Is my voice just a familiar voice? An electrophysiological study

Jérôme Graux, Marie Gomot, Sylvie Roux, Frédérique Bonnet-Brilhault, and Nicole Bruneau

INTRODUCTION

We are constantly exposed to hearing our own voices (OVs) from the moment we speak and to our own faces as often as we look in the mirror. Some authors therefore consider such over-learned stimuli as highly familiar. However, there is mounting evidence suggesting that self-perception is ‘special’, i.e. it involves systems that are physically and functionally distinct from those involved in the perception of familiar and even unfamiliar stimuli. In this field own face is the self-stimulus that has received most of the attention. Own face perception involves a specific neural network that is different from that involved in perception of familiar and unfamiliar faces, as demonstrated in many imaging (Turk et al., 2002; Platek et al., 2004, 2006; Sugiura et al., 2005; Platek and Kemp, 2009; Taylor et al., 2009) and electrophysiological studies (Caharel et al., 2002; Sui et al., 2006; Keyes et al., 2010). Two imaging studies in the auditory modality (PET and functional magnetic resonance imaging) have suggested that processing of one’s OV and FVs involves different neural networks (Nakamura et al., 2001; Kaplan et al., 2008). However, the dynamics of these differential processes has not been examined using an electrophysiological approach which provides the temporal resolution to track the sequence of brain events.

In a recent electrophysiological study, we used an auditory oddball paradigm to compare the brain processes involved in OV discrimination compared with discrimination of unfamiliar voices (Graux et al., 2013). A recording of one’s OV can be used as stimuli in such oddball paradigm. Indeed, while hearing a recording of our OV is distinct from hearing the sound of our own natural voice while speaking because of bone conduction difference, a recording of our OV can be accurately recognized and discriminated from other voices [see Graux et al. (2013) for more information on this topic]. Although the mismatch negativity (MMN) evoked by the OV of the participant and by the unknown (unfamiliar) voice displayed similar features, the subsequent P3a clearly distinguished the two conditions as its amplitude was significantly smaller for OV than for unfamiliar voice discrimination. Moreover, OV automatic detection was associated with an early response (‘pre-MMN’) involving a component recorded over the left frontal region, the activity of which lasted throughout the whole time course of the discriminative response. This pattern was proposed as an index of self-processing (which must be differentiated from auditory sensory attenuation to self-produced speech underpinned by the efference copy mechanism). However, it might equally reflect familiarity processes as one’s OV is a highly familiar stimulus. Interestingly, using a similar oddball paradigm, Beauchemin et al. (2006) showed that features of MMN and P3a were different for FV and unfamiliar voice discrimination, with greater MMN and P3a amplitude to FVs compared with unfamiliar voices. In the study presented here we assessed the discriminative processes involved in automatic detection of OV compared with FVs under the hypothesis that distinct brain processes underlie OV and FV discrimination processes. Moreover, our protocol also allowed us to distinguish between both familiar/unfamiliar and own/unfamiliar voice discrimination processes.

MATERIALS AND METHODS

Participants

Fifteen participants (eight male and seven female) aged from 21 to 28 years (mean ± s.e.m. = 23 ± 2) participated in the experiment. All participants were native French speakers and had normal hearing as assessed by a subjective audiometric test performed prior to event-related potential (ERP) recordings. Subjects were not included if they had a history of neurological disorders, alcohol or substance dependence or mental retardation. The Ethics Committee (CCP) of the University Hospital of Tours approved the protocol. Each participant gave informed written consent after the experiment had been explained. The study was conducted according to the ethical principles of the World Medical Association (Declaration of Helsinki; BMJ 1991; 302: 1194).

Stimuli

A pool of voices was constituted by recording 100 female and male adults. All voices were recorded with an Apex® 180 microphone by asking the subjects to sustain (about 1 s) the French sequence ‘a’, ‘e’, ‘i’, ‘o’, ‘u’ with neutral prosody. Only the vowels ‘a’ were used for the ERP study. Voice recordings were edited using Adobe Audition® 2.0.
EEG data were recorded from 64 electrodes using the ActiveTwo EEG recording head was 75 dB SPL. They were not aware that they were going to hear their OV or a FV in a movie on a TV screen and they were instructed to pay attention to the sequences that they would hear a series of different voices. They watched a silent armchair located in a soundproofed dimly lit room and were informed the sequences was counterbalanced across the participants. Each sequence comprised a block of 1000 stimuli including a standard (probability of occurrence: $P = 0.86$) and two deviant voices ($P = 0.07$ each). Stimuli were delivered through speakers with a constant (onset to onset) interstimulus interval of 700 ms. A sequence lasted 12 min and the whole paradigm around 40 min. Analyses of sound power in the spectral and time domains were performed using one-way ANOVA at each frequency, time frequency and time point (Figure 1A–C, respectively) using Matlab. No significant power differences were found between the three voices (OV, FV and UV) in either spectral or time domains. In the time frequency, only significant effects of voice category were observed around the 1000 and 8000 Hz frequency bands.

Procedure
At the beginning of the session, subjects were asked to participate in the recording of their OVs to constitute a ‘library’ of voices, and this was performed according to the previously described method. They subsequently performed the audiometric test followed by the ERP session. During ERP recording, subjects were seated in a comfortable armchair located in a soundproofed dimly lit room and were informed that they would hear a series of different voices. They watched a silent movie on a TV screen and they were instructed to pay attention to the video to tell the story to the experimenter at the end of the session. They were not aware that they were going to hear their OV or a FV in the auditory sequences. Overall, intensity of sounds at the subject’s head was 75 dB SPL.

EEG recording
EEG data were recorded from 64 electrodes using the ActiveTwo system (BioSemi, The Netherlands). Additional electrodes were applied on the nose for offline analysis and below the right eye to record the electro-oculographic activity. The signal was recorded with a sampling frequency of 512 Hz, filtered at 0–104 Hz and stored for offline analysis. Data were referenced offline to the potential recorded on the nose.

Data analysis
EEG data were edited with NeuroScan® equipment (Compumedics) after data conversion using Polyrex. A 0.3 Hz digital high-pass filter was applied. Movement artifacts were discarded manually, and automatic correction of eye movements was applied using a spatial filter transform developed by Neuroscan® (Neuroscan, 2003). EEG epochs were averaged for each stimulus condition over a 700 ms analysis period, including 100 ms before the stimulus. ERP were corrected with a $-100$ to 0 ms baseline before stimulus onset and were digitally filtered (low-pass 30 Hz). They were then analyzed with the ELAN® software (Aguera et al., 2011).

In the first step, we compared the responses to the three types of standard stimuli (OV, FV and UV) at each electrode and at each time point using permutation tests implemented in ELAN® software. To restrict the period of analysis of the ‘standard’ ERPs, we considered two windows of interest according to the morphology of the ERPs to voice, i.e. the early positivity between 0 and 200 ms after stimulus onset and the later slow negativity occurring in the 200–450 ms latency range. To take multiple comparisons into account in the time interval of interest, and thus to assess the actual significance of time intervals of apparently significant differences, we used a randomization procedure proposed by Blair and Karniski (1993) and inspired by Guthrie and Buchwald (1991).

Moreover, the brain activity specifically elicited by automatic discrimination of the participant’s OV was obtained by subtracting ERPs to OV stimuli when presented as standard (sequence 2) from ERPs to OV stimuli when presented as deviant (sequence 1). This response is termed ‘OV discriminative response’ (OVDR). Likewise, the ‘FV discriminative response’ (FVDR) was obtained by subtracting ERPs to FVs when presented as standard (sequence 3) from ERPs to FV stimuli when presented as deviant (sequence 1).

The ‘unfamiliar voice discriminative response’ (UVDR) was also computed by subtracting ERPs to unfamiliar voices when presented as standard (sequence 1) from ERPs to UV stimuli when presented as deviant (in the sequence where FV was the standard).

OVDR, FVDR and UVDR amplitude were compared using permutation tests with correction for multiple tests in two windows of interest: the latency window of MMN between 100 and 210 ms, and the latency window of P3a between 210 and 400 ms. OVDR, FVDR and UVDR latency measured at fronto-central sites (FCz) were compared using Student’s t-test.

Scalp potential (SP) maps were generated using a two-dimensional spherical spline interpolation (Perrin et al., 1989), and a radial projection from Cz (top views), which respects the length of the meridian arcs. To disentangle the distribution of multiple components overlapping in potential fields, scalp current density (SCD) maps were computed from the second spatial derivatives of the spline functions used for potential mapping (Perrin et al., 1987, 1989) obtained using Tikhonov regularization.

Topographical comparisons were performed on normalized data to avoid any bias from the amplitude effect. The normalization method involved dividing the value at each electrode by the norm of the vector in the electrode space for each subject and each condition (McCarthy and Wood, 1985). Topographical comparisons between OVDR, FVDR and UVDR were performed in the same latency range as previously described using permutation tests with correction for multiple tests.

RESULTS
ERPs to standard and deviant stimuli (OV, FV and UV) displayed the ‘classical’ successive deflections: P1, N1, P2 peaking at around 70, 115, 160 ms, respectively. These responses were followed by a large negative deflection peaking at around 300 ms and a late sustained negative activity. No significant differences were found between the amplitude of the three standard responses (Figure 2, left).

Figure 2 (right) shows the superimposition of the grand average to OVDR, FVDR and UVDR. In the three conditions, the grand-averaged fronto-central waveforms displayed a negative deflection peaking at around 170–190 ms, with polarity reversal at temporo-mastoid sites.
followed by a positive deflection peaking at around 260–280 ms. These responses corresponded to MMN and P3a deflections, respectively.

**OV<sup>DR</sup> and FV<sup>DR</sup>**

There were no significant differences between OV<sup>DR</sup> and FV<sup>DR</sup> in either MMN peak latency [OV<sup>DR</sup>: 180 ± 18 ms; FV<sup>DR</sup>: 178 ± 18 ms; t(14) = 0.5, P > 0.6] or P3a peak latency [OV<sup>DR</sup>: 267 ± 19 ms; FV<sup>DR</sup>: 269 ± 20 ms; t(14) = 0.4, P > 0.7] measured at FCz where MMN and P3a culminated. ERP differences between OV<sup>DR</sup> and FV<sup>DR</sup> were evaluated on normalized data by performing permutation tests with correction for multiple comparisons. Significant differences were found between OV<sup>DR</sup> and FV<sup>DR</sup> from 230 to 320 ms, i.e. in the latency window of P3a, at left fronto-temporal electrodes (F5, F3, FC5, FC3, T7, C5, C3, CP3) with greater positive fields to FV<sup>DR</sup> compared with OV<sup>DR</sup> (Figures 2 and 3). Comparison of P3a distributions recorded in the two conditions was performed using two-way ANOVAs on normalized mean amplitudes at three different left-frontal (F5, F3, FC5, FC3, C5, C3) medio-frontal (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) right-frontal (F4, F6, FC4, FC6, C4, C6) regions. The results indicated significant condition differences [F(1,14) = 8.7, P = 0.01] but no significant Condition × Electrodes set interaction [F(2,28) = 1.6, P > 0.2], indicating no topographical difference between the two conditions.

Descriptive analysis of corresponding SCD in the P3a latency range indicated left and right fronto-temporal and central source currents in the two conditions (Figure 3). On the left side, source currents displayed greater amplitude in FV<sup>DR</sup> compared with OV<sup>DR</sup>.

**OV<sup>DR</sup> and UV<sup>DR</sup>**

There were no significant differences between OV<sup>DR</sup> and UV<sup>DR</sup> in either MMN or P3a peak latency measured at FCz. Significant amplitude differences were found between OV<sup>DR</sup> and UV<sup>DR</sup> from 200 to 270 ms in the latency window of P3a, at right (AF4, F2, F4, F6, FC2, FC4, FC6, C2, C4, C6) and left (C1, C3) fronto-central electrodes with greater positive fields to UV<sup>DR</sup> compared with OV<sup>DR</sup> (Figures 2 and 3).

**FV<sup>DR</sup> and UV<sup>DR</sup>**

There were no significant differences between FV<sup>DR</sup> and UV<sup>DR</sup> in either MMN or P3a peak latency measured at FCz. Significant amplitude differences were found between FV<sup>DR</sup> and UV<sup>DR</sup> from 180 to 210 ms, i.e. in the latency window of MMN, at midline frontal electrodes (AF4, Fz, FCz, AF3, F1, F3) with greater negative fields to FV<sup>DR</sup> compared with UV<sup>DR</sup>.

**DISCUSSION**

Our results showed that the neural processes involved in the discrimination of one’s OV are clearly distinguishable from those involved in the discrimination of a FV as the P3a amplitude was significantly smaller in response to OV than to a FV. These results complement those obtained in a previous study in which we found that the amplitude of the P3a evoked by OV was smaller than that evoked by an unfamiliar voice (Graux et al., 2013). This confirms that the processes involved in perception of self-related stimuli are at least partially different from those involved in the perception of familiar and
non-familiar stimuli, as was previously found for face processing (Caharel et al., 2002; Turk et al., 2002; Platek et al., 2004, 2006; Sui et al., 2006; Platek and Kemp, 2009; Taylor et al., 2009; Keyes et al., 2010). However, it is possible that the term ‘familiarity’ encompasses different types of information involving different brain processes. For instance, it seems logical to think that a famous voice, the voice of a friend, a voice learned for a test or the voice of a sibling do not engage the same neural processes. Further studies are, therefore, needed to determine whether the processes involved in the perception of OV are radically different from those involved in all FVs regardless of the type of familiarity to exclude the possibility that the processes involved in the discrimination of one’s OV may be similar to some types of FVs.

From a topographical point of view, SCD maps obtained in the P3a latency range showed that the discriminative response to the FV and that to OV involved the same brain regions. However, the discriminative response to the FV evoked a response of greater amplitude in the left fronto-temporal region than that to OV. Imaging studies that have investigated the neuroanatomical correlates of perception of FVs have mainly highlighted activation of a neural network involved in episodic memory (Nakamura et al., 2001; Shah et al., 2001). However, some authors have stressed the need to distinguish the familiarity judgment of a voice (is that voice familiar or not?) and the identification of a FV which involves the recall of contextual and biographical information on the recognized person (Yonelinas, 2002). Both types of processing may recruit different neural processes, the former being thought to be underpinned by faster and more robust processes. Consistent with this, Birkett et al. (2007) showed that the region of the left auditory associative cortex (left middle temporal gyrus) was preferentially activated in tasks involving judgment of voice familiarity. This left temporal activation for FVs was obtained for both implicit and explicit tasks, indicating that this activation occurred even if participants were not asked to make a judgment of familiarity, as previously shown in a behavioral study (Burton and Bonner, 2004). There is thus several evidences for the existence of very fast, implicit processing of familiarity judgment affecting the auditory sensory integrative process. Our results suggest that these brain processes may be more involved in the discrimination of FVs than in the discrimination of one’s OV, confirming that familiarity judgment and self-perception involve different neuronal networks.

MMN and P3a evoked by an unfamiliar voice were also compared with responses to OV and a FV. Consistent with our previous study (Graux et al., 2013), we showed that P3a amplitude evoked by OV was significantly smaller than that evoked by an unfamiliar voice. Moreover, we showed in this study that MMN amplitude evoked by a FV was significantly greater than that evoked by an unfamiliar voice. Passive discrimination of familiar or personally significant stimuli has been reported to elicit greater MMN and/or P3a amplitude than unfamiliar stimuli or stimuli without personal significance (Frangos et al., 2005; Jacobsen et al., 2005; Beauchemin et al., 2006; Holeckova et al., 2006; Roye et al., 2007). Enhanced MMN to familiar/personally significant stimuli has been reported to be elicited greater MMN and/or P3a amplitude than unfamiliar stimuli or stimuli without personal significance (Frangos et al., 2005; Beauchemin et al., 2006). Our results confirmed these findings (with enhanced MMN to OV vs unfamiliar voice) and suggest that long-term memory traces for familiar/personally significant stimuli can impact on MMN elicitation. Moreover, enhanced P3a to familiar/personally significant stimuli has been reported to be associated with the involvement of involuntary attention switching toward deviants with personal significance (Roye et al., 2007). In our study, reduced P3a amplitude to OV compared to FV or unfamiliar voice suggested that OV does not trigger the same attention as unknown or FVs and must be differentiated from other familiar/personally significant stimuli. As suggested in Graux et al. (2013), this highlights the
difference between two kinds of self: the self as self-image representation (OV) and self as something self-related (FV).

REFERENCES

Agner, P.-E., Jerbi, K., Caclin, A., Bertrand, O. (2011). ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Computational Intelligence and Neuroscience*, 2011, 158970.

Beauchemin, M., De Beaumont, L., Vannasing, P., et al. (2006). Electrophysiological markers of voice familiarity. *The European Journal of Neuroscience*, 23, 3081–6.

Birkett, P.B., Hunter, M.D., Parks, R.W., et al. (2007). Voice familiarity engages auditory cortex. *Neuroreport*, 18, 1375–8.

Blair, R.C., Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, 30, 518–24.

Burton, A., Bonner, L. (2004). Familiarity influences judgments of sex: the case of voice recognition. *Perception*, 33, 747–52.

Caharel, S., Poiroux, S., Bernard, C., Thibaut, F., Lalonde, R., Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *The International Journal of Neuroscience*, 112, 1499–512.

Compumedics Neuroscan (2003). Offline Analysis of Acquired Data. (SCAN 4.3 – Vol. II, EDIT 4.3). [Software Manual]. El Paso, TX.

Frangos, J., Ritter, W., Friedman, D. (2005). Brain potentials to sexually suggestive whistles show meaning modulates the mismatch negativity. *Neuroreport*, 16, 1313–7.

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., Bruneau, N. (2013). My voice or yours? An electrophysiological study. *Brain Topography*, 26, 72–82.

Guthrie, D., Buchwald, J.S. (1991). Significance testing of difference potentials. *Psychophysiology*, 28, 240–4.

Holeclova, I., Fischer, C., Giard, M.H., Delpeuch, C., Morlet, D. (2006). Brain responses to a subject’s own name uttered by a familiar voice. *Brain Research*, 1082, 142–52.

Jacobsen, T., Schroger, E., Winkler, I., Horváth, J. (2005). Familiarity affects the processing of task-irrelevant auditory deviance. *Journal of Cognitive Neuroscience*, 17, 1704–13.

Kaplan, J.T., Aziz-Zadeh, L., Uddin, L.Q., Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, 3, 218–23.

Keyes, H., Brady, N., Reilly, R.B., Foxe, J.I. (2010). My face or yours? Event-related potential correlates of self-face processing. *Brain and Cognition*, 72, 244–54.

McCarthy, G., Wood, C.C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62, 203–8.

Nakamura, K., Kawashima, R., Sugiuira, M., et al. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia*, 39, 1047–54.

Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72, 184–7.

Perrin, F., Pernier, J., Bertrand, O., Giraud, M.H., Echallier, J.F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, 66, 75–81.

Platek, S.M., Keenan, J.P., Gallup, G.G., Mohamed, F.B. (2004). Where am I? The neurological correlates of self and other. *Brain Research. Cognitive Brain Research*, 19, 114–22.

Platek, S.M., Kemp, S.M. (2009). Is family special to the brain? An event-related fMRI study of familiar, familial, and self-face recognition. *Neuropsychologia*, 47, 849–58.

Platek, S.M., Loughhead, J.W., Gur, R.C., et al. (2006). Neural substrates for functionally discriminating self-face from personally familiar faces. *Human Brain Mapping*, 27, 91–8.

Royer, A., Jacobsen, T., Schröger, E. (2007). Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones. *European Journal of Neuroscience*, 26, 784–90.

Shah, N.J., Marshall, J.C., Zafiris, O., et al. (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain*, 124, 804–15.

Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., Kawashima, R. (2005). Cortical mechanisms of visual self-recognition. *Neuropsychologia*, 43, 143–9.

Taylor, M.I., Arsalidou, M., Bayless, S.I., Morris, D., Evans, J.W., Barbeau, E.J. (2009). Neural correlates of personally familiar faces: parents, partner and own faces. *Human Brain Mapping*, 30, 208–20.

Turk, D.J., Heatherton, T.F., Kelley, W.M., Funnell, M.G., Gazzaniga, M.S., Macrae, C.N. (2002). Mike or me? Self-recognition in a split-brain patient. *Nature Neuroscience*, 5, 841–2.

Yonelinas, A.P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.