Electrophysiological evidence for an early processing of human voices

Ian Charest, Cyril R Pernet, Guillaume A Rousselet, Ileana Quiñones, Marianne Latinus, Sarah Fillion-Bilodeau, Jean-Pierre Chartrand, Pascal Belin

BMC Neuroscience

Research article

Electrophysiological evidence for an early processing of human voices

Ian Charest*†[1], Cyril R Pernet†[2], Guillaume A Rousselet†[1], Ileana Quiñones[3], Marianne Latinus†[1], Sarah Fillion-Bilodeau[4], Jean-Pierre Chartrand[4,5] and Pascal Belin†[1,5]

Address: [1]Centre for Cognitive NeuroImaging (CCNi) & Department of Psychology, University of Glasgow, Glasgow, UK, [2]SFC Brain Imaging Research Centre, Division of Clinical Neurosciences, University of Edinburgh, Edinburgh, UK, [3]Cuban Neuroscience Centre, Department of Cognitive Neuroscience, Havana, Cuba, [4]Département de Psychologie, Université de Montréal, Montréal, QC, Canada and [5]International Laboratory for Brain, Music and Sound Research, Université de Montréal & McGill University, Montreal, Canada

Email: Ian Charest* - i.charest@psy.gla.ac.uk; Cyril R Pernet - cyril.pernet@ed.ac.uk; Guillaume A Rousselet - g.rousselet@psy.gla.ac.uk; Ileana Quiñones - ileana@cneuro.edu.cu; Marianne Latinus - m.latinus@psy.gla.ac.uk; Sarah Fillion-Bilodeau - sarah.fillion-bilodeau@umontreal.ca; Jean-Pierre Chartrand - jean-pierre.chartrand@umontreal.ca; Pascal Belin - p.belin@psy.gla.ac.uk

* Corresponding author †Equal contributors

Published: 20 October 2009

BMC Neuroscience 2009, **10**:127 doi:10.1186/1471-2202-10-127

Received: 20 May 2009
Accepted: 20 October 2009

This article is available from: http://www.biomedcentral.com/1471-2202/10/127

Abstract

Background: Previous electrophysiological studies have identified a "voice specific response" (VSR) peaking around 320 ms after stimulus onset, a latency markedly longer than the 70 ms needed to discriminate living from non-living sound sources and the 150 ms to 200 ms needed for the processing of voice paralinguistic qualities. In the present study, we investigated whether an early electrophysiological difference between voice and non-voice stimuli could be observed.

Results: ERPs were recorded from 32 healthy volunteers who listened to 200 ms long stimuli from three sound categories - voices, bird songs and environmental sounds - whilst performing a pure-tone detection task. ERP analyses revealed voice/non-voice amplitude differences emerging as early as 164 ms post stimulus onset and peaking around 200 ms on fronto-temporal (positivity) and occipital (negativity) electrodes.

Conclusion: Our electrophysiological results suggest a rapid brain discrimination of sounds of voice, termed the "fronto-temporal positivity to voices" (FTPV), at latencies comparable to the well-known face-preferential N170.

Background

The field of study of cortical processing of complex sounds has been highly productive in the recent past both in humans and monkeys. A model similar to the "what" and "where" segregation of the visual processing network has been suggested for auditory processes of sound identification and localization [1,2]. Regions within the superior and middle temporal cortices and the inferior prefrontal gyrus have been identified as candidates for the "what" pathway of the auditory stream, whereas a "where" pathway would rely on the posterior temporal cortex and the inferior and superior parietal cortices [3-6]. Within the auditory "what" pathway, functional magnetic resonance imaging (fMRI) studies have identified the

'temporal voice areas' (TVA - [7]), i.e. bilateral auditory areas situated along the superior temporal sulcus (STS) showing a greater response to human vocalisations than to other sound categories [8,9]. These regions were later found to be species-specific as they elicited stronger responses to human vocalisations compared to non-human vocalisations [10,11]. In a recent fMRI study in macaques, Petkov et al. (2008) found a region of the secondary auditory cortex which showed a comparable preference for conspecific vocalisations over vocalisations from other species or non-vocal sounds. These findings suggest a long evolutionary history of voice-preferential processing [12]. Yet, the time course of voice processing remains unclear.

In studies using intracranial electrophysiological recordings in human participants, early responses to sound stimulation were shown to reach the primary auditory cortex (A1) as early as 15 ms after sound onset [13-15] and differences between sound categories have been observed as soon as 55 ms after this early response to sound stimulation. Using evoked related potentials (ERPs) and an oddball paradigm, Murray et al. (2006) reported early ERP differences between man-made (sound of a bicycle bell, glass shattering, telephone...) and living auditory objects (baby cries, coughing, birdsong, cat vocalization...) as early as 70 ms after stimulus onset [16]. Although most of the living sounds in that study consisted of vocalisations, it remains unclear whether the effect was driven by voices or not. ERP studies providing evidence directly relevant to the speed of voice/non-voice categorisation are scarce. Two studies found a larger response to sung voices when compared to instrumental sounds at a latency of 320 ms after stimulus onset, with a fronto-central distribution, which was termed the "voice-specific response" (VSR) [17,18]. To further assess the VSR, Gunji et al. (2003) used magnetoencephalography (MEG) and analysed two components of the evoked response: the N1m and the sustained field observed 400 ms after stimulus onset. They observed no difference in the magnitude of the N1m component between voices and instrumental sounds; however the source strength of the 400 ms sustained field was larger for vocal sound than for instrumental sounds [19]. Both components had sources in Heschl's gyrus in both hemispheres.

Although previous studies did not address directly the voice/non-voice discrimination process, some of them suggest the existence of earlier correlates of voice processing. Indeed, effects of voice familiarity [20], voice gender adaptation [21], human vs. computer voice [22], voice priming [23], speech vs. tones [24], and speaker identity [25] have been observed between 150 ms to 200 ms.

The relatively long latency of voice vs. non-voice ERP differences (320 ms) stands in strong contrast with these results and with the early living/non-living distinction reported by Murray et al. (2006).

In the present study, we investigated the speed of voice processing by measuring ERPs in response to sounds from three categories -- voices, bird songs and environmental sounds -- while participants were performing an incidental target (pure tone) detection task. We hypothesised that since neural correlates of voice paralinguistic characteristics were observed in the range of 150 to 200 ms, investigating neural correlates of voice recognition by directly comparing neural responses to voices with those of sounds from other categories should lead to differences in the same latency range or earlier, in contrast with the previously reported 320 ms.

Methods

Participants

Thirty-two French-speaking adults (15 females, mean = 27.25 y/o, std. 8.23), participated in the study. They all reported normal audition and no neurological problems. They all gave informed written consent and the study was approved by the University of Montreal ethics committee. Participants were compensated 30 Canadian dollars for their time. Some of the subjects were included in the 'novice' group of a study focusing on differences between ornithologists and novices [26], although they were never informed of the expertise nature of this study, and only informed to press a response button as fast as they could when they heard a 1000 Hz pure tone target.

Stimuli and design

Stimuli consisted of 450 sound samples, 150 in each of three categories: bird songs, human vocalisations (73 speech items, 77 vocalisations) and environmental sounds (30 natural sounds, 60 instruments and 60 mechanical sounds). The bird songs were selected from the « Chants d'oiseaux du Québec et de l'Amérique du Nord » (2004; Peterson Guides coll, Broquet/Cornell laboratory of ornithology) audio CD. Other sound stimuli came from commercially available sources and from recordings in the laboratory. Sounds were edited using Cool Edit Pro (Syntrillium Corporation, Phoenix, Arizona, USA) to a sampling rate of 22050 Hz, a 16-bit resolution, and duration of 200 ms with a 10-ms linear attack and decay. They were all root mean square (RMS) normalised using Matlab (The MathWorks Inc., Natick, Massachusetts, USA). A sample of the stimuli is available for consultation online http://vnl.psy.gla.ac.uk/resources/speed_of_voice_sounds/.

Analyses of sound power in the temporal, spectral and time-frequency domains were performed using one-way ANOVAs at each time, frequency, or time-frequency bin (11.6 ms, 43 Hz) using Matlab (figure 1). In the time domain, power differences between the three sound cate-

gories were observed from 11 to 35 ms (minimum F = 8.05, p < 0.05) and from 76 to 100 ms (minimum F = 8.08, p < 0.05; figure 1c). Post-hoc tests showed that differences in the temporal domain were driven by voices from 11 to 35 ms and by environmental sounds from 76 to 100 ms, with significantly less power than the other two categories(p < 0.05).

In the frequency domain, significant effects were found around 500 Hz (minimum F = 25.52, p < 0.05) and 4000 Hz (minimum F = 20.09, p < 0.05; figure 1a), reflecting the smaller power in low frequencies (0-2000 Hz) and greater power at intermediate frequencies (2000-8000 Hz) for bird songs relative to the other two categories (p < 0.05); and more power at high frequencies (> 8000 Hz) for environmental sounds compared to the other two categories although this last difference was not significant (p > 0.05).

In the time-frequency domain, significant effects of sound category were observed across all frequencies for early latencies only (0-30 ms; p < 0.05), whilst differences in frequency bands around 500 Hz and 4000 Hz were observed at all time-bins (p < 0.05; figure 1b).

Procedure: sound identification

10 participants were included in a verification study in which they listened to each sound from the three stimulus

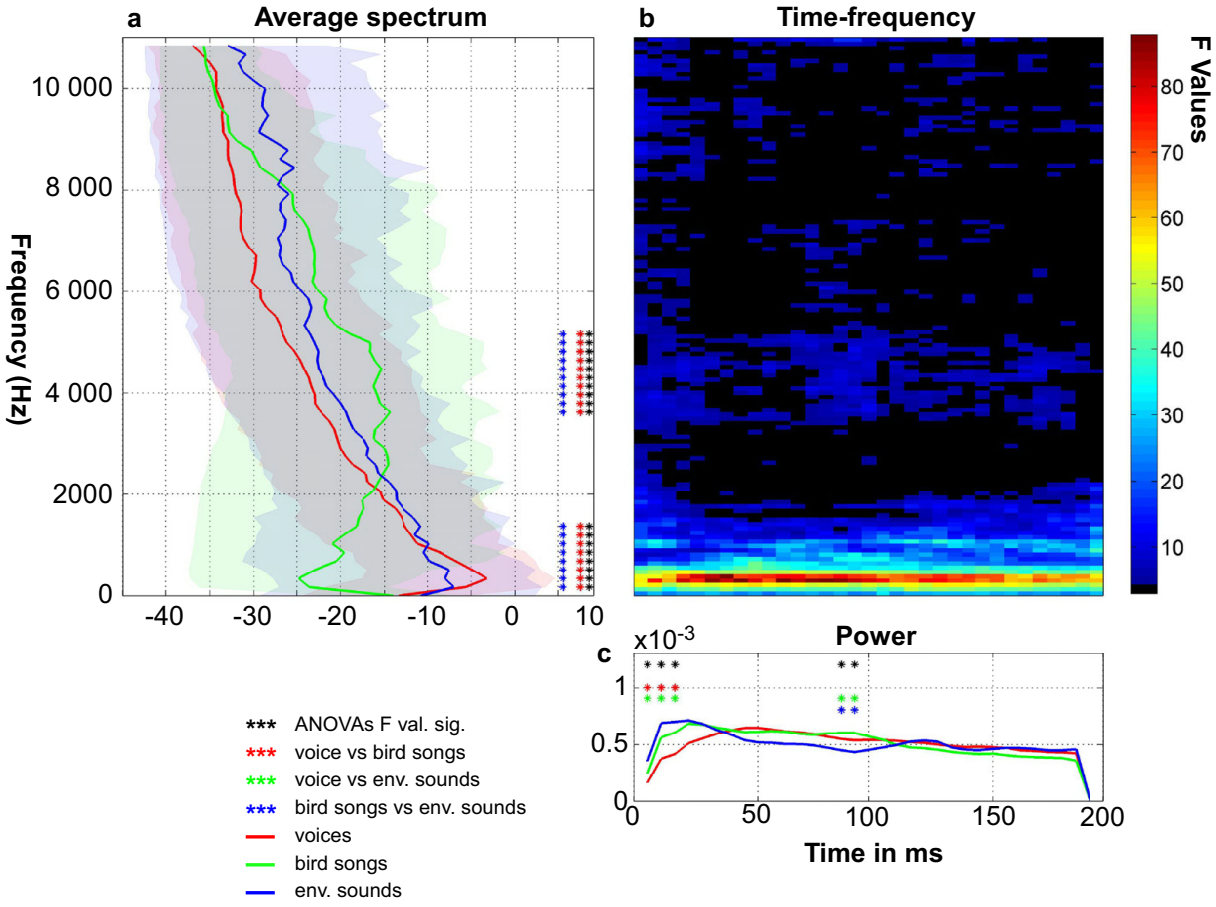

Figure 1

Acoustical differences between sound categories. a) Red, blue and green lines show the average power spectrum of their respective sound categories (see legend). Stars highlight frequency bands showing a significant effect of sound category. 95% confidence intervals are indicated by shaded areas. b) Time-frequency analysis. The significance of the sound categorical difference effect is indicated by the colormap. Black areas indicate non-significant results. Frequency axis as in figure 1a. c) Power Analysis in time. Red, blue and green lines show the average power of the respective sound categories over the 200 ms duration. Black stars indicate time points showing a significant effect of sound category; blue and green stars indicate significant post-hoc tests. Note that the largest effect of sound category is a higher power in the low frequencies for voices and environmental sounds compared to bird songs.

categories and performed a three alternative forced choice categorisation task. All categories showed similar levels of recognition (percent correct ± standard deviation: 92 ± 4% for voices, 90 ± 10% for bird songs and 85 ± 8% for environmental sounds). Statistical analyses testing the null H0 hypothesis with a bootstrap method (1000 samples with replacement of the categorical labels; procedure explained in more details in the EEG recordings and analysis section) showed no significant differences between the three categories ($p > 0.05$).

Procedure: Task

Participants were seated in a sound-proof cabin and were presented with each sound from the three categories in a pseudo-random order with a 3000-3500 ms random inter-stimulus-interval (ISI). Each sound was played twice, in two different runs. Stimuli were presented via Beyerdynamic DT 770 headphones at a self-adjusted comfortable level of about 65 dB sound level, as measured using a Lutron Sl-4010 sound level meter. Participants were instructed to detect a 1000 Hz sinusoidal pure sound target with a 10% probability of occurrence. They were instructed to press a button each time they heard the target stimulus, and also to minimise blinking, head motion and swallowing.

Procedure: EEG recordings and analysis

Electroencephalography (EEG) data were recorded continuously at a 250 Hz sampling frequency using a BrainAmp amplifier (Brainproduct-MR 64 channel-Standard; 62 EEG electrodes, one EOG, one ECG, Brain Products, Munich, Germany) using a 0.5-70 Hz band-pass filter. The 64 Ag/AgCl electrodes were attached using a BrainCap 10-20 array, with an on-line reference at electrode FCz, a ground electrode on the midline posterior to Oz, an ECG electrode attached above the left collar bone, and an EOG electrode attached above the zygomatic bone, below the left eye. The electrode impedances were kept below 10 kΩ throughout the recording.

EEG recordings were analysed using EEGLAB v6.01 [27], under Matlab (The MathWorks Inc., Natick, Massachusetts, USA). Trials with abnormal activities were excluded based on a detection of extreme values (± 100 μV for all channels), abnormal trends (trial's slope larger than 75 μV/epoch and a regression R^2 larger than 0.3), and abnormal distribution (when the trial's kurtosis fell outside five standard deviations of the kurtosis distribution for each single electrode or across all electrodes) [27,28]. Data were then re-referenced to the average of all electrodes and band-pass filtered in the range 1-30 Hz. For each stimulus category, EEG epochs of 1200 ms, starting 200 ms before stimulus onset, were averaged, and the mean pre-stimulus activity was subtracted from the activity at each time point.

Statistical inferences on amplitude differences between sound categories were performed for each electrode and time point using bootstrap procedures implemented in Matlab (The MathWorks Inc., Natick, Massachusetts, USA). The null hypothesis H0 that the three conditions were sampled from populations with similar means was evaluated by sampling with replacement, independently for each subject, among the three categories. The samples consisted of the full electrodes by time-points matrices. This was followed by averaging the ERP across subjects for each resampled condition, and then computing the differences between the means of two fake conditions. This process was repeated 9999 times, leading to a distribution of 10000 bootstrapped estimates of the mean difference across subjects between two ERP conditions. Then the 99.9% percent confidence interval was computed (alpha = 0.001). Finally, the difference between two sample means was considered significant if it was not contained in the 99.9% null hypothesis confidence interval [29,30]. The statistical analyses were restrained to a [-200 to 500 ms] time-window as we were mainly interested in rapid brain discrimination processes.

In order to evaluate the relative contribution of speech and non-speech vocal sounds, post-hoc analyses were performed on a sample of 50 speech sounds (vowel, word, consonant, etc.) and 50 non-speech vocal sounds (cough, laughter, yawn, gargle, etc.) selected from the voice category (the 50 most ambiguous were excluded from the analysis). Electrophysiological responses to these two categories were then compared independently to the average of bird songs and environmental sounds using the bootstrap procedure described above.

Results

Behavioral results: target detection task during EEG recordings

Mean reaction times (RT) to correctly detected targets were 563 ms across subjects. In terms of accuracy, 97.82% of the targets were followed by a button press (hits) and only 0.04% of the non-target events were followed by a button press (false alarms (FA)).

Responses to targets were split according to whether targets followed voices, bird songs or environmental sounds. Repeated measures ANOVAs were implemented on RT for correct responses, proportion of hits and false alarms using SPSS (SPSS, Chicago, Illinois). We did not observe differences in proportion of hits ($F(2,62) = 0.532$, $p = 0.590$), nor in FA rate ($F(2,62) = 1.882$, $p = 0.161$), but we observed a significant difference in RT ($F(2,62) = 6.745$, $p = 0.002$). Paired samples t-tests indicated significantly longer reaction times in response to the pure tone following presentation of bird songs than for voices ($t = -4.63$, $df = 31$; $p < 0.001$) and environmental sounds ($t = 2.76$, $df =$

31; $p < 0.01$), which did not differ from each other ($t = -0.357$, $df = 31$; $p = 0.724$). Mean values and their standard deviations for the behavioral results are reported in Table 1.

Event Related Potentials

Voices vs. bird songs and environmental sounds

ERPs in all participants showed the classical N1-P2 waveform components with central topographic distribution [31-34]. Results from the comparison of ERPs to voice vs. the other categories are shown in figure 2. The earliest amplitude differences were observed for the voice vs. bird song comparison, emerging around 64 ms after sound onset at electrodes O1, PO3 and PO4 (figure 2a), and about 10 ms later at most of the occipital and frontal electrodes. These differences peaked at about 200 ms, and lasted until 300 ms after stimulus onset. The earliest amplitude differences between voices and environmental sounds were observed at 120 ms on fronto-temporal electrodes FT8 and FT7, peaking at about 200 ms and lasting until 400 ms after stimulus onset (figure 2b). Significant amplitude differences between voices and environmental sounds were also observed around 200 ms on several occipital electrodes (figure 2b). A conjunction of these two differences revealed a broadly distributed pattern of ERPs with a preferential response to voices (figure 2c). While bilateral fronto-temporal electrodes (FC5, FC6) showed a greater positivity in response to voices, a larger negativity was observed at occipital locations (PO7, PO8).

At fronto-temporal electrode FC6 (right hemisphere), a significant amplitude difference showing a smaller negative potential for voice compared to both bird songs and environmental sounds was observed between 132 ms and 152 ms (figure 3a). Following this smaller negativity, a larger positive ERP amplitude elicited by voice sounds compared to bird and environmental sounds was observed at fronto-temporal electrode FC6 as early as 164 ms extending to 280 ms (figure 3a), and from 188 ms to 268 ms at fronto-temporal electrode FC5 (left hemisphere; figure 3c). At occipital electrodes, larger ERP negativities for voices compared to both bird songs and environmental sounds were observed from 200 ms to 220 ms with maximal difference at 200 ms at electrode PO8 (right hemisphere, figure 3g), and from 212 ms to 232 ms with maximal difference at 212 ms at electrode PO7 (left hemisphere, figure 3h).

Speech contribution to the voice effect

In order to evaluate the contribution of speech information to the voice-preferential response observed in the time-window of the auditory P2, voice stimuli were separated in speech vocal sounds (clear presence of articulated speech, n = 50) and non-speech vocal sounds (coughs, laughs, etc., n = 50). Similar patterns of amplitude difference were observed for speech and non-speech voice stimuli when compared to other categories (figure 4a and 4b). Both showed a significant difference around the P2 component starting at 164 ms (figure 4c) although effects were broader and longer lasting for speech than non-speech sounds. Finally, as one can expect, speech vs. non-speech stimuli showed significant differences over many electrodes (figure 4d). Importantly, effects appeared first between 80 and 120 ms and later on from 224 ms over fronto-temporal electrodes and from 272 ms over occipital electrodes (reversed polarity). These differences between speech and non-speech stimuli were clearly different from those observed for voices (speech and/or non-speech stimuli) compared to environmental and bird sounds.

Other categorical effects

In addition to the stronger voice responses reported above, another categorical effect was observed: bird songs elicited smaller N1 (100 ms) and P2 (200 ms) components than voices and environmental sounds at the vertex electrode Cz (figure 3e).

Discussion

Scalp recordings were measured in 32 healthy adult participants to investigate the time-course of brain activity associated with the presentation of 3 categories of brief (200 ms) sounds - voices, bird songs and environmental sounds - while participants performed a pure tone detection task. We observed significantly larger ERP amplitudes for voices compared to other sound categories at fronto-

Table 1: Results of the target detection task during ERPs

	RT (ms)	*% Hits*	*% Miss*	*% FA*
Voices	553 ± 174	98.12 ± 4.79	1.88 ± 4.79	0.08 ± 0.32
bird songs	581 ± 165	97.71 ± 4.87	2.29 ± 4.87	0.02 ± 0.05
environmental sounds	556 ± 156	97.63 ± 6.17	2.37 ± 6.17	0.14 ± 0.18

Average reaction times (RT, in ms), and proportion of hits and miss were split depending on whether the pure tone target was preceded by the presentation of voices, bird songs or environmental sounds. The table also presents the relevant standard deviations from the mean. Proportion of false alarms following the presentation of the 3 sound categories are also presented in the table with standard deviations from the mean.

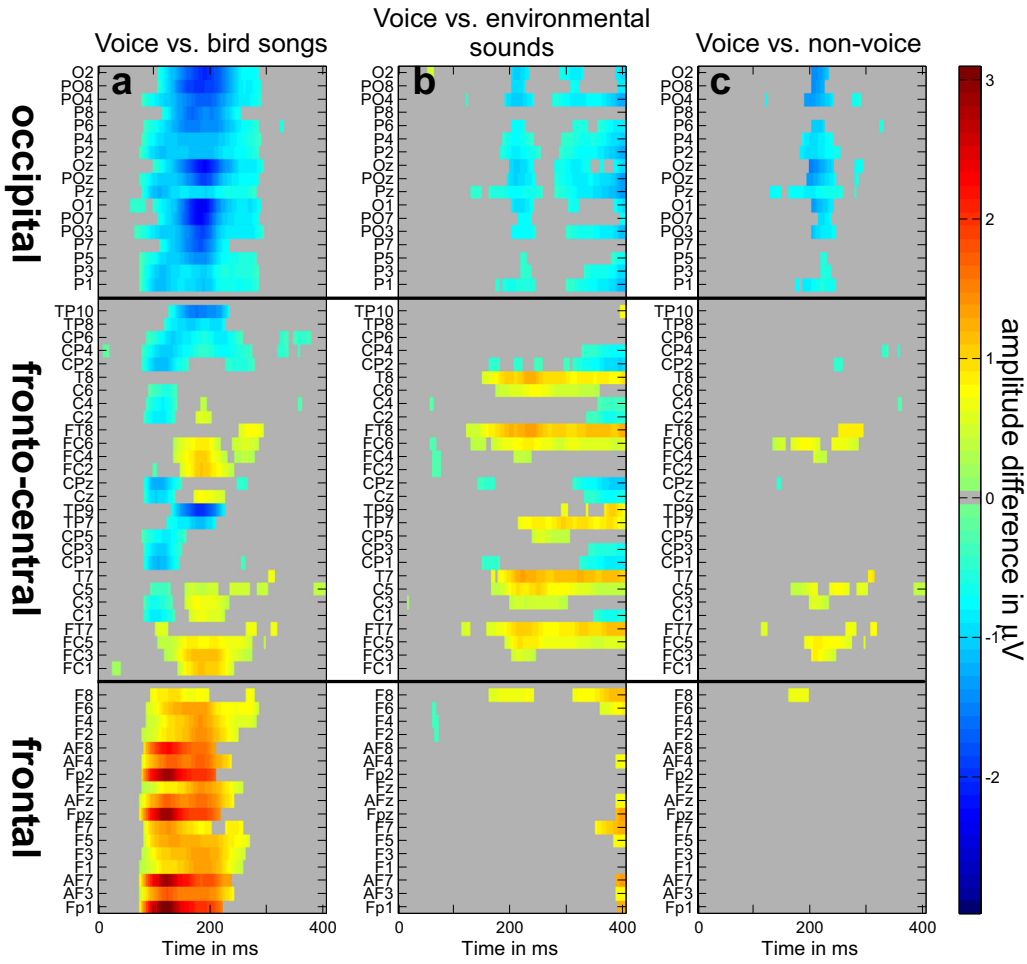

Figure 2
ERP bootstrap results. Significant ERP amplitude difference at all electrodes (vertical axis). Gray areas represent non-significant ERP amplitude differences. Significant average ERP amplitude differences (µV) are represented in an increasing gradient from blue to red (jet64). Bootstrap tests revealed significant amplitude differences between (a) voices vs. birdsongs, (b) voices vs. environmental sounds, and (c) voices vs. both birdsongs and environmental sounds (conjunction of (a) and (b)).

temporal (positivity) and occipital (negativity) electrodes, emerging as early as 164 ms after stimulus onset and peaking around 200 ms (see figures 2, 3), an electrophysiological response termed the "fronto-temporal positivity to voices" (FTPV; [35,36]).

Lack of voice sensitive response at electrode Cz

Results from electrode Cz did not show a consistent preference for voice over other sound categories at any latency. The bootstrap results indicated that ERPs to voices at Cz were never simultaneously larger than both bird songs and environmental sounds (figure 3e). On the contrary, bird songs elicited smaller amplitudes than both environmental sounds and voices on the N1 (~100 ms) and P2 (~200 ms) components. This effect is in line with (i) the difference in RTs showing that targets following bird songs were processed slower than when following voices or environmental sounds; (ii) subjects might have been less familiar with bird songs than with voices or environmental sounds, in line with recent findings showing an enhancement by familiarity of the N1 and the P2 components, which is predictive of the effects we observed, considering that subjects were more familiar to voices and environmental sounds [37]; (iii) acoustic analyses showing that bird songs were the most distinctive category (figure 1).

In our study, the absence of the "VSR" reported by Levy et al. (2001) around 320 ms after sound onset, could be explained by differences in materials, or experimental design, or both. In order to recognise the target and perform the task as fast as they could, subjects had to maintain their attention on every stimulus that was presented to them. This is consistent with Levy et al., (2003) who

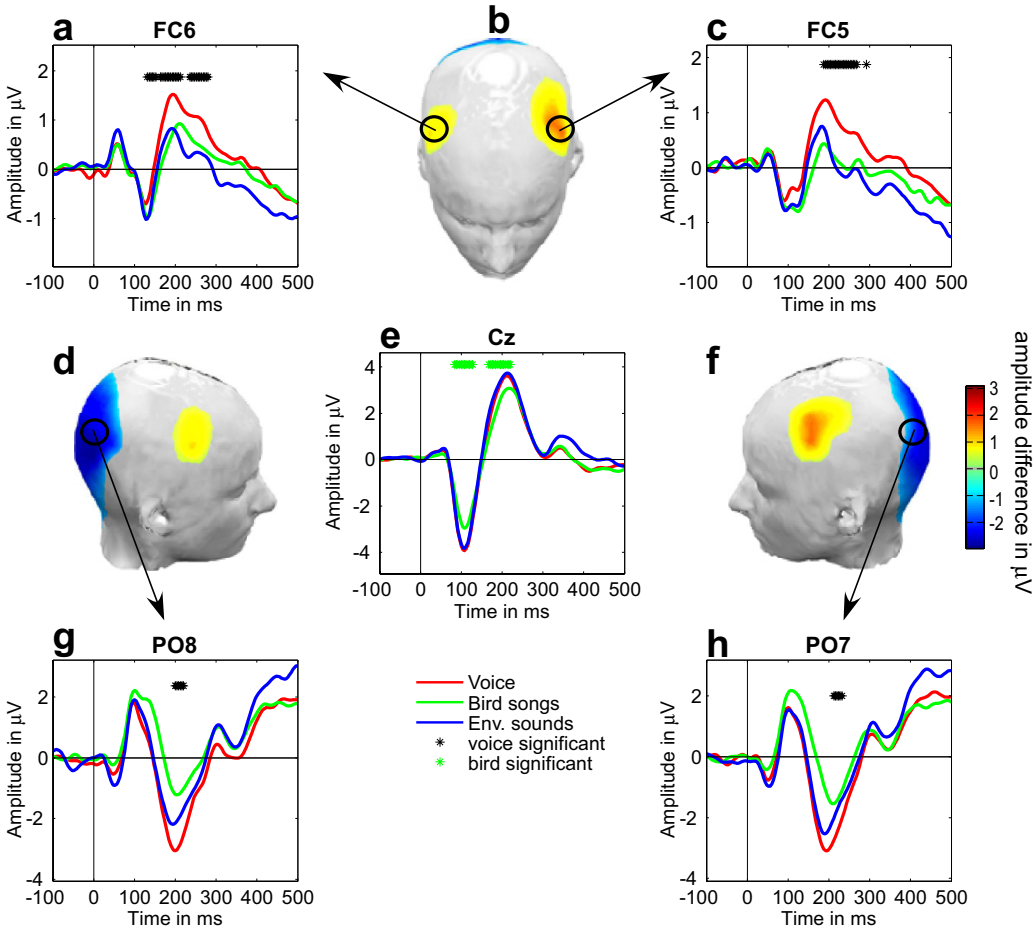

Figure 3

Electrophysiological responses and bootstrap results. a, c, e, g, h) Colored lines indicate ERP waveforms for each category. Black stars indicate 4 ms time bins at which electrophysiological responses to voices was significantly stronger than both the two other stimulus categories (voice > non-voice). Green circles indicate 4 ms time-bins at which ERPs bird songs were significantly greater than to both voice and environmental sounds. b, d, f) Headplot showing the voice > non-voice significant difference at 200 ms and electrode positions for panels a, c, g and h. Colormap indicates significant amplitude difference (in µV) with greater ERP amplitude (positive or negative) for voices compared to both bird songs and environmental sounds. Gray areas were not significant.

mentions that when participants attended to the stimulation sequence focusing on a feature other than timbre, the "VSR" was absent. In our experiment, because the target was a 1000 Hz pure tone, participants might have focused on pitch features, thus explaining the differences found for bird songs on the N1 and P2 (figure 3e) and the absence of a "VSR" in the latencies suggested by Levy et al., (2001). Another potential aspect leading to the differences in the scalp localisation of the voice related effects we observed and the results reported by Levy et al., (2001, 2003) is the choice of a common reference to the nose, whereas we opted for an off-line average reference [38]. Finally, the difference in findings between the studies by Levy et al. (2001, 2003) and the present study may lie in their choice of a target tone (piano tone) that belongs to the category of musical instruments, like all their stimuli except for the voices, whereas we were careful in the present study to choose a target (pure tone) that clearly did not belong to any of the three compared stimulus categories.

Early categorical differences

The earliest categorical difference was observed at the N1 latency (starting as early as 80 ms), with smaller magni-

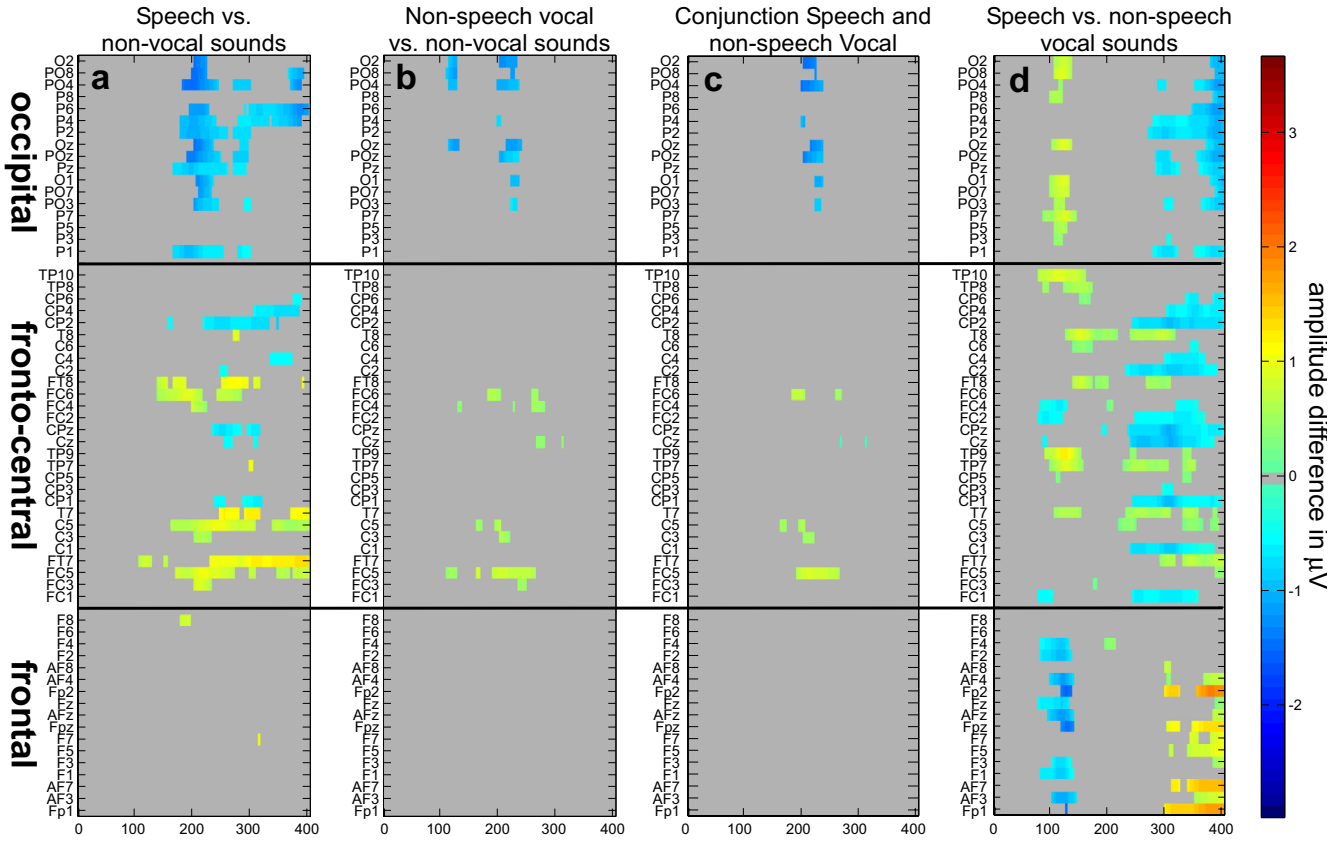

Figure 4

Post-hoc results. ERP analyses at all electrodes for the speech and non-speech vocal sounds comparisons. The color code is shown on the right side. Electrodes are stacked along the vertical axis. The horizontal black lines separate the different groups of electrodes organised in frontal, fronto-central and occipital electrodes. Gray areas were not significant. Significant ERP differences (amplitude difference in µV averaged across subjects) are represented in an increasing gradient from blue to red (jet64). Bootstrap analyses revealed significant amplitude differences between (a) speech sounds and the two non-voice categories grouped together in a conjunction test, (b) non-speech vocal sounds and the two non-voice categories grouped together in a conjunction test, c) the conjunction of a) and b), and d) the speech vs. non-speech vocal sounds test.

tude in response to bird songs at central and some occipital electrodes (figure 3e). These early categorical differences are comparable to latencies reported by Murray et al., (2006), who found a categorical difference between man-made and living auditory objects as early as 76 ms post sound onset.

We interpret the early birdsong ERP difference as reflecting acoustical differences between sound categories: whereas acoustic energy was concentrated at low frequencies for both voices and environmental sounds (figure 1a), it peaked at much higher frequencies for birdsongs. This finding is consistent with the well-established sensitivity of the auditory N1 component to acoustical structure [32,33]. Another early categorical difference was observed emerging 132 ms after sound onset on fronto-central electrode FC6, with a smaller negative potential in response to voices (figure 3a). This difference on the N1 component at fronto-temporal electrode FC6 could also be related to acoustical differences: lower power was observed for voices at sound onset (figure 1c). This is consistent with findings that relate the acoustic energy at stimulus onset with ERP amplitude on the N1 components [39-41].

The FTPV: a rapid brain discrimination of sounds of voice emerging at 164 ms

The larger amplitude observed at fronto-temporal electrodes FC5 and FC6 in response to voice stimuli is consistent with our hypothesis of an early time-course for voice discrimination. As early as 164 ms post stimulus-onset, ERPs at electrodes FC5 and FC6 were consistently larger for voices than bird songs and environmental sounds. The same pattern was observed at similar latencies at occipital

electrodes PO7 and PO8, with reverse polarities. Figure 3 shows an early ERP to voice at electrode locations FC5, FC6, PO8 and PO7, but this effect was also observed at the vast majority of occipital electrodes and at some frontal electrodes (figure 2c). At 200 ms, this electrophysiological response to voices reached nearly twice the amplitude of ERPs to other sounds, especially at fronto-temporal electrodes FC5 and FC6. This effect does not relate to acoustical differences, which were observed at Cz on the N1 and P2 components and at FC6 on the N1. In addition, the ERP voice response does not require subjects to make explicit discrimination among sound categories (see Methods).

The latency of this electrophysiological marker of human voice processing is in keeping with previous studies addressing complementary questions. For example, Beauchemin et al. (2006), found EEG sensitivity to voice familiarity in the time-range of the auditory P2. Although they used different stimuli, different experimental design (an oddball paradigm eliciting a Mismatch Negativity), and different EEG recording procedure (linked mastoids vs. average reference) the latencies they report are consistent with the voice effects we observed in the latencies of the auditory P2. Therefore, neuronal activity in the time window of the auditory P2 seems to be sensitive to voice vs. non-voice differences, and also higher level cognitive processes such as voice familiarity, voice identity, and voice gender [20-25].

Voice-related brain mechanisms

This early FTPV probably corresponds to activity originating from the "what" part of the auditory stream [3-6], and most likely in the temporal voice areas [7-9]. These brain regions have been reported to be very close (two or three synapses) to core auditory regions [42] and could potentially include areas that contain a large amount of voice-selective cells as it has been demonstrated for face selective neuronal patches in the macaque brain [43], although this is very speculative and remains to be verified.

An auditory counterpart of the face-preferential N170?

The well established face-preferential N170 ERP is characterised by larger amplitude in response to faces compared to other visual object categories from ~130 ms to 200 ms and peaking at around 170 ms after stimulus onset [44-46]. The time course of the present FTPV shows some temporal coincidence with that of the N170: significant voice/non-voice amplitude differences emerged at 164 ms post onset and were well present at several electrodes at 170 ms. Although onsets are delayed by about 30 ms, this time-course similarity between face-preferential and voice-preferential responses offers interesting avenues for future studies. Because the same broad types of information -speech, identity, affect- are typically integrated across face and voice in social interactions, a parsimonious principle of organisation would be that unimodal preferential effects for faces and voices emerge at a comparable timeframe, well-suited for integrative mechanisms [47,48]. Thus, we suggest that the FTPV could provide an auditory analogue of the well known N170 [44,49].

The role of speech sounds

As illustrated on figure 4, speech sounds contained in the voice category contributed strongly to the FTPV. However, it also appears that vocalisations (non-speech) elicited a preferential response, although the pattern of activation was restricted to a few electrodes (in fact observed on electrodes showing the strongest averaged effect). Both observations are consistent with previous fMRI results that have shown i) a greater activity throughout the auditory cortex to speech sounds compared to their scrambled versions and ii) a greater activity to non-speech vocal sounds compared to their scrambled version restricted to the right anterior superior temporal gyrus/sulcus [50]. Indeed, our results extend fMRI ones, showing i) that the voice preferential response is not speech dependant; ii) that the preferential response for speech and non-speech stimuli has a similar time course; and iii) that the preferential response evoked for non-speech vocal stimuli compared to other sound categories is bilateral and more localised. Further experiment are needed to test if the effects observed specifically for non-speech stimuli, here for electrodes PO3/FC3-FC5, PO4/FC4-FC6, Oz-POz, correspond to activations of the anterior superior temporal gyri/sulci [50].

Limitations

Although this study highlights for the first time an early electrophysiological response to voices, the degree of selectivity of the FTPV remains to be established. To demonstrate the robustness of the preferential electrophysiological responses in the face perception domain, several experiments were designed in order to account for the variety of visual objects [45,51-53] and uncontrolled low-level differences [46,54-58]. Future studies on voice categorisation should use a greater number of sound categories in order to better assess the robustness of this potentially selective response. Because natural sound categories are necessarily characterised by acoustical differences that may contribute to ERP differences, sound categories consisting of acoustical controls such as scrambled versions [50], or sinusoidally amplitude-modulated noise [59,60] could be used in order to rule out the contribution of factors such as amplitude modulation on the ERP.

Another way to better understand the early voice discrimination process would be to design an experiment with two stimulus categories (e.g. voice and monkey vocalisations) and at least two tasks (e.g. (i) human vs. monkey

discrimination and (ii) expressions vs. no-expression discrimination) and a baseline condition, which would allow us to define whether the effect we report is specific, selective or preferential to voices, as described in [61].

Finally, an interesting possibility that remains to be tested is whether the FTPV is driven by increased attention to voices. As Levy et al. (2003) did in their study it would be interesting to manipulate attention in order to test its effect on the rapid brain discrimination of sounds of voice.

Conclusion

We searched for early ERP markers of voice. Our results provide the first evidence of an early electrophysiological response to sounds of human voices termed the "fronto-temporal positivity to voices" (FTPV). This rapid brain response to voices appears in the latency range of the auditory P2, which is comparable to the well-known face preferential N170.

Authors' contributions

IC, SFB, JPC, and PB designed the study. IC, SFB and JPC collected the data. IC conducted the analyses and wrote the manuscript. GAR, CRP, ML and IQ helped analyse the data. CRP, GAR, ML and PB helped revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Rebecca Watson for her help and comments on this study. The study was supported by grants from the UK's Economical and Social Research Council (ESRC), Medical Research Council (MRC), Royal Society, Biotechnology and Biological Sciences Research Council (BBSRC), from the Canadian Foundation for Innovation, the Natural Sciences and Engineering Research Council of Canada, the Canadian institute of Health Research and from France-Télécom.

References

1. Courtney SM, Ungerleider LG, Keil K, Haxby JV: **Object and Spatial Visual Working Memory Activate Separate Neural Systems in Human Cortex.** *Cereb Cortex* 1996, **6:**39-49.
2. Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL: **The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations.** *J Neurosci* 1994, **14:**6336-6353.
3. Alain C, Arnott SR, Hevenor S, Graham S, Grady CL: **"What" and "where" in the human auditory system.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98:**12301-12306.
4. Rauschecker JP, Tian B: **Mechanisms and streams for processing of "what" and "where" in auditory cortex.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97:**11800-11806.
5. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP: **Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex.** *Nat Neurosci* 1999, **2:**1131-1136.
6. Wessinger CM, Van Meter J, Tian B, Van Lare J, Pekar J, Rauschecker JP: **Hierarchical Organization of the Human Auditory Cortex Revealed by Functional Magnetic Resonance Imaging.** *Journal of Cognitive Neuroscience* 2001, **13:**1-7.
7. Pernet C, Charest I, Belizaire G, Zatorre RJ, Belin P: **The temporal voice areas: spatial characterization and variability.** *NeuroImage* 2007, **36:**.
8. Belin P: **Voice processing in human and non-human primates.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2006, **361:**2091-2107.
9. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B: **Voice-selective areas in human auditory cortex.** *Nature* 2000, **403:**309-312.
10. Fecteau S, Armony JL, Joanette Y, Belin P: **Is voice processing species-specific in human auditory cortex? An fMRI study.** *NeuroImage* 2004, **23:**840-848.
11. von Kriegstein K, Smith DRR, Patterson RD, Ives DT, Griffiths TD: **Neural Representation of Auditory Size in the Human Voice and in Sounds from Other Resonant Sources.** *Current Biology* 2007, **17:**1123-1128.
12. Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK: **A voice region in the monkey brain.** *Nat Neurosci* 2008, **11:**367-374.
13. Brugge JF, Volkov IO, Garell PC, Reale RA, Howard MA III: **Functional Connections Between Auditory Cortex on Heschl's Gyrus and on the Lateral Superior Temporal Gyrus in Humans.** *J Neurophysiol* 2003, **90:**3750-3763.
14. Godey B, Schwartz D, de Graaf JB, Chauvel P, Liégeois-Chauvel C: **Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients.** *Clinical Neurophysiology* 2001, **112:**1850-1859.
15. Liegeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P: **Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components.** *Electroencephalogr Clin Neurophysiol* 1994, **92:**204-214.
16. Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S: **Rapid Brain Discrimination of Sounds of Objects.** *J Neurosci* 2006, **26:**1293-1302.
17. Levy DA, Granot R, Bentin S: **Processing specificity for human voice stimuli: electrophysiological evidence.** *Neuroreport* 2001, **12:**2653-2657.
18. Levy DA, Granot R, Bentin S: **Neural sensitivity to human voices: ERP evidence of task and attentional influences.** *Psychophysiology* 2003, **40:**291-305.
19. Gunji A, Koyama S, Ishii R, Levy D, Okamoto H, Kakigi R, Pantev C: **Magnetoencephalographic study of the cortical activity elicited by human voice.** *Neuroscience Letters* 2003, **348:**13-16.
20. Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M: **Electrophysiological markers of voice familiarity.** *European Journal of Neuroscience* 2006, **23:**3081-3086.
21. Zaske R, Schweinberger SR, Kaufmann J, Jurgen M, Kawahara H: **In the ear of the beholder: neural correlates of adaptation to voice gender.** *European Journal of Neuroscience* 2009, **30:**527-534.
22. Lattner S, Maess B, Wang Y, Schauer M, Alter K, AD F: **Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception.** *Human Brain Mapping* 2003, **20:**13-21.
23. Schweinberger SR: **Human brain potential correlates of voice priming and voice recognition.** *Neuropsychologia* 2001, **39:**921-936.
24. Tiitinen H, Sivonen P, Alku P, Virtanen J, Näätänen R: **Electromagnetic recordings reveal latency differences in speech and tone processing in humans.** *Cognitive Brain Research* 1999, **8:**355-363.
25. Titova N, Naatanen R: **Preattentive voice discrimination by the human brain as indexed by the mismatch negativity.** *Neuroscience Letters* 2001, **308:**63-65.
26. Chartrand J-P, Filion-Bilodeau S, Belin P: **Brain response to birdsongs in bird experts.** *Neuroreport* 2007, **18:**335-340.
27. Delorme A, Makeig S: **EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.** *Journal of Neuroscience Methods* 2004, **134:**9-21.
28. Delorme A, Sejnowski T, Makeig S: **Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis.** *NeuroImage* 2007, **34:**1443-1449.
29. Wilcox RR: *Introduction to robust estimation and hypothesis testing* 2nd edition. San Diego, CA: Academic Press; 2005.
30. Wilcox RR: **New Designs in Analysis of Variance.** *Annual Review of Psychology* 1987, **38:**29-60.

31. Bruneau N, Roux S, Garreau B, Lelord G: **Frontal auditory evoked potentials and augmenting-reducing.** *Electroencephalogr Clin Neurophysiol* 1985, **62:**364-371.
32. Jacobson GP, Lombardi DM, Gibbens ND, Ahmad BK, Newman CW: **The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1.** *Ear and Hearing* 1992, **13:**300-306.
33. Näätänen R, Picton T: **The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure.** *Psychophysiology* 1987, **24:**375-425.
34. Shahin A, Roberts L, Miller L, McDonald K, Alain C: **Sensitivity of EEG and MEG to the N1 and P2 Auditory Evoked Responses Modulated by Spectral Complexity of Sounds.** *Brain Topography* 2007, **20:**55-61.
35. Rogier O, Roux S, Barthélémy C, Bruneau N: **Specific temporal response to human voice in young children.** In *10th International Conference on Cognitive Neuroscience; Bodrum, Turkey* Frontiers in Human Neuroscience; 2008.
36. Rogier O, Roux S, Barthélémy C, Bruneau N: **Electrophysiological correlates of voice processing in young children.** *International Journal of Psychophysiology* 2008, **69:**274-275.
37. Kirmse U, Jacobsen T, Schröger E: **Familiarity affects environmental sound processing outside the focus of attention: An event-related potential study.** *Clinical Neurophysiology* 2009, **120:**887-896.
38. Dien J: **Issues in the application of the average reference: Review, critiques, and recommendations.** *Behavior Research Methods, Instruments & Computers* 1998, **30:**34-43.
39. Kodera K, Hink RF, Yamada O, Suzuki JI: **Effects of rise time on simultaneously recorded auditory-evoked potentials from the early, middle and late ranges.** *Audiology* 1979, **18:**395-402.
40. Näätänen R: **The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function.** *Behavioral and Brain Sciences* 1990, **13:**201-288.
41. Woods DL: **The component structure of the N1 wave of the human auditory evoked potential.** *Electroencephalogr Clin Neurophysiol Suppl.* 1995, **44:**102-109.
42. Kaas JH, Hackett TA: **Subdivisions of auditory cortex and processing streams in primates.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97:**11793-11799.
43. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS: **A Cortical Region Consisting Entirely of Face-Selective Cells.** *Science* 2006, **311:**670-674.
44. Bentin S, Allison T, Puce A, Perez E, McCarthy G: **Electrophysiological studies of face perception in humans.** *Journal of Cognitive Neuroscience* 1996, **8:**551-565.
45. Itier RJ, Taylor MJ: **N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs.** *Cereb Cortex* 2004, **14:**132-142.
46. Rousselet GA, Husk JS, Bennett PJ, Sekuler AB: **Time course and robustness of ERP object and face differences.** *J Vis.* 2008, **8(12):**1-18.
47. Belin P, Fecteau S, Bedard C: **Thinking the voice: neural correlates of voice perception.** *Trends in Cognitive Sciences* 2004, **8:**129-135.
48. Campanella S, Belin P: **Integrating face and voice in person perception.** *Trends in Cognitive Sciences* 2007, **11:**535-543.
49. Bentin S, Taylor MJ, Rousselet GA, Itier RJ, Caldara R, Schyns PG, Jacques C, Rossion B: **Controlling interstimulus perceptual variance does not abolish N170 face sensitivity.** *Nat Neurosci* 2007, **10:**801-802.
50. Belin P, Zatorre RJ, Ahad P: **Human temporal-lobe response to vocal sounds.** *Brain Res Cogn Brain Res* 2002, **13:**17-26.
51. Bötzel K, Schulze S, Stodieck SR: **Scalp topography and analysis of intracranial sources of face-evoked potentials.** *Exp Brain Res.* 1995, **104(1):**135-143.
52. Carmel D, Bentin S: **Domain specificity versus expertise: factors influencing distinct processing of faces.** *Cognition* 2002, **83:**1-29.
53. McCarthy G, Puce A, Belger A, Allison T: **Electrophysiological Studies of Human Face Perception. II: Response Properties of Face-specific Potentials Generated in Occipitotemporal Cortex.** *Cereb Cortex* 1999, **9:**431-444.
54. Goffaux V, Gauthier I, Rossion B: **Spatial scale contribution to early visual differences between face and object processing.** *Cognitive Brain Research* 2003, **16:**416-424.
55. Johnson JS, Olshausen BA: **Timecourse of neural signatures of object recognition.** *Journal of Vision* 2003, **3:**499-512.
56. Rousselet GA, Husk JS, Bennett PJ, Sekuler AB: **Single-trial EEG dynamics of object and face visual processing.** *NeuroImage* 2007, **36:**843-862.
57. Rousselet GA, Mace MJ, Thorpe SJ, Fabre-Thorpe M: **Limits of Event-related Potential Differences in Tracking Object Processing Speed.** *Journal of Cognitive Neuroscience* 2007, **19:**1241-1258.
58. Van Rullen R, Thorpe SJ: **The Time Course of Visual Processing: From Early Perception to Decision-Making.** *J Cogn Neurosci* 2001, **13:**454-461.
59. Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A: **Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing.** *Cereb Cortex* 2004, **14:**247-255.
60. Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A: **Representation of the Temporal Envelope of Sounds in the Human Brain.** *J Neurophysiol* 2000, **84:**1588-1598.
61. Pernet C, Schyns PG, Demonet JF: **Specific, selective or preferential: Comments on category specificity in neuroimaging.** *NeuroImage* 2007, **35:**991-997.