Discriminating Male and Female Voices: Differentiating Pitch and Gender

Article *in* Brain Topography · November 2011

DOI: 10.1007/s10548-011-0207-9 · Source: PubMed

2 authors:

Marianne Latinus
French Institute of Health and Medical Research
83 PUBLICATIONS **3,503** CITATIONS
SEE PROFILE

Margot J. Taylor
Hospital for SIck Children; University of Toronto
432 PUBLICATIONS **23,169** CITATIONS
SEE PROFILE

Brain Topogr
DOI 10.1007/s10548-011-0207-9

ORIGINAL PAPER

Discriminating Male and Female Voices: Differentiating Pitch and Gender

Marianne Latinus · Margot J. Taylor

Received: 19 July 2011 / Accepted: 24 October 2011

Abstract Gender is salient, socially critical information obtained from faces and voices, yet the brain processes underlying gender discrimination have not been well studied. We investigated neural correlates of gender processing of voices in two ERP studies. In the first, ERP differences were seen between female and male voices starting at 87 ms, in both spatial–temporal and peak analyses, particularly the fronto-central N1 and P2. As pitch differences may drive gender differences, the second study used normal, high- and low-pitch voices. The results of these studies suggested that differences in pitch produced early effects (27–63 ms). Gender effects were seen on N1 (120 ms) with implicit pitch processing (study 1), but were not seen with manipulations of pitch (study 2), demonstrating that N1 was modulated by attention. P2 (between 170 and 230 ms) discriminated male from female voices, independent of pitch. Thus, these data show that there are two stages in voice gender processing; a very early pitch or frequency discrimination and a later more accurate determination of gender at the P2 latency.

Keywords ERPs · Voice · Auditory · Frequency · N1 · P2

M. Latinus
Faculté de Médecine de Rangueil, CerCo, Université Toulouse 3-CNRS, 31062 Toulouse, France

M. Latinus (✉)
Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, United Kingdom
e-mail: marianne.latinus@glasgow.ac.uk

M. J. Taylor
Diagnostic Imaging and Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G1X8, Canada

Introduction

Typically adults can easily and accurately extract gender from acoustical information in voices. In particular, the perception of voice gender primarily relies on the fundamental frequency [f0—(Lavner et al. 2000; Mullennix et al. 1995)] that is on average higher by an octave in female than male voices; yet, pitch overlaps considerably between male and female voices (Hillenbrand et al. 1995). Using behavioural adaptation paradigms, however, Schweinberger and colleagues (2008) established that the representation of voice gender was relatively independent of low-level acoustic information, as aftereffects were abolished with sinusoidal tones matched for fundamental frequency (Schweinberger et al. 2008). This demonstrated that, although voice pitch and gender are linked, other information is used to recognise an individual's gender from his/her voice. Other acoustic parameters that may contribute to gender identification include formant frequencies that reflect the filtering action of the vocal tract on voice production (Ghazanfar and Rendall 2008; Latinus and Belin 2011) and which are also lower in male voices (Andrews and Schmidt 1997; Whiteside 1998). Studies have demonstrated that gender recognition also relies on temporal information (Fu et al. 2004), as temporal inversion of voices decreases gender recognition (Bedard and Belin 2004). These studies demonstrate that although voice gender seems to be primarily discerned through voice pitch, other factors can be used to accurately recognise gender, indicating that the perception of pitch and gender can be dissociated (Lattner et al. 2005).

Functional magnetic resonance imaging (fMRI) studies have highlighted regions located along the superior temporal sulcus (STS) responsible for processing of voices, for both linguistic and extra-linguistic information in humans

Published online: 12 November 2011

(Belin et al. 2000; Scott et al. 2000; von Kriegstein et al. 2003) and macaques (Petkov et al. 2008). The processing of extra-linguistic aspects of voices engaged primarily the anterior STS—the temporal pole—of the right hemisphere, as only this region discriminated vocal from non-vocal sounds in the absence of speech information (Belin et al. 2002; von Kriegstein et al. 2003). Investigation of gender processing of voices with fMRI has produced inconsistent results (Sokhi et al. 2005; Lattner et al. 2005). Lattner et al. (2005) showed that female voices produced stronger bilateral response than male voices, with a right hemisphere dominance, especially in the superior temporal gyrus (STG), whereas Sokhi et al. (2005) reported that female voice processing involved the STG while male voices produced a larger response in the right precuneus. Lattner et al. (2005) also investigated pitch perception regardless of voice gender, and showed that voice pitch involved a network of regions localised closed to Heschl's gyrus. They showed that high-pitch voices activated a neural network similar to female voices whereas low-pitch voices induced a larger activity in the left anterior cingulate gyrus; pitch judgement involved the right prefrontal cortex.

The time course of neural voice processing of voice characteristics is not well understood, and the literature, again, reports inconsistent results. An event-related potential (ERP) study highlighted a voice-specific response (VSR) 320 ms after stimulus onset (Levy et al. 2001); the VSR, a frontal positive deflection larger for vocal than non-vocal stimuli, was found to be attention-dependent as the difference between vocal and non-vocal sounds disappeared when auditory stimuli were unattended (Levy et al. 2003; Gunji et al. 2003). Since that first study, others have reported an earlier signature of voice processing (Charest et al. 2009; De Lucia et al. 2010; Rogier et al. 2010). While Murray and colleagues (2006) reported very early effects in a living/non-living categorisation, Charest et al. (2009) using a range of vocal stimuli, including non-speech vocalisation and animal vocalisation (bird cries), showed a preferential response to voices starting at 120 ms after stimulus onset and peaking at 200 ms, i.e. in the latency range of the auditory P2 component. Modulations of the auditory complex at the P2 latency have been described using complex stimuli: it was modulated by speech (Tiitinen et al. 1999) and by identity priming in a voice recognition paradigm (Schweinberger 2001). Studies that investigated the time course of voice gender processing are sparse, with one study showing a modulation of the N1/P2 complex following adaptation with gender-congruent vocal adaptors (Zaske et al. 2009). The amplitude of the N1 was reduced for male voices following adaptation to male voices, while the P2 to female voices was reduced after adaptation with female voices. Thus, there is little information on gender discrimination of voices, and none that have determined the spatial–temporal brain patterns that index this critical human skill.

In the present studies, we explored the time course of voice gender processing using ERPs. We proposed the following hypotheses: (1) the perception of pitch and gender are linked but can be dissociated (Lattner et al. 2005); (2) pitch processing occurs earlier than gender perception as suggested by studies demonstrating modulation of early auditory ERPs (P50) by sound frequency (Liegeois-Chauvel et al. 1994; Pantev et al. 1995); (3) neural activity sensitive to gender would be seen as greater activation to female voices, over right anterior sites (Lattner et al. 2005; Zaske et al. 2009). To address these hypotheses, we measured the neural activity related to gender categorisation of voices and the role of pitch in gender discrimination using ERPs. Participants performed gender categorisation on audio clips of voices. The role of fundamental frequency, perceived as pitch, in gender categorisation of voices was determined in a second study, using low- and high-pitched voices as well as normal voices. To avoid redundancy with the overlapping issues in the two studies, we present the methods and results of the two studies followed by one general discussion.

Materials and Methods

Subjects

Nineteen English-speaking adults (9 females) aged between 20 and 35 years (mean = 26.4 years), participated in the study. None of the subjects reported any hearing problems. They all gave informed written consent and the study was approved by the Sunnybrook Health Sciences Research Ethics Board.

Stimuli

Auditory stimuli were 14 monosyllabic French words spoken by six different speakers (3 females and 3 males) recorded using CoolEdit Pro (stereo; 22.05 kHz; 32 bits). The speakers also spoke the words using high- and low-pitched voices; speakers were instructed to speak the words, making their natural pitch a higher or lower frequency, but not forcing their voices—keeping them as natural-sounding as possible, while making clearly audible changes in the pitch. All speakers were able to do so. Thus, there were 252 stimuli divided into six categories (42 stimuli per condition): female high-pitch voices (HF), female low-pitch voices (LF), female normal voices (NF) and male high-pitch voices (HM), male low-pitch voices (LM), male normal voices (NM). As we were not interested in semantic processing, the experiments were run on native, mono-lingual English speaking adults with French words. The stimuli were filtered using a high-pass filter (20 Hz) to remove low-frequencies not related to the stimuli. All the stimuli used in the experiment were normalized for energy (root mean square) using Matlab (The MathWorks, Inc.). The duration of the stimuli was on average 263 ms. To prevent the perceptual effect of clicks at onset and offset, an envelope of 10 ms rise and fall times was applied to all stimuli. In the first experiment, only stimuli in the speakers' normal voice were used; there were 84 vocal stimuli, 42 by female and 42 by male voices. In the second experiment, the full series of 252 auditory stimuli was used; there were 42 in each of the 6 categories. All of the subjects completed the experiment with normal voices first and then the study with the pitch-altered voices.

Sound Analysis

Several acoustical parameters of the vocal stimuli were measured: mean pitch (f0), f0 range (difference between the minimum and the maximum of f0 for each stimulus) and formant frequencies (F1 to F4) plus sound duration and words' start time. These parameters were measured using Praat (Boersma and Weenick 2001) and mean values are shown in Table 1. Two repeated measures ANOVAs were run: the first on the normal voices only, and the second one on all six categories. Voice gender was a between-subject factor, while word was a repeated factor (14 levels), when all six categories were included pitch was also a repeated factor with 3 levels.

When comparing only the normal voices, an effect of gender was seen on the fundamental frequency (mean f0 − $F(1,4) = 56.34$, $P = 0.002$), which was higher for female than male voices and on the f0 range ($F(1,4) = 8.84$, $P = 0.041$), which was significantly larger for male voices (see Table 1). Words affected mean frequency of the first three formants ($F(13,52) = 5.85$, $P = 0.031$; $F(13,52) = 8.03$, $P = 0.015$; $F(13,52) = 4.98$, $P = 0.014$ for mean F1, F2, and F3, respectively), in line with previous reports (Hillenbrand et al. 1995). All other acoustical parameters, F4 frequency, sound duration and word start time were not affected by words or speakers' gender.

In experiment 2, analyses of f0 revealed an expected effect of gender ($F(1,4) = 12.61$, $P = 0.024$) and pitch ($F(2,8) = 42.67$, $P = 0.002$), with no interactions; female voices were on average higher pitched than male voices, and f0 was highest for high-pitch voices, while it was the lowest for low-pitch voices (Table 1). f0 range was still larger for male than female voices as shown by a speakers' gender effect ($F(1,4) = 8.58$, $P = 0.043$), yet it was not modulated by pitch. Formant analysis revealed that F1, F2, and F3 frequencies differed with words. All other acoustical parameters, F4 frequency, sound duration and word start time were not affected by words, speakers' gender or pitch.

Tasks and Design

Stimuli were presented binaurally via headphones at normal speaking levels (68 ± 5 dB); inter-stimulus intervals varied randomly between 1,300 and 1,600 ms. The presentation order of stimuli was randomised across participants. During the tasks, a central fixation cross was shown on a screen 60 cm in front of the subjects, who were asked to maintain central fixation and refrain from making eye movements. Participants pressed one key for male voices and another for female voices (counter-balanced across subjects); in both experiments, participants were instructed to respond as accurately and as quickly as possible. Instructions for the task in the second study informed the subjects that the pitch of the voices may be altered and, thus may not be a valid cue to discriminate gender.

ERP Recordings and Analyses

The experiments were run in a dimly lit sound-attenuating booth. ERPs were recorded using an ANT system and a 64 electrode cap, including three ocular electrodes to monitor

Table 1 Sound analysis: sound duration (ms), start time (ms), mean fundamental frequency (f0) and mean f0 range, and formants 1 to 4

	Female voices			Male voices		
	High-pitched	Normal	Low-pitched	High-pitched	Normal	Low-pitched
Sound duration	269.88 ± 16.5	263.33 ± 12.8	278.74 ± 4.5	242.79 ± 16.5	246.98 ± 12.5	273.55 ± 4.5
Start time	6.39 ± 1.72	10.08 ± 1.92	6.39 ± 2.21	9.94 ± 2.21	6.73 ± 1.92	7.69 ± 1.72
f0	398.15 ± 13.6	211.11 ± 7.41	175.86 ± 2.51	324.37 ± 7.78	148.02 ± 9.18	140.91 ± 9.31
f0 range	71.66 ± 12.88	79.95 ± 18.86	19.92 ± 2.21	63.5 ± 8.01	121.57 ± 20.7	100.39 ± 23.1
F1	668 ± 41.86	611 ± 31.65	524 ± 26.44	550 ± 27.46	638 ± 37.39	660 ± 59.98
F2	1696 ± 71.6	1768 ± 78.44	1673 ± 91.65	1577 ± 71.42	1726 ± 62.12	1768 ± 79.02
F3	2967 ± 40.23	2923 ± 39.92	2907 ± 50.35	2855 ± 39.81	2932 ± 37.6	2849 ± 46.67
F4	3988 ± 36.14	4094 ± 37.07	4012 ± 35.43	3913 ± 58.49	3942 ± 44.82	3770 ± 49.42

f0 range and all frequencies are expressed in Hz. Mean values per categories ± standard error of the mean vertical and horizontal eye movements. Impedances were kept below 5 kΩ. The sampling acquisition rate was 1024 Hz. FCz was the reference during acquisition; an average reference was calculated off-line. Continuous EEG was epoched into 600 ms sweeps including 100 ms pre-stimulus, used as baseline. Trials containing ocular and muscle artefacts or an amplitude shift greater than 100 µV, were rejected from analyses as well as trials including an incorrect answer. Epochs were averaged by condition and filtered using a bandpass filter 1–30 Hz; subject's data were retained for analysis if the average ERPs included at least 20 epochs for each condition. The average number of trials (± standard deviation) per conditions to create the averages was, in experiment 1, 35 ± 5 for female voices, 37 ± 5 for male voices. In experiment 2, the average number of trials per conditions were 36 ± 6, 36 ± 7, 31 ± 6, 34 ± 6, 33 ± 5, 37 ± 5. There was no significant difference in the number of trials across the six conditions.

A preliminary two-way ANOVA was completed on the ERP data from both studies to assess interactions between subject's gender and brain activity for the different conditions. These analyses revealed differences between ERPs for female and male subjects; females had consistently larger amplitude responses than males. However, no interactions on the ERPs between subjects' gender and gender categorisation were observed. Consequently ERP analyses were collapsed across the gender of the participants.

Peak analyses were performed on the individual average ERP, for each condition, on peaks classically described in auditory ERP literature: N1—negative peak around 100 ms, P2—positive deflection around 200 ms (Näätänen et al. 1988) and the VSR, a positive peak around 340 ms (Levy et al. 2001). Individual subject's peaks were measured in a ±30 ms time-window centred at the grand average latency and at the electrode sites where the component was maximal (Picton et al. 2000). N1 was measured in a time-window centred at 118 ms, at CP1/CP2, C1/C2, FC1/FC2 and Cz. The time-window for P2 was centred at 215 ms, and measured at electrodes F1/F2, F3/F4, FC1/FC2 and FC3/FC4. VSR was measured at FC1/FC2, F1/F2, F3/F4, AF3/AF4 and Fz centred at 350 ms.

Statistical Analyses

Behavioural data and peak latencies and amplitudes were submitted to repeated measures analyses of variance (using SPSS11). In the first experiment, the within-subjects factor was voice gender (2 levels); in the second experiment, the within-subjects factors were voice gender (2 levels) and pitch (3 levels). For peak analyses, the supplementary within-subjects factors were hemisphere (2 levels) for latency measures plus electrode (different levels depending on the component) for amplitude measures. For significant interactions we performed paired comparison and post-hoc tests to determine the factors leading to the effects.

Spatial–temporal effects were assessed by comparing brain activity for the different conditions, at each time point and electrode. These analyses determine when brain activity differs significantly between conditions, allowing ERP differences to be identified independently of peak measures (Giard and Peronnet 1999; Fort et al. 2002; Charest et al. 2009; Latinus et al. 2010). Repeated measures ANOVA within the general linear model framework were run on the ERPs using Matlab7.2 with task and stimulus as inter-subject factors at each time point and electrode. To estimate the statistical significance of the ANOVA, we calculated a data-driven distribution of F-values using a bootstrap-F method; this method makes no assumption on the normality of the data distribution and is therefore robust to normality violations (Berkovits et al. 2000; Wilcox 2005; Rousselet et al. 2009). The data were centred at 0 to be under the null hypothesis that conditions do not differ from 0. ANOVAs at each time point and electrode were run on the centred data after resampling the subjects with replacement. We stored the bootstrapped F-values for each time point and electrode independently. This operation was repeated 999 times to obtain a distribution of 1000 bootstrapped estimates of F-values under the null hypothesis (Berkovits et al. 2000). Corrections for multiple comparisons were performed using the spatial–temporal clustering methods described in Pernet et al. (2011). The 2D clustering was performed using the function implemented in the LIMO EEG toolbox (Pernet et al. 2011), derived from the clustering technique developed in Fieldtrip (http://fieldtrip.fcdonders.nl). A spatial–temporal cluster is considered significant, if the sum of F-values within the cluster is superior to the threshold bootstrap cluster sum. For each bootstrap, we calculated the maximal sum of F-values, which allowed us to build a distribution of the spatial–temporal cluster values, and therefore extract the threshold value used to assess the significance of an observed cluster value (Pernet et al. 2011). Post-hoc tests, paired comparisons for the three pitch conditions (collapsed for gender), were run for the Pitch factor whenever the ANOVA was significant. Data-driven confidence intervals were calculated for each comparison (HIGH vs. LOW, HIGH vs. NORMAL and LOW vs. NORMAL), using the same protocol as described above, with bootstrapping, random sampling and confidence interval estimation.

Results

Study 1: Gender Categorisation of Voices

Subjects performed well on gender discrimination of voices; percent correct was at 95% regardless of voice gender.

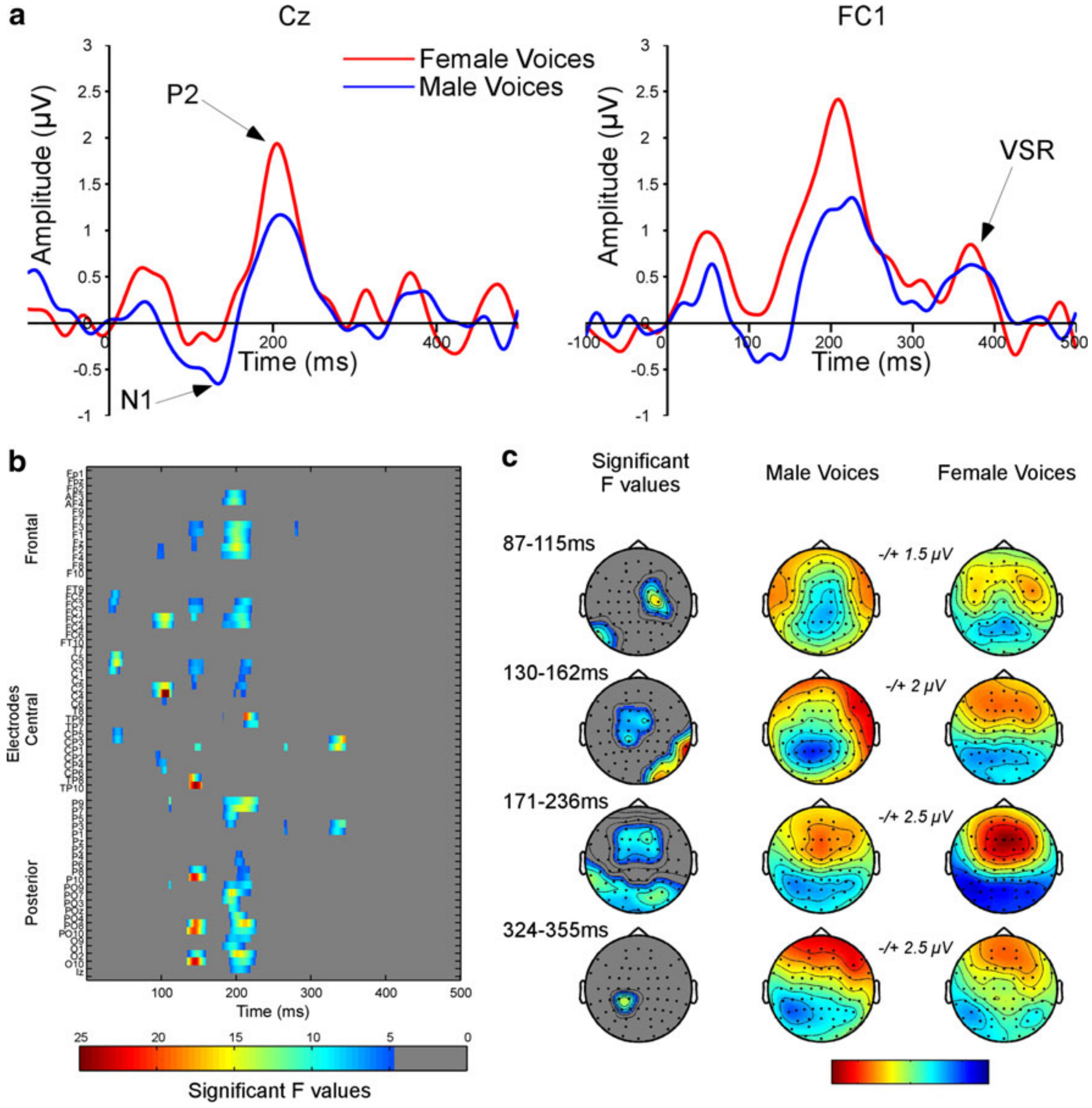


Fig. 1 Grand average ERPs and topographies for voices from study 1. **a** ERPs to normal female (*red line*) and male (*blue line*) voices averaged over electrodes Cz and FC1. **b** Results of the repeated measures ANOVA. Representation of the significant *F*-values ($F(1,18) > 4.41$) after correction for multiple comparisons. **c** Topographies for each time interval where the ANOVAs were significant. First column: significant *F*-values of the difference between male and female voices second column: male voices, third column: female voices. *NB* For the *F*-value topographies, non-significant values are in *grey*

RTs were longer for female (719 ± 21 ms) than for male (690 ± 24 ms) voices ($F(1,18) = 6.75$, $P = 0.018$).

Latencies of neither N1 nor P2 were affected by voice gender. N1 amplitude was larger for male than female voices ($F(1,18) = 18.35$, $P < 0.001$; Fig. 1a), whereas P2 was larger ($F(1,18) = 8.56$, $P < 0.009$) for female than male voices (2.9 vs. 2.0 μV, respectively) (Fig. 1a). VSR latency and amplitude were not modulated by voice gender (Fig. 1a).

Spatial–temporal analyses of brain activity revealed differences between the processing of female and male voices starting at 87 ms (Fig. 1b). Between 87 and 115 ms, the topographies for male and female voices were dissimilar, reaching significance in central and posterior regions. This was due to greater negativity to male than female voices (Fig. 1c). Topographic differences in this latency range and the N1 modulation by voice gender (a larger N1 to male voices) suggest that the N1 observed for male voices may arise from different brain sources than that to female voices. In the time window 130–162 ms, the topographies observed for female and male voices again differed. As seen in Fig. 1c (2nd row), the topography to male voices was more similar to that of N1 (Fig. 1c—1st row) whereas the topography for female voices more similar to the topography of the P2 (Fig. 1c—3rd row). This may reflect an earlier onset of P2 from N1 for female voices (P2 latency: female voices: 212 ms, male voices 218 ms, $F(1,18) = 2.62$, $P = 0.123$). Between 171 and 236 ms, robust differences between female and male voices demonstrated overall larger amplitudes for female voices than male voices consistent with the peak analysis (Fig. 1). Late modulations of the brain potentials, in the latency range of the VSR, reached significance over parieto-central regions (Fig. 1c—4th row); they reflected a larger activity (positive in frontal regions and negative in parietal regions) to male than female voices.

Study 2: Does Voice Gender Categorisation Rely on Pitch?

Although in experiment 1 several disparities were observed in the ERPs to female and male voices, the source of these differences is difficult to assess as female and male voices differed significantly in fundamental frequency (by 63 Hz in the present study) and in f0 range (by 41 Hz). Fundamental frequency, one of the most obvious cues to discriminate gender (Lavner et al. 2000; Mullennix et al. 1995), is known to modulate early ERP components (Liegeois-Chauvel et al. 1994). Consequently, the observed differences could stem either from pitch perception, i.e. low-level processing, or from an effective gender categorisation, i.e. a more abstract representation. We hypothesised that early ERP modulations reflected pitch processing while later effects (at the P2 latency) may indicate an abstract processing of voice gender (Zaske et al. 2009). Thus, we ran a second study including the pitch-altered voices (high-pitch and low-pitch voices) in addition to the original voices, to distinguish between pitch and gender. Subjects were instructed that voice pitch would be altered—i.e., that there may be unusual pitch/gender mapping, thus directing their attention away from pitch as the cue for gender.

Behavioural Results

Accuracy on normal-pitch voices was high (~96%) regardless of gender. Recognition of male voices was more accurate than female voices ($F(1,17) = 20.29$, $P < 0.001$), due to the pitch-altered voices (i.e. high- and low-pitched voices) disrupting female more than male categorisation (pitch × voice gender: $F(2,34) = 14.04$, $P < 0.001$) (Table 2). Categorisation of high-pitch voices was the least accurate ($F(2,34) = 53.09$, $P < 0.001$), especially for female voices. Low-pitch voices were categorised as accurately as normal voices for male voices but not for female voices (see Table 2).

RTs were equivalent for female and male voices ($F(1,17) = 0.354$, n.s.) but were modulated by subject gender, as responses were faster to same-sex voices (subject × voice gender: $F(1,17) = 5.24$, $P = 0.035$—Table 2). RTs differed as a function of pitch and voice gender ($F(2,34) = 36.91$, $P < 0.001$): the fastest responses were for low-pitch voices categorised as male, whereas RTs to high-pitch voices were the fastest for female voices and slowest for male voices (see Table 2).

Behavioural results suggest that pitch is an important factor of gender perception as coherent pitch information helps verify the recognition of voice gender; yet it is not the only factor as gender recognition was still performed accurately in the absence of typical or reliable pitch information.

Peak Analyses

In this study, N1 showed no sensitivity to pitch or voice gender in latency or amplitude. P2 was earlier ($F(1,18) = 21.25$, $P < 0.001$) and larger ($F(1,18) = 14.95$, $P = 0.001$) for female compared to male voices (Fig. 2a). A general effect of pitch was observed on P2 latency ($F(2,36) = 6.33$, $P = 0.007$) and amplitude ($F(2,36) = 7.69$, $P = 0.003$), driven by the P2 to high-pitch voices having the longest latency and largest amplitude (Fig. 2). Voice-selective response (VSR) was delayed

Table 2 Mean percentage of correct answers (hits) and reaction times for study 2 for each condition

	Female voices			Male voices		
	High-pitched	Normal	Low-pitched	High-pitched	Normal	Low-pitched
Hits ± sem	76.80 ± 2.34	96.11 ± 0.88	84.64 ± 1.67	87.78 ± 1.32	96.55 ± 1.99	97.91 ± 1.03
RTs ± sem						
Female Ss	746 ± 37	792 ± 46	775 ± 49	879 ± 47	749 ± 42	736 ± 44
Male Ss	802 ± 35	828 ± 43	839 ± 47	850 ± 45	771 ± 40	760 ± 42

NB Percentage of correctly identified stimuli was particularly low for high-pitched voices, especially for female voices. RTs to female high-pitched voices were the fastest, whereas for male voices, low-pitched voices induced the fastest response by 5 ms for male voices reflecting the delay observed at the P2 latency, shown by a peak-to-peak analysis. VSR was earlier over the left hemisphere for high-pitch voices, and over the right hemisphere for low-pitch voices, while no lateralisation was observed for normal-pitch voices (frequency × hemisphere: $F(2,36) = 5.94$, $P = 0.006$). VSR amplitude was not affected by pitch or gender.

Spatial–Temporal Analyses

The spatial–temporal ANOVAs suggested that the processing of pitch and voice gender were independent, as no interactions were seen between them at any time points (Fig. 2b). While in the first experiment, significant differences between male and female voices were observed as early as 87 ms, in this second study, the earliest reliable differences between conditions, attributable to gender, were found at the P2 latency, i.e. between 157 and 232 ms. These differences were driven by a larger activity to female voices than to male voices over central and right temporo-occipital regions (Fig. 2b, c—collapsed data for pitch). A major difference between male and female voices is fundamental frequency (f0) that is on average higher for female than male voices (see methods). Thus, in order to determine if the P2 difference was attributable to gender categorisation, a one-way ANOVA was run on brain topography between high-pitch male voices (320 Hz) and normal female voices (195 Hz). This comparison showed a similar pattern of results as the comparison of brain topography evoked by female and male voices, collapsed for pitch, despite the reversal of the usual pitch/gender pairing (Fig. 2d). Consequently, although in this particular case pitch was higher for male than female voices, the topography of the significant differences was similar to the gender comparison with a fundamental frequency higher for female than for male voices. These results reinforce that the neural activity subtending P2 may be a neural correlate for gender discrimination of voices.

Pitch categories affected brain potentials at several spatial–temporal clusters starting as early as 27 ms (Fig. 2b), producing results similar to the gender categorisation observed in experiment 1. Between 27 and 63 ms, significant differences were seen in left anterior and right posterior regions. These early significant differences were due to high-pitch voices inducing larger amplitudes than low-pitch and normal voices, shown by post-hoc analysis (Fig. 3, 1st row). High-pitch voices showed no hemispheric asymmetry, whereas the pattern of activity for low-pitch and normal voices showed right-frontal lateralisation (Fig. 3, 1st row). Pitch did not affect the ERPs at the peak of the P2 component, but slightly after between 212 and 272 ms over left fronto-central and posterior electrodes; these differences were due to a larger absolute response to pitch-altered voices compared to normal voices (Fig. 3, 2nd row). Pitch effects within the VSR latency range (307–351 ms) were due to less activation for low-pitch voices compared to high-pitch and normal voices (Fig. 3, 3rd row).

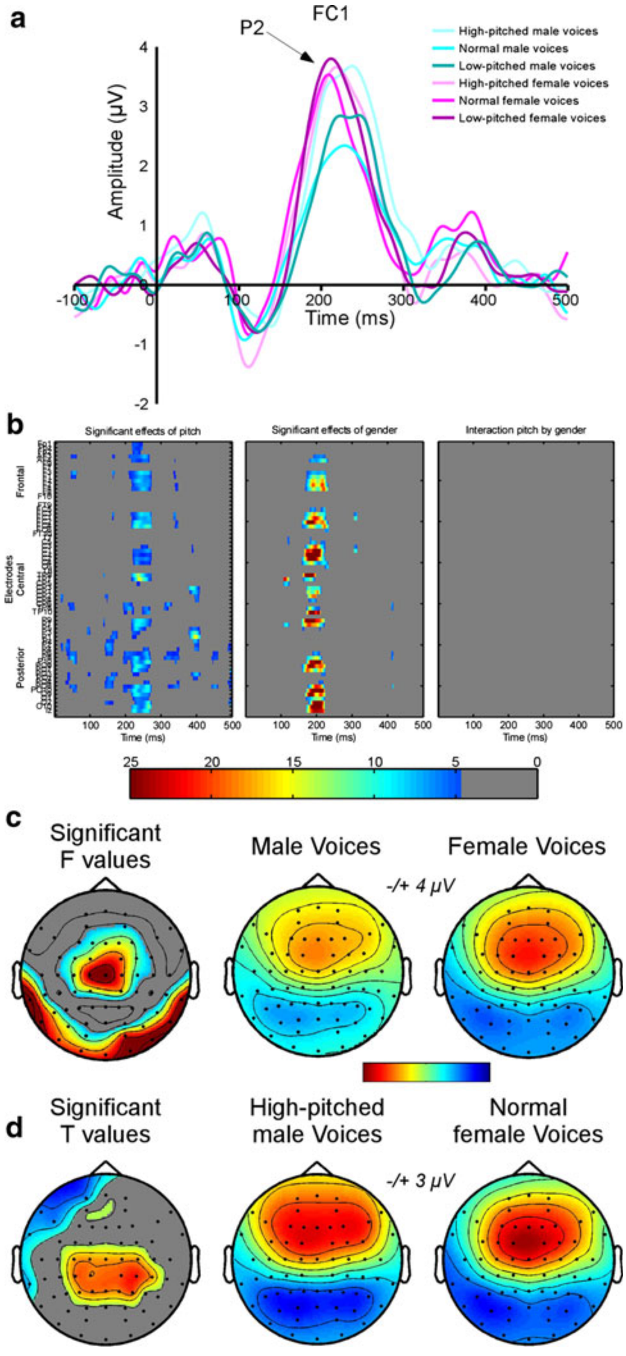
Discussion

We report two studies investigating the neural correlates of voice gender perception. In the first study, participants listened to female and male voices, while performing a gender categorisation; in the second study, pitch-altered voice stimuli were included to dissociate pitch processing from higher-level gender representation processing. These two studies revealed significant differences between the processing of female and male voices, both behaviourally and neurophysiologically.

At the behavioural level, gender differences were seen in accuracy only for the pitch-altered voices with pitch modifications having a greater influence on female than male categorisation. In the first study, RTs were longer for female than male voices. In study 2, participants categorised same-sex voices faster than opposite sex voices, as reported with faces (Yamaguchi et al. 1995; Cellerino et al. 2004). Moreover, participants categorised the voices with the most typical pitch fastest, i.e., high-pitch for female voices and low-pitch for male voices. These results showed that a person's gender is in part derived from fundamental frequency (pitch), and provide behavioural evidence that high f0 are typical of a feminine voice and, vice versa (Whiteside 1998; Murry and Singh 1980; Mullennix et al. 1995). Yet other parameters, such as temporal information or formant frequency, are clearly used to perform voice gender categorisation as it remains reliable even in the absence of a customary f0 (Schweinberger et al. 2008; Fu et al. 2004). The f0 range could be a cue for gender that remains even after the pitch alteration, as it was larger for male voices regardless of pitch. Our results also demonstrated differences in the processing of female and male voices, as pitch modification seems to disrupt female more than male voice categorisation. This could seem at odds with results showing that that female voice perception relies more on temporal information than does the recognition of male voices (Murry and Singh 1980), yet, the stimuli used in our two experiments were short, and temporal information may have been reduced.

Analyses of the spatial–temporal brain patterns were critical in revealing differences in the processing of male and female voices; while study 1 revealed differences at several spatial–temporal clusters, study 2 clarified that only the effects within the P2 latency range were specifically related to gender perception, in accordance with Zaske

Fig. 2 Peak and spatial–temporal analyses for study 2. **a** Grand▶ average ERPs for each condition at FC1 (P2 is represented). Female voices are in *pink*, male voices are in *blue*. The *lighter* to *darker lines* represent higher pitch to lower pitch voices. Note the delayed and smaller P2 for male voices. **b** Results of the bootstrapped repeated measures ANOVA for the 2 factors (pitch and gender) and their interaction. The scale represents *F*-values, when the 2-way ANOVA was significant (gender: $F(1,18) > 4.4155$; pitch and interaction: $F(2,36) > 3.26$) after correction for multiple comparisons. Non-significant *F*-values are presented in *grey*. **c** Topographies of the significant *F*-values between 157 and 232 ms (left) and for male and female voices. **d** Comparison between activity to male high-pitch voices and female normal voices, and maps of the significant *T* values of the paired *T* test. Note that the female voices evoked a larger activity than male voices with a similar distribution of effects as seen in the gender results (**c**), even though in this comparison the female voices had an f0 125 Hz lower than the male voices



et al. (2009). Early ERP differences, starting at 30 ms post-stimulus onset in study 2 and at 87 ms in study 1, were attributable to pitch processing, but not gender processing per se. It has previously been demonstrated that the Pa or P50, a positive potential occurring in this latency range, was sensitive to stimulus frequency and its topography reflects changes in dipole orientation with increasing frequency (Liegeois-Chauvel et al. 1994; Pantev et al. 1995). This change in topography has been proposed to reflect the tonotopy of the primary auditory cortex (Pantev et al. 1995). Thus, in the present studies, early effects mostly seen in topography changes due to pitch, likely reflect frequency processing differences in the auditory cortex between high and normal-to-low pitch voices.

In study 1, male voices evoked a larger N1 than female voices and this modulation by voice gender was also evident in topographical differences. In study 2, however, N1 was not affected by pitch or gender, nor was any spatial difference observed within this latency range. N1 reflects the processing of physical and temporal aspects of auditory stimuli (Näätänen and Picton 1987) including frequency (Näätänen et al. 1988; Zaske et al. 2009). Amplitude of the auditory N1 has been shown to be sensitive to the physical similarity between stimuli in adaptation designs (Zaske et al. 2009). N1 latency and amplitude decrease with increasing frequency using pure tone stimuli, especially for unattended tones (Crottaz-Herbette and Ragot 2000; Jacobson et al. 1992; Näätänen and Picton 1987; Alho et al. 1994) consistent with the results of our study 1. It has also been shown that selective attention influences the N1 component (Neelon et al. 2006), and that attention to pitch masks the N1 modulation by frequency (Alho et al. 1994). This suggests that the smaller N1 for female voices seen in study 1 corresponds to automatic pitch processing; this was not observed in study 2 due to attention being directed away from pitch as it was not predictive and subjects were informed that pitch had been modified. This difference between the two studies is consistent with classic studies showing that the auditory N1 is very sensitive to attention effects (Näätänen and Picton 1987; Alho et al. 1994).

Female voices evoked an earlier and/or larger P2 than male voices in both studies: between 170 and 230 ms differences were observed over fronto-central brain areas that encompass the P2 component. An earlier P2 to female voices was reported in a previous study (Zaske et al. 2009) and was proposed to reflect higher fundamental frequencies in female voices. Our results are in contradiction with this hypothesis, as high-pitched voices, with the highest

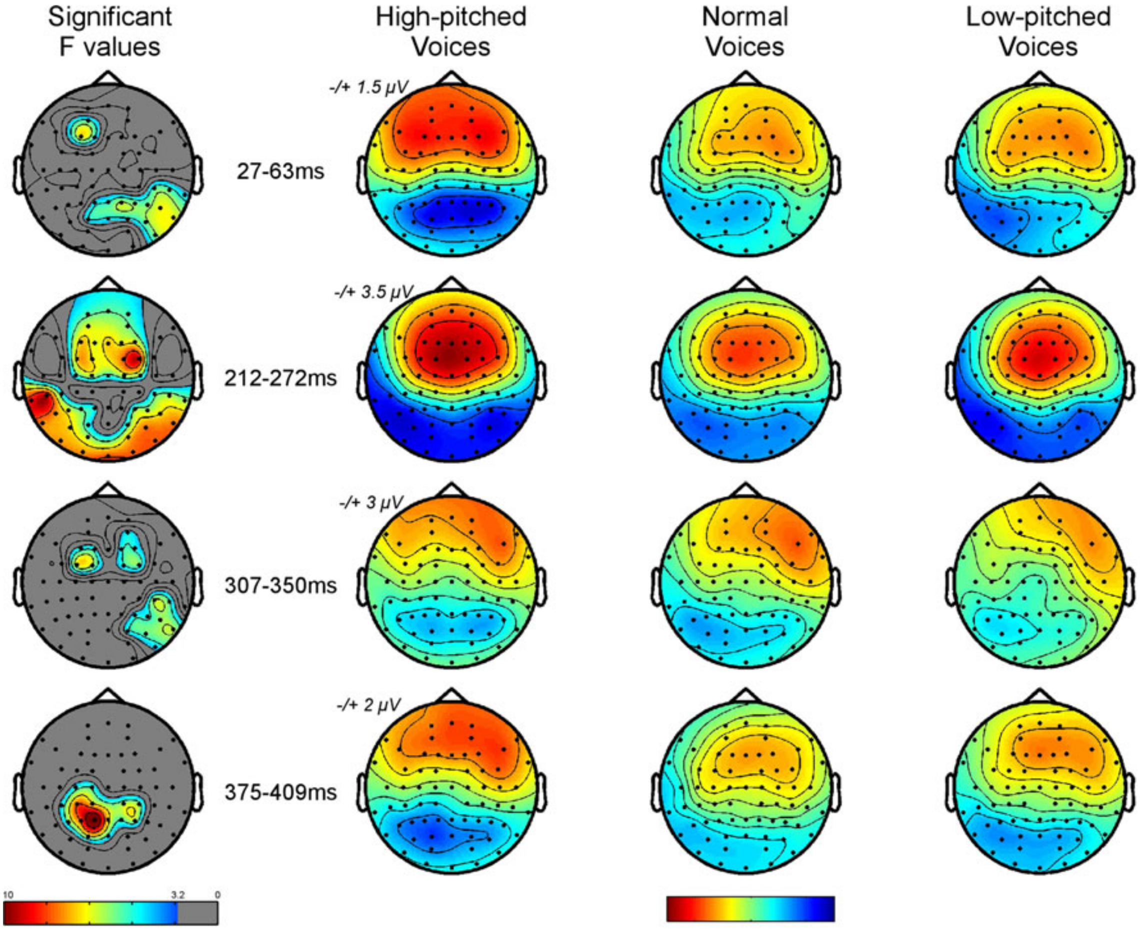


Fig. 3 Modulation of brain topographies by the pitch factor. First column: topographies of the significant *F*-values ($F(2,36) > 3.26$). Second to fourth columns: high-pitched, normal and low-pitched voices, respectively, collapsed for voice gender

fundamental frequency, evoked the latest P2. Yet, Zaske and colleagues (2009) also proposed that P2 could index a perceptual rather than a physical processing of female gender. In our studies, comparison between brain topographies to high-pitch male voices and normal female voices showed that male voices, even with a higher f0, yielded smaller responses than female voices at the same location and latency range. Thus, the combined results of the two studies suggest that neural mechanisms that underlie P2 are involved in a voice gender representation relatively abstracted from low-level, i.e. pitch, information. The P2 component has also been shown to index voice processing (Charest et al. 2009; Rogier et al. 2010; Lattner et al. 2003), as its amplitude was higher for vocal than non-vocal sounds. Lattner et al. (2003) demonstrated that a violation of listeners' expectations led to a voice-specific brain response 200 ms after stimulus onset. It has been suggested that P2 is an index for speech processing, as P2 is larger to vowels than tones (Tiitinen et al. 1999) and is sensitive to voice priming in a voice recognition paradigm (Schweinberger 2001). These effects on P2 amplitude may reflect voice sensitivity rather than speech processing. Although the aim of our article was not to study neural correlates of voice detection, our results support the hypothesis that P2 may reflect general voice processing (Charest et al. 2009). Sokhi et al. (2005) reported that female voices activated the right anterior STG whereas male voices activated the precuneus. This was not evident in our study as topographies to male and female voices were comparable, suggesting that a common brain source is at the origin of the P2 for male and female voices. However, as fMRI data does not provide temporal information, the brain areas described by Sokhi et al. (2005) may well be activated at different latencies such that activation of the STG around 200 ms led to a larger P2 for female voices and the precuneus activation may occur later and drive differences we observed at the VSR latency in left posterior regions.

Both study 1 and 2 revealed a positive deflection around 320 ms that may have been the VSR, although we did not use a voice/non-voice comparison (Levy et al. 2001). It has been suggested that VSR indexes the discrimination of human voice stimuli (Levy et al. 2003), a discrimination that according to the current results does not require or include gender or voice frequency, as neither altered the VSR per se. Modulations of brain activity were observed within this latency range but not at the electrodes where the VSR was measured.

Although we found significant effects in these two studies, we acknowledge some limitations. First, it is noteworthy that only three voices per gender were used in the study, with fourteen items per voice. This is a low number of voices, although not uncommon for this type of research (for example, five speakers per gender in Zaske et al. 2009 and four speakers in Schweinberger et al. 2008). Future studies should not only include more recordings of different voices, but an interesting question would be to use voices across the age range to determine if the age of the speaker impacts the discrimination of the sex of the speaker. Second, it would be better to have more trials per average, to obtain even clearer discrimination of the spatial–temporal pattern. The risk of this would be habituation of the responses. The fact that we found significant effects, with a decent number of subjects and using robust statistics, consistent with and expanding upon other studies in the literature, does give us confidence that the findings are veridical.

Conclusion

In conclusion, these studies revealed that auditory ERPs index both pitch and gender processing of voices: pitch processing starts very early and is modulated by attention, particularly at the N1 latency, while gender discrimination occurs around 200 ms and is likely associated with other aspects of voice processing (Charest et al. 2009; Zaske et al. 2009). Thus, we propose that gender processing of voices has two stages. An early tonotopically-sensitive stage estimates the pitch of the incoming sound; this can be an effective estimate of voice gender. However, once pitch information is accounted for, it appears that differences at the P2 latency remain at fronto-central regions, suggesting that gender discrimination of voices takes place at this latency. We suggest that true voice gender processing occurs at the P2 latency while pitch processing, which could be a more rapid surrogate for gender processing, occurs much earlier.

Acknowledgments Marianne Latinus gratefully acknowledges the salary support from the Fondation pour la Recherche Médicale. We thank Dr. Nancy J. Lobaugh for her generosity in allowing us full access to her ERP lab, and the help provided with the studies by Dr. Lobaugh and Erin Gibson. The authors also want to thank Dr. Rufin VanRullen for constructive comments on the manuscript.

References

Alho K, Teder W, Lavikainen J, Naatanen R (1994) Strongly focused attention and auditory event-related potentials. Biol Psychol 38(1):73–90

Andrews ML, Schmidt CP (1997) Gender presentation: perceptual and acoustical analyses of voice. J Voice 11(3):307–313

Bedard C, Belin P (2004) A "voice inversion effect?". Brain Cogn 55(2):247–249

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403(6767):309–312

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13(1):17–26

Berkovits I, Hancock GR, Nevitt J (2000) Bootstrap resampling approaches for repeated measure designs: relative robustness to sphericity and normality violations. Educ Psychol Measur 60(6): 877–892

Boersma P, Weenick D (2001) Praat, a system for doing phonetics by computer. Glot Int 5(9/10):341–345

Cellerino A, Borghetti D, Sartucci F (2004) Sex differences in face gender recognition in humans. Brain Res Bull 63(6):443–449

Charest I, Pernet CR, Rousselet GA, Quinones I, Latinus M, Fillion-Bilodeau S, Chartrand JP, Belin P (2009) Electrophysiological evidence for an early processing of human voices. BMC Neurosci 10:127. doi:10.1186/1471-2202-10-127

Crottaz-Herbette S, Ragot R (2000) Perception of complex sounds: N1 latency codes pitch and topography codes spectra. Clin Neurophysiol 111(10):1759–1766

De Lucia M, Clarke S, Murray MM (2010) A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci 30(33):11210–11221. doi:10.1523/JNEUROSCI.2239-10.2010

Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14(1):20–30

Fu QJ, Chinchilla S, Galvin JJ (2004) The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. J Assoc Res Otolaryngol 5(3):253–260

Ghazanfar AA, Rendall D (2008) Evolution of human vocal production. Curr Biol 18(11):R457–R460

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11(5):473–490

Gunji A, Koyama S, Ishii R, Levy D, Okamoto H, Kakigi R, Pantev C (2003) Magnetoencephalographic study of the cortical activity elicited by human voice. Neurosci Lett 348(1):13–16

Hillenbrand J, Getty LA, Clark MJ, Wheeler K (1995) Acoustic characteristics of American english vowels. J Acoust Soc Am 97:3099–3111

Jacobson GP, Lombardi DM, Gibbens ND, Ahmad BK, Newman CW (1992) The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. Ear Hear 13(5):300–306

Latinus M, Belin P (2011) Human voice perception. Curr Biol 21(4):R143–R145. doi:10.1016/j.cub.2010.12.033

Latinus M, VanRullen R, Taylor MJ (2010) Top-down and bottom-up modulation in processing bimodal face/voice stimuli. BMC Neurosci 11:36. doi:10.1186/1471-2202-11-36

Lattner S, Maess B, Wang Y, Schauer M, Alter K, Friederici AD (2003) Dissociation of human and computer voices in the brain: evidence for a preattentive gestalt-like perception. Hum Brain Mapp 20(1):13–21

Lattner S, Meyer ME, Friederici AD (2005) Voice perception: sex, pitch, and the right hemisphere. Hum Brain Mapp 24(1):11–20

Lavner Y, Gath I, Rosenhouse J (2000) The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun 30:9–26

Levy DA, Granot R, Bentin S (2001) Processing specificity for human voice stimuli: electrophysiological evidence. Neuroreport 12(12): 2653–2657

Levy DA, Granot R, Bentin S (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40(2):291–305

Liegeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P (1994) Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol 92(3):204–214

Mullennix JW, Johnson KA, Topcu-Durgun M, Farnsworth LM (1995) The perceptual representation of voice gender. J Acoust Soc Am 98(6):3080–3095

Murry T, Singh S (1980) Multidimensional analysis of male and female voices. J Acoust Soc Am 68(5):1294–1300

Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24(4):375–425

Näätänen R, Sams M, Alho K, Paavilainen P, Reinikainen K, Sokolov EN (1988) Frequency and location specificity of the human vertex N1 wave. Electroencephalogr Clin Neurophysiol 69(6): 523–531

Neelon MF, Williams J, Garell PC (2006) The effects of auditory attention measured from human electrocorticograms. Clin Neurophysiol 117(3):504–521

Pantev C, Bertrand O, Eulitz C, Verkindt C, Hampson S, Schuierer G, Elbert T (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. Electroencephalogr Clin Neurophysiol 94(1):26–40

Pernet CR, Chauveau N, Gaspar C, Rousselet GA (2011) LIMO EEG: a toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. Comput Intell Neurosci 2011:831409. doi: 10.1155/2011/831409

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11(3):367–374

Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R, Jr, Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ (2000) Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37(2):127–152

Rogier O, Roux S, Belin P, Bonnet-Brilhault F, Bruneau N (2010) An electrophysiological correlate of voice processing in 4- to 5-year-old children. Int J Psychophysiol 75(1):44–47. doi: 10.1016/j.ijpsycho.2009.10.013

Rousselet GA, Husk JS, Pernet CR, Gaspar CM, Bennett PJ, Sekuler AB (2009) Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. BMC Neurosci 10:114

Schweinberger SR (2001) Human brain potential correlates of voice priming and voice recognition. Neuropsychologia 39(9):921–936

Schweinberger SR, Casper C, Hauthal N, Kaufmann JM, Kawahara H, Kloth N, Robertson DM, Simpson AP, Zaske R (2008) Auditory adaptation in voice perception. Curr Biol 18(9):684–688

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123(Pt 12):2400–2406

Sokhi DS, Hunter MD, Wilkinson ID, Woodruff PW (2005) Male and female voices activate distinct regions in the male brain. Neuroimage 27(3):572–578

Tiitinen H, Sivonen P, Alku P, Virtanen J, Naatanen R (1999) Electromagnetic recordings reveal latency differences in speech and tone processing in humans. Brain Res Cogn Brain Res 8(3): 355–363

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17(1): 48–55

Whiteside SP (1998) Identification of a speaker's sex: a study of vowels. Percept Mot Skills 86(2):579–584

Wilcox RR (2005) Introduction to robust estimation and hypothesis testing, 2d edn. Academic Press, San Diego

Yamaguchi MK, Hirukawa T, Kanazawa S (1995) Judgment of gender through facial parts. Perception 24(5):563–575

Zaske R, Schweinberger SR, Kaufmann JM, Kawahara H (2009) In the ear of the beholder: neural correlates of adaptation to voice gender. Eur J Neurosci 30(3):527–534