than those of the potential distributions (Pernier et al., 1988), we aimed at better separating the overlapping responses to voice.

2. Results

Fig. 1 presents the grand-averaged ERPs to Voc and NVoc stimuli at a subset of illustrative electrodes. The waveforms for both types of stimuli display the three classical successive fronto-central N1, P2 and N2 deflections peaking at around 120, 200 and 400 ms, respectively. In contrast to responses recorded at midline and left fronto-temporal electrode sites, responses recorded at right fronto-temporal electrodes (F8 and FT4) and occipito–temporo–parietal electrodes (O1, P3, T5, M1) were clearly dissociated.

The differences between Voc and NVoc responses were statistically significant from 74 to 296 ms at F8 and from 104 to 290 ms at FT4 (gray shading in Fig. 1), and also on F4 and T4 electrodes for shorter periods (see Fig. 2a). This voice effect, isolated in the [Voc–NVoc] difference wave, appeared as a positive deflection at right fronto-temporal electrodes that overlapped the classical N1 and P2 auditory responses. As can be seen in Fig. 3, the potential distribution of the [Voc–NVoc] voice effect remained stable during the entire period presenting a significant effect (between 74 and 294 ms post-stimulus onset), independently of the successive N1 and P2 components.

Significant stimulus-related differences were also found at occipito–temporo–parietal–sites from 52 to 192 ms on the left hemisphere (O1:60–192 ms; P3: 84–174 ms; T5: 78–174 ms, M1: 52–176 ms) (see Fig. 2a and gray shading in Fig. 1). This voice effect appeared as a negative deflection that overlapped the classical N1 and P2 auditory responses. In order to clarify the components underlying the negative fields recorded at occipital sites, scalp potential and scalp current density mapping was performed in response to Voc, NVoc and to [Voc-NVoc] difference (Fig. 4) in the 74–184 ms time-window. These maps showed that the posterior negative field was specifically elicited by Voc stimuli and that it was underlain by a sink-source current pattern at occipito–temporo–parietal sites.

2.1. *Sub-category effect*

ERPs to Voc sounds were further compared to ERPs to two homogeneous sub-categories of NVoc sounds (music sounds and warning sounds) extracted from the sequence of non-vocal environmental sounds. The waveforms obtained are shown in Fig. 5. Similar pattern of amplitude differences were observed when Voc sounds were compared to these two sub-categories and statistical results are shown on Fig. 2b and c. The right fronto-temporal voice effect emerges at around 75 ms or 110 ms when comparing Voc ERPs with ERPs to warning sounds or with ERPs to music sounds, respectively. The occipito–temporo–parietal voice effect emerges at around 80 ms when comparing Voc ERPs with warning or music responses, the voice effect being bilateral when compared to warning responses. No significant difference was found when ERPs to music and warning sounds were compared.

3. Discussion

The results of this study indicate that, compared to environmental sounds, vocal sounds elicit a specific electrophysiological pattern of responses even if no auditory task is required. Both stimuli elicited the successive N1, P2, N2 sensory responses, and these responses were overlapped by two components specifically elicited by vocal sounds, one recorded at right fronto-temporal sites from 75 to 300 ms and the other recorded at occipito–temporo–parietal sites from 60 to 190 ms.