Brain and Cognition 92 (2014) 112–117

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Who is speaking? Implicit and explicit self and other voice recognition

Michela Candini [a,1], Elisa Zamagni [a,1], Angela Nuzzo [a], Francesco Ruotolo [b], Tina Iachini [b], Francesca Frassinetti [a,c,*]

[a] Department of Psychology, University of Bologna, 40127 Bologna, Italy
[b] Department of Psychology, Laboratory of Cognitive Science and Immersive Virtual Reality, Second University of Naples, 81100 Caserta, Italy
[c] Fondazione Salvatore Maugeri, Clinica del Lavoro e della Riabilitazione, IRCCS – Istituto Scientifico di Castel Goffredo, 46042 Mantova, Italy

ARTICLE INFO

Article history:
Accepted 1 October 2014
Available online 10 November 2014

Keywords:
Self
Other
Voice
Implicit recognition
Explicit recognition

ABSTRACT

In the domain of self-recognition, voice is a critical feature for self/other distinction. The aim of this study was to explore if people have an implicit and/or explicit knowledge of their voice. A group of healthy participants were submitted to an implicit and an explicit self-voice recognition task. They listened to pairs of pre-recorded auditory stimuli (words or pseudowords) pronounced by themselves, by a familiar or an unfamiliar person. Afterwards, in the "Implicit task" participants had to judge whether the pair of stimuli were pronounced by same or different speakers; in the "Explicit task" they had to identify if one of the stimuli was or not their own voice.

Results showed a difference between Implicit and Explicit tasks since participants were more accurate in implicit than explicit self voice-recognition. Moreover, in the Implicit task, participants had the same level of accuracy when they had to judge stimuli pronounced with self or others' voice, whereas when an explicit voice-recognition was required, they were less accurate with self than with others' voice.

1. Introduction

Recognition of people is a fundamental biological function for human species. To support this function, in our daily life we usually use a complex multimodal system based on multisensory (e.g. visual, auditory, etc.) recognition channels. Specifically, the voice signals carry different types of information, for example when we hear a voice we are informed about both linguistic information and speaker's personal characteristics such as gender, age and identity. Regarding the latter one, the importance of voices as a "channel" to convey the identity of the speaker is clear when speaker's visual characteristics are not available, for example, when we answer the phone or when we listen to someone speaking from an adjacent room. In these kinds of situations at least two different cognitive abilities are involved: the processing of voice's characteristics and the recognition of a familiar voice. The first process allows us to distinguish between vocal and non-vocal sounds (e.g. words with respect to natural sounds or animal cries), whereas the second process allows us to recognize a familiar voice among the others.

As far as the processing of voice's characteristics, fMRI studies have demonstrated the existence of "temporal voice areas" (TVAs), located bilaterally along the superior temporal gyrus, more involved in processing vocal than non-vocal sounds (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009; Gervais et al., 2004; Linden et al., 2011). Recently, Bestelmeyer, Belin, and Grosbras (2011) showed that voice/non voice discrimination ability was impaired when rTMS was applied over the right upper bank of the superior temporal sulci (STS) compared with a control site stimulation that was the right supramarginal gyrus (SMG). Moreover, this effect was not observed in a control task in which participants had to judge the loudness of different sounds. In this latter task, performance was not affected by rTMS at either stimulation sites. Other studies have demonstrated that different regions of the superior temporal sulcus (STS) are involved in different aspects of voice processing (Andics et al., 2010; Blank, Anwander, & von Kriegstein, 2011; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; von Kriegstein & Giraud, 2004). von Kriegstein et al. (2003) conducted an fMRI study in which subjects had to recognize a target speaker or a verbal content in sentences spoken by familiar and unfamiliar speakers. Results put in evidence that posterior regions of the STS were more involved in the verbal content task and anterior regions were responsive to voice recognition.

* Corresponding author at: Department of Psychology, University of Bologna, Viale Berti Pichat, 5 – 40127 Bologna, Italy. Fax: +39 051 243086.
E-mail address: francesc.frassinetti@unibo.it (F. Frassinetti).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.bandc.2014.10.001
0278-2626/

Regarding voice recognition, studies using different approaches have shown that voice recognition is dissociated from other aspects of identity such as, for example, face recognition (see Gainotti, 2011, 2013 for reviews). Neuropsychological evidence shows that the right fusiform gyrus is prevalently involved in face recognition (Gauthier et al., 2000; Kanwisher, McDermott, & Chun, 1997) whereas the right superior temporal gyrus is mainly involved in voice recognition (Belin, 2006; Belin et al., 2000).

Moreover, famous voice recognition can be dissociated from unfamiliar voice discrimination. Studies on brain damaged patients have shown that patients with a right posterior peri-sylvian lesion were impaired in famous voice recognition, whereas patients with a bilateral temporal lobe lesion were impaired in unfamiliar voice discrimination (Van Lancker, Cummings, Kreiman, & Dobkin, 1988; Van Lancker, Kreiman, & Cummings, 1989). Accordingly, neuroimaging data support the notion that familiar and unfamiliar voices are processed in different areas showing that the anterior part of the right superior temporal sulcus (STS) is more activated for familiar than for unfamiliar voices (Belin & Zatorre, 2003; von Kriegstein & Giraud, 2004; von Kriegstein et al., 2003). In a PET study (Nakamura et al., 2001) using a familiar/unfamiliar decision task with friends' voices or unknown persons' voices, higher activities in the right temporal and left frontal pole, in the entorhinal cortex and in the left precuneus were found during the recognition of familiar voices.

Whereas the majority of studies provides evidence about familiar/unfamiliar voices discrimination, few neuroimaging (Allen et al., 2005; Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008; Nakamura et al., 2001; Rosa, Lassonde, Pinard, Keenan, & Belin, 2008) and neurophysiological studies (Graux et al., 2013) have investigated one's own voice recognition. These studies reported the involvement of different brain areas for self as compared to other's voice recognition.

Allen et al. (2005), in an fMRI study, found that listening to self-generated words was associated with more activation in the left inferior frontal and right anterior cingulate cortex, whereas listening to other people's generated words was associated with greater engagement of the lateral temporal cortex bilaterally. Analogously, a difference in the mismatch negativity (MMN) was found in an Event-Related-Potentials (ERP) study (Graux et al., 2013) when participants passively heard three recordings pronounced by themselves with respect to when they were pronounced by two unknown persons.

Overall, these previous findings suggest that self, familiar and unfamiliar voices are processed as distinct information and are subtended by different cerebral areas.

However, it is worth noting that all the mentioned studies with the only exception of the ERP study by Graux et al. (2013) used a paradigm of explicit self-voice recognition.

This is particularly relevant since in the domain of self-recognition there is much evidence suggesting a dissociation between implicit and explicit self-processing. In this respect, it was recently demonstrated an intriguing dissociation between implicit and explicit self-body recognition. The interesting point is that participants, who implicitly recognized images representing self body-parts in a visual matching task, failed when an explicit recognition of self-body images was required. Indeed, when participants had to indicate which of two vertically aligned images (high or low) matched a central target stimulus, they performed better with self rather than other' body-parts. By contrast, a lack of this facilitation was observed when participants were required to explicitly judge if the upper or the lower image corresponded to their own body-effector (Ferri, Frassinetti, Costantini, & Gallese, 2011; Frassinetti, Ferri, Maini, Benassi, & Gallese, 2011).

In line with this evidence, the authors proposed the existence of two-way access to our self-body knowledge which involves different mechanisms. Indeed, a sensorimotor body-representation is engaged in the implicit, but not in the explicit, recognition of one's own body-parts (Ferri, Frassinetti, Ardizzi, Costantini, & Gallese, 2012). Thus, the aim of the present study was to verify whether also for the voice there appeared a dissociation between implicit and explicit self-voice recognition.

To this aim, a group of healthy subjects was submitted to an implicit and an explicit self-voice recognition task. In both tasks, participants listened to pairs of auditory stimuli (words or pseudo-words). For both types of stimuli, speaker's voice could belong to: the participant, a familiar or an unfamiliar person. In the Implicit task, participants had to judge whether stimuli were pronounced by the same speaker or different speakers, whereas in the Explicit task participants had to identify if there was, or not, their own voice.

In line with previous evidence, we postulate that the implicit recognition of one's own voice, similarly to implicit self-body recognition, relies upon a sensory network, whereas the explicit recognition of one's own voice is based on cognitive mechanisms. For this reason, the explicit processing could be more fallacious than the implicit one (Berti & Rizzolatti, 1992; Tranel & Damasio, 1985). Thus, we hypothesize a facilitation for one's own voice in the implicit but not in the Explicit task. This would prove that implicit and explicit self-voice processing is subtended by different mechanisms. Alternatively, if implicit and explicit self-voice processing is based on the same mechanisms, no difference should emerge in the two tasks.

2. Material and methods

2.1. Participants

Sixty-two right-handed healthy participants (21 males, mean age = 43.2 ± 15.1 years and 41 females, mean age = 37.6 ± 14.7 years; Mann–Whitney *U* test $p = .10$) without auditory or neurological pathology participated in the study. All participants were Italian speakers and were naive to the purpose of the research. Written informed consent was obtained from all participants.

The study was approved by the local ethics committee and all procedures were in agreement with the 2008 Helsinki Declaration.

2.2. Stimuli

In a first session, voices were recorded by the same experimenter in a silent and quite room by using a recorder (Panasonic RR-XS420) positioned at a fixed distance (60 cm) from participant's trunk. Participants' voice was recorded while they pronounced Italian words and pseudowords presented on a sheet of paper (A4 format). They were invited to maintain a flat tone of voice and to pronounce items as clearly as possible. If the experimenter judged the recorded items not easily discriminable, asked participants to repeat them until they were. Words were disyllabic and high-frequency stimuli (in Italian: cane, foca, lupo, alce, rana e topo; in English: dog, seal, wolf, elk, frog, mouse) belonging to the same semantic category (animal). Pseudowords were obtained from the words by means of two letters replacement (cona, faco, lusa, leca, tupi e rona). Subsequently, each vocal stimulus was digitized at 44,100 Hz, 16 bit, stereo modality, and elaborated using a dedicated software (WavePad Sound Editor) to adjust overall sound pressure and to balance the volume. The mean duration of each stimulus was 663.18 ms (SD = 100.36; range = 451–894 ms). Each stimulus could represent participant's voice (A stimulus), the voice of a familiar other (B stimulus) or the voice of an unfamiliar other (C stimulus). Each participant was asked to bring