NeuroImage 52 (2010) 1528–1540

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Neural mechanisms for voice recognition

Attila Andics [a,b,c,*], James M. McQueen [a,d], Karl Magnus Petersson [a,b,e], Viktor Gál [c,f], Gábor Rudas [c], Zoltán Vidnyánszky [c,f]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands
[c] MR Research Center, Szentágothai János Knowledge Center, Semmelweis University, Budapest, Hungary
[d] Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, The Netherlands
[e] Cognitive Neuroscience Research Group, Institute for Biotechnology & Bioengineering, CBME, University of Algarve, Portugal
[f] Neurobionics Research Group, Hungarian Academy of Sciences – Péter Pázmány Catholic University – Semmelweis University, Budapest, Hungary

ARTICLE INFO

Article history:
Received 19 February 2010
Revised 12 May 2010
Accepted 14 May 2010
Available online 27 May 2010

Keywords:
fMRI
Voice recognition
Category learning
Voice typicality
Superior temporal sulcus
Anterior temporal pole

ABSTRACT

We investigated neural mechanisms that support voice recognition in a training paradigm with fMRI. The same listeners were trained on different weeks to categorize the mid-regions of voice-morph continua as an individual's voice. Stimuli implicitly defined a voice-acoustics space, and training explicitly defined a voice-identity space. The pre-defined centre of the voice category was shifted from the acoustic centre each week in opposite directions, so the same stimuli had different training histories on different tests. Cortical sensitivity to voice similarity appeared over different time-scales and at different representational stages. First, there were short-term adaptation effects: increasing acoustic similarity to the directly preceding stimulus led to haemodynamic response reduction in the middle/posterior STS and in right ventrolateral prefrontal regions. Second, there were longer-term effects: response reduction was found in the orbital/insular cortex for stimuli that were most versus least similar to the acoustic mean of all preceding stimuli, and, in the anterior temporal pole, the deep posterior STS and the amygdala, for stimuli that were most versus least similar to the trained voice-identity category mean. These findings are interpreted as effects of neural sharpening of long-term stored typical acoustic and category-internal values. The analyses also reveal anatomically separable voice representations: one in a voice-acoustics space and one in a voice-identity space. Voice-identity representations flexibly followed the trained identity shift, and listeners with a greater identity effect were more accurate at recognizing familiar voices. Voice recognition is thus supported by neural voice spaces that are organized around flexible 'mean voice' representations.

Introduction

The ecological significance of voices is reflected in the existence of regions in the primate (Petkov et al., 2008) and human cortex (Belin et al., 2000) that are specially tuned to conspecifics' vocalizations. Voices are used very efficiently for person recognition (e.g., Schweinberger et al., 1997). To do that, listeners need to link variable voice encounters to stable voice-identity categories. But how the brain could represent voice identities is still largely unknown. That is the central question of this paper.

To identify mechanisms that support voice recognition, one needs to separate voice-identity representations from earlier levels of voice processing. It has been suggested that a voice structural processing stage which is sensitive to voice-acoustic changes is anatomically separable from a voice-identity processing stage which is sensitive to changes in voice-identity (Belin et al., 2004; Campanella and Belin, 2007). Voice-acoustic analysis has been proposed to take place in voice-sensitive regions of the bilateral superior temporal sulci (Belin et al., 2000; Belin et al., 2002; von Kriegstein et al., 2003, 2005), and voice-identity analysis has been linked to regions of the right anterior temporal lobe (Nakamura et al., 2001; von Kriegstein et al., 2003, 2005; von Kriegstein and Giraud, 2004; Belin and Zatorre, 2003; Lattner et al., 2005; Sokhi et al., 2005).

Although this previous research has contributed considerably to our understanding of the separation of different voice processing stages, the precise nature of the underlying neural mechanisms at each of these stages is still unknown. One aim of this study was to address this issue. Furthermore, there is a common difficulty in the interpretation of many of the studies that have claimed to distinguish voice-identity representations from earlier levels of voice processing. This is that their critical contrasts were based on acoustic manipulations (e.g., Belin and Zatorre, 2003; Belin et al., 2000; Belin et al., 2002), task changes (e.g., Stevens, 2004; von Kriegstein et al., 2003), or both (e.g. von Kriegstein and Giraud, 2004). The proposed separation of voice processing stages

* Corresponding author. Szentágothai Knowledge Center – Semmelweis University MR Research Center, H-1083 Budapest, Balassa u. 6, Hungary. Fax: +36 1 459 1580.
E-mail address: attila.andics@gmail.com (A. Andics).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.05.048

may possibly reflect these acoustic and/or task differences. A second aim of the present study was therefore to try to distinguish between these processing stages with acoustic and task differences controlled. Several other cortical regions have also been implicated in voice processing in both primates and humans, including the anterior insular cortex (Remedios et al., 2009; Wong et al., 2004), the ventrolateral prefrontal cortex (Romanski et al., 2005; Fecteau et al., 2005), and paralimbic regions including the amygdala (Lloyd and Kling, 1988; Fecteau et al., 2007). A third aim was to try to clarify the role of these areas in voice recognition.

A useful voice processing mechanism positions voice stimuli in an object space. fMRI evidence on natural object processing suggests that stimuli that are more typical within an object space elicit reduced neural responses (Loffler et al., 2005; Myers, 2007; Belizaire et al., 2007). A possible neural mechanism for object space representation is based on neural sharpening: with experience, the coding of central values in relevant object dimensions becomes sparser (for a recent review, see Hoffman and Logothetis, 2009). Neural sharpening reflects long-lasting cortical plasticity and is thus suitable for positioning stimuli in an object space over the long term. Long-term neural sharpening has been demonstrated in a face space (Loffler et al., 2005). In a study on face-identity processing, reduced haemodynamic responses were found in the fusiform face area for central stimuli only when those were also central in the long-term stored face space of the viewer (referred to as 'mean face' stimuli, Loffler et al., 2005), suggesting that long-term central faces are encoded more sparsely. Based on these results and on behavioural findings that have indicated a prototype-centered representation of voices in long-term memory (Papcun et al., 1989; Mullennix et al., 2009; Bruckert et al., 2010), we can expect a typicality-based neural sharpening mechanism for voices similar to that found for faces.

But long-term neural sharpening is not the only mechanism that can explain response reduction for central stimuli. Another candidate mechanism is short-term neural adaptation: in case of fast and balanced stimulus presentation, neural response reduction for central stimuli can be a consequence of the on-average greater physical similarity of preceding events to central than to peripheral stimuli (Aguirre, 2007; Epstein et al., 2008). Short-term adaptation, just like neural sharpening, is sensitive to the object's relative position among similar objects, but in this case sensitivity is restricted to a very limited time scale. Short-term adaptation, in contrast with long-term neural sharpening, presupposes no long-term stored knowledge about the centre of the object space. But voice recognition cannot be successful without long-term stored information on person identity, that is, long-lasting voice-identity representations. Voice-acoustic analysis, on the contrary, might be based on short-term mechanisms exclusively, or it might be supported by an automatically formed, long-term stored voice-acoustics space, with a 'mean voice' as its centre. No previous studies have found evidence for the existence of such 'mean voice' representations. Here we attempted to identify long-lasting voice representations, and separate them from short-term stimulus similarity effects.

The present study evaluated two hypotheses. First, we attempted to confirm the hypothesis that person recognition from vocal information is mediated by anatomically separable stages of voice analysis (i.e., voice-acoustic analysis and voice-identity analysis). Second, we tested the hypothesis that voice analysis at each of these stages is supported by neural representations of the stimulus space such that long-term stored typical values are coded more sparsely than atypical values, that is, that there are both voice-acoustic and voice-identity spaces. To achieve these goals, we applied a learning–re-learning paradigm. Listeners were trained to categorize the middle part of several voice-morph continua as a certain person's voice. Because perceptually relevant inter-speaker and intra-speaker variation are largely based on the same acoustic cues (Potter and Steinberg, 1950; Nolan, 1997; Benzeghiba et al., 2007), the stimuli, although they were made by morphing between voices, nevertheless modeled natural within-voice variability in the way each individual produces spoken words. The training hence simulated normal voice learning, where the same voice-identity must be linked to variable tokens of words. The trained voice-identity category was associated with a different interval on the voice-morph continua on each of 2 weeks for every listener. The voice-acoustics space was defined implicitly by the stimulus continuum used throughout the experiment, while the voice-identity space was defined by explicit feedback during training. Training was followed by fMRI tests each week.

We thus investigated two equivalent contrasts with the same subjects, the same stimuli and the same task. One contrast measured voice-acoustic sensitivity and the other measured voice-identity sensitivity. We predicted that if a neural region is sensitive to deviations from long-term stored typical values in either the voice-acoustic or the voice-identity space, then that region will respond less strongly to acoustically central or trained identity-internal stimuli than to acoustically peripheral or trained identity-external stimuli respectively, while remaining insensitive to short-term adaptation effects. To reveal the contribution of long-term and short-term mechanisms behind these sensitivities, we separated the effect of stimulus similarity to the directly preceding voice stimulus from longer-lasting effects.

Materials and methods

Participants

Twenty-five Hungarian listeners (14 females, 11 males, 19–31 years) with no reported hearing disorders were paid to complete the experiment. Written informed consent was obtained from all participants. One person was excluded because of a failure to perform the task during training. The analyses presented below were based on the remaining twenty-four subjects.

Stimuli

Recording

We recorded two young female non-smoking native Hungarian speakers with no speech disorders, saying the Hungarian words "bú" [sadness], "fű" [grass], "ki" [out], "lé" [liquid], "ma" [today] and "se" [neither] in standard Hungarian with no recognizable regional accent (voiceA and voiceB). These monosyllables were selected to cover various types of segmental content, with consonants varying in manner and place of articulation and in voicing, and with vowels varying in height, backness, roundedness and length. Speakers were similar in pitch (voiceA: 195 Hz, voiceB: 179 Hz), as shown by measurements averaged across the six words. Recordings were made in a soundproof booth using a Sennheizer Microphone ME62, a MultiMIX mixer panel, and Sony Sound Forge. All stimuli were digitized at a 16 bit/44.1 kHz sampling rate and were volume balanced using Praat software (Version 4.2.07; Boersma and Weenink, 2005).

Morphing

Voice morphing was then performed between the natural endpoint tokens of the two speakers, making one 100-step continuum per word (voiceA = morph0, voiceB = morph100). Intermediate steps were made using the morphing algorithms of STRAIGHT (Kawahara, 2006).

Perceptual rescaling

To ensure approximately equal perceptual distances between neighbouring steps on each of the stimulus continua, the morphs for each of the six words were subjected to perceptually-informed rescaling. A behavioural pretest was carried out in order to acquire psychophysical data which could then be used for re-labelling the morph steps. In this pretest, ten repetitions of seven steps (5, 20, 35,

50, 65, 80 and 95) of each of the six morph continua were presented, in random order, to 10 naive listeners who performed a forced-choice voiceA or voiceB categorization task (these listeners did not take part in the main experiment). There was no training or feedback provided. The test directly followed an initial voice-to-response-button assignment, in which listeners were presented with a single repetition of all six natural endpoint tokens of each speaker. Group-averaged 'voiceB' response proportions per level for each continuum were then subjected to linear interpolation, to get estimates of how each step of each continuum would be perceived. All morph steps were then re-labelled to best match the corresponding, interpolated 'voiceB' response proportions. For example, after perceptual rescaling, morph20 for each word refers to the morph step on that word continuum whose identification proportion as 'voiceB' was closest to 20% in this pretest. Example stimuli are available as Supplementary data.

Training

Design

The voices were unfamiliar to all listeners. Listeners were trained to categorize the middle parts of the voice-morph continua as a certain person's voice (we call this the trained voice identity). They had to perform an A or not-A categorization task on each stimulus (Ashby and Maddox, 2005). They were asked whether the presented stimulus was an exemplar of the trained voice identity or of a different voice. A within-subject training manipulation was applied. The trained voice-identity category was associated with a different interval on the voice-morph continua on each of 2 weeks for every listener, namely either the morph20–morph60 range or the morph40–morph80 range — these will be referred to as 'voice20–60 training' and 'voice40–80 training', respectively. The whole continuum was sampled each week, and listeners were presented with exactly the same stimuli (with a different trial order) during the two training sessions. The difference between the training conditions was restricted to the feedback that was provided. The order of the training sessions was counterbalanced: half of the listeners had voice20–60 training on the first week and voice40–80 training on the second week, while the other half of the listeners had the reverse order.

During training, 25 stimuli from each of the six 0–100 voice-morph continua were presented, sampling the continua at approximately equal perceptual distances (a difference of 4 steps). The steps used were morphs 2, 6, 10, … , 90, 94, and 98. To maximize any training effect, the 8 stimulus steps that were closest to the critical 20, 40, 60 or 80 levels (i.e., those that were used at test) were presented twice as often as the rest (these steps were 18, 22, 38, 42, 58, 62, 78, 82). There was, however, no difference in presentation frequency between central and peripheral stimuli. In each of two weeks participants received 80 min of training over 2 days, with 4 training sessions of 16 min each on day 1 and a single training block on day 2. The first two blocks were blocked by word; in subsequent blocks the words were mixed. Training was followed by an fMRI test session on day 2 in each week.

Procedure

Trial onsets were signaled with a question mark displayed in the middle of the screen for 300 ms. The auditory stimulus (a voice morph of one of the six words) began 200 ms after trial onset and lasted on average 456 ms. A response had to be made within 1800 ms of stimulus onset. Listeners received feedback on every trial. This feedback consisted of two parts. First, they saw an evaluation of their performance (i.e., whether the response was correct, incorrect or late) between 2000 and 2250 ms after trial onset. Second, this visual feedback was followed by auditory and visual reinforcement of learning. Listeners were presented with a repetition of the auditory stimulus, starting at 2700 ms after trial onset. This auditory reinforcement was accompanied with temporally synchronized visual reinforcement (a picture) presented between 2700 and 3450 ms after trial onset. If the stimulus morph was within the pre-defined trained voice-identity category (in 42% of all trials), then this picture was a face (positive feedback). If the stimulus morph was outside the trained voice-identity category, then a scrambled picture (matched in size, colour and contrast) was presented instead of the face (negative feedback). The same female face and the same scrambled picture were shown to all listeners in all training sessions on both weeks. We used the same face throughout the experiment in order to model natural voice learning, where acoustic variability in the realization of spoken words has to be mapped onto the same voice-identity. The manipulation appeared to be successful in that all participants reported, after the experiment, that they thought that they had heard various exemplars of natural voices only and that they were convinced that the trained voice was an actual person's voice. The face was unfamiliar to all listeners before the experiment. They were told that it was the trained talker's face at the beginning of a half-minute long practice session on the training task which was presented before the first training session. The procedure ensured that every training stimulus was immediately repeated after the listener had made their choice, but for the second time with a visually disambiguated talker identity. No response had to be made on the repeated stimulus. Trials had a duration of 5500 ms.

Conditions of interest

The critical stimuli in the fMRI test were morphs 20, 40, 60 and 80. The categorization training defined identity membership of these stimuli (internal, boundary and external), although these specific morph levels were not presented during training. During voice20–60 training, morph40 stimuli were category-internal, morph80 stimuli were category-external, and morph20 and morph60 stimuli were at the category boundaries. In contrast, during voice40–80 training, morph60 stimuli were internal, morph20 stimuli were external, and morph40 and morph80 stimuli were at the boundaries. Voice-identity membership was trained by giving explicit feedback on every trial. Feedback was always positive for stimuli within the artificially determined trained voice-identity interval, and it was always negative for stimuli outside this interval. During voice20–60 training, for example, morph steps greater than 20 but smaller than 60 were trained as internal through positive feedback, and morph steps smaller than 20 or greater than 60 were trained as external through negative feedback. An analogous procedure was used for voice40–80 training. As a consequence, out of the trained morph levels corresponding to the internal, boundary and external conditions at test, the proportion of morphs with positive feedback was 100, 50 and 0%, respectively. This defined the identity space. The stimuli therefore also differed in categorization ambiguity: it was expected that internal and external stimuli would be categorized less ambiguously and more accurately than boundary stimuli.

The critical voice morphs also differed in terms of their distributional position on the stimulus continua: morph40 and morph60 were close to the middles of the continua, while morph20 and morph80 were close to the endpoints — these morphs will be referred to as acoustic-central and acoustic-peripheral stimuli, respectively. Identity-internal stimuli were always acoustic-central, identity-external stimuli were always acoustic-peripheral, and identity-boundary stimuli were acoustic-central and acoustic-peripheral equally often. See Fig. 1A for an overview of the training and test design.

Analyses of training data

Voice-category training data were collapsed across training blocks and days, and binned around the nine morph levels used at test (10, 20, …, 90) applying a $+/-4$ morph step interval. This was done to enable a direct comparison of the training data to the fMRI test data (see Figs. 1B,C). The trained morph levels 2 and 98 were not included

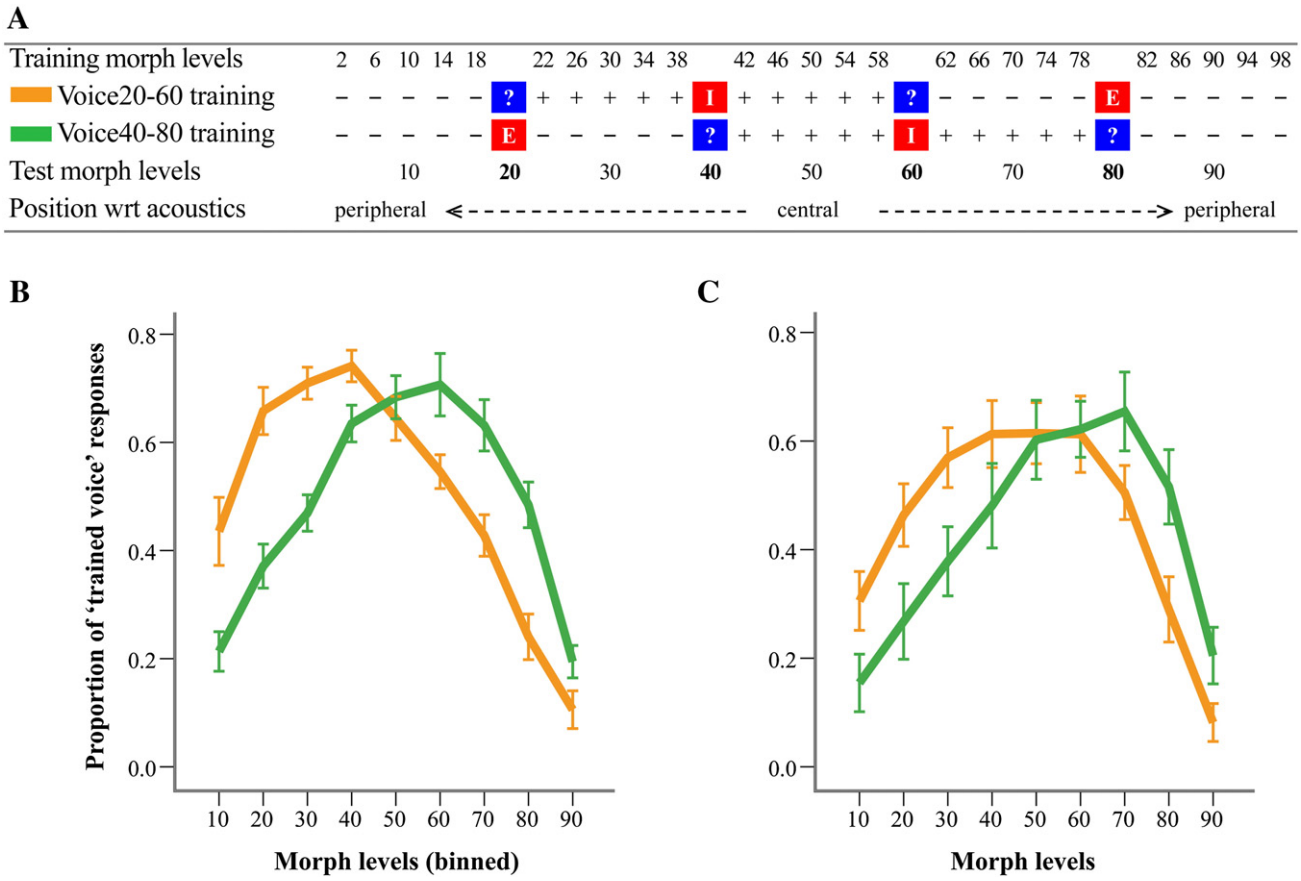

Fig. 1. Design and behavioural results. (A) Training and test design. Stimulus position with respect to identity was defined via feedback during training: internal morphs were associated with positive feedback (+) and external morphs with negative feedback (−). For critical test stimuli (morphs 20, 40, 60 and 80; in bold), which were not presented during training, stimulus position with respect to identity was in half of the cases (red boxes) internal (I) for more central and external (E) for more peripheral stimuli, while in the other, stimulus-matched half of the cases (blue boxes) stimulus position was at the voice category boundary (?) for both central and peripheral stimuli. (B) Proportion of 'trained voice' responses across binned morph levels during training, collapsing over all training blocks in each condition. Error bars represent the standard error of the mean ($n = 24$). (C) Proportion of 'trained voice' responses across morph levels at fMRI test, for each training condition. Error bars represent the standard error of the mean ($n = 24$).

in any bins. The non-critical morph level bins (10, 30, 50, 70, 90) comprised three stimuli that were actually used in training (the morph in the middle of the bin plus those in a 4-step distance in both directions, e.g., bin 10 comprised data corresponding to stimulus levels 6, 10 and 14). Each of these non-critical bins corresponded to 90 trials per condition, per subject (30 trials per stimulus level). The critical morph level bins 20, 40, 60 and 80 comprised two actually trained stimulus levels, in a 2-step distance in both directions (e.g., bin 20 comprised data corresponding to stimulus levels 18 and 22 — the actual morph level 20 was only presented at test). Every critical bin corresponded to 120 trials per condition, per subject (60 trials per stimulus level, as the number of repetitions on these critical stimulus levels was doubled).

fMRI test

Design and procedure

At fMRI test the task was the same as during training ("do you hear the trained voice identity or another voice?"), but no feedback was given. The 10-minute test contained 216 trials (four repetitions of six word continua, sampling each continuum with 9 morph levels, namely 10, 20, 30, 40, 50, 60, 70, 80 and 90), and a button-press response was expected after each stimulus. Trials had a duration of 2500 ms. Stimulus presentation was blocked by word continuum: all 9 levels of a word continuum were presented in each 9-trial-miniblock. Morph levels were therefore evenly distributed throughout the trial sequence. The word was different in consecutive miniblocks, and stimuli in consecutive trials were physically different. Stimulus ordering was otherwise random and varied across listeners. An example of a miniblock is: "lé" [30] — "lé" [80] — "lé" [10] — "lé" [50] — "lé" [40] — "lé" [90] — "lé" [20] — "lé" [70] — "lé" [60].

We explored the role of the task in an additional test in which subjects had to perform a word-repetition detection task by pressing a button when two consecutive words were the same. For this task the trained voice-category-membership properties (i.e. whether they were exemplars of the trained voice identity or of another voice) were irrelevant. Two 9-minute runs with stimuli from the six trained word continua, sampled with the critical morph levels 20, 40, 60 and 80, were presented. At this test stimulus presentation was blocked by morph level, in 7-trial-miniblocks. Every miniblock contained each of the six words, and exactly one of them was repeated in each miniblock, in a randomly-chosen position within the block. An example of a miniblock at the irrelevant-task test is: "ki" [40] — "lé" [40] — "bú" [40] — "fű" [40] — "fű" [40] — "ma" [40] — "se" [40]. A response was expected for the second "fű" stimulus but not for the other six stimuli in the block. Subjects were not informed about the frequency of word repetitions.

This irrelevant-task test preceded the relevant-task test each week. The constant order of tests was preferred to a balanced ordering because our focus was not on a direct comparison of the two tasks, but rather on a direct comparison of training effects across weeks within each test. We assumed that a constant order of tests would reduce noise caused by variation in listening history and in the amount of time already spent in the fMRI scanner.

Further tests included a single localizer run for voice-sensitive regions in the first week (including blocks of vocal and non-vocal sounds, using the stimuli from Pernet et al., 2007, with passive listening), and one for face-sensitive regions (including blocks of faces, houses, objects and matched scrambled objects, with a picture repetition detection task) in the second week.

Stimuli were presented at a standard, comfortable volume. Stimuli were controlled using Presentation software (Version 10.2; www.neurobs.com). During imaging, stimulus presentation was synchronized by a TTL trigger pulse with the data acquisition. Stimuli were delivered binaurally through MRI-compatible headphones (MR Confon, Magdeburg, Germany).

Data acquisition

MRI measurements were performed on a Philips Achieva 3 T whole body MR unit (Philips Medical Systems, Best, The Netherlands) equipped with an eight-channel Philips SENSE head coil. For the main tests EPI-BOLD fMRI time series were obtained from 27 transverse slices covering temporal lobes and the inferior part of the frontal lobes with a spatial resolution of $3.5 \times 3.5 \times 3$ mm, including a 0.5 mm slice gap, using a single-shot gradient-echo planar sequence (parallel imaging; ascending slice order; acquisition matrix 64×64; FOV = 224 mm; TR = 2500 ms; TA = 1763 ms (i.e., 737 ms silent gap); TE = 32.3 ms; and flip angle = 90°). That is, the acquisition of each volume was followed by a 737 ms gap when the scanner was silent. Compared to standard sparse sampling methods, this close-to-continuous sampling method not only increased statistical power by increasing the number of data points, but also made it possible to haemodynamically model each individual stimulus. At the same time it was possible to present all auditory stimuli in silence (stimulus onset time coincided with scanner silent gap onset). The relevant and irrelevant-task runs included 265 and 225 volumes respectively.

For the voice localizer there were 29 transverse slices and a longer silent gap between acquisitions (TR = 10,000 ms, including 2000 ms acquisition and 8000 ms silent gap; TE = 36.5 ms). For the face localizer we used continuous scanning with 31 transverse slices (TR = 2200 ms; TE = 37 ms). The voice and face localizer runs included 63 and 200 volumes respectively. All other parameters were identical to the main test settings.

In addition to the functional time series, a standard T1-weighted three-dimensional scan using a turbo-field echo (TFE) sequence with 180 slices covering the whole brain was collected for anatomical reference at the end of the second scanning session, with $1 \times 1 \times 1$ mm spatial resolution.

Data analysis

Image preprocessing and statistical analysis were performed using SPM5 (www.fil.ion.ucl.ac.uk/spm). The functional EPI-BOLD images were realigned, slice-time corrected (except for the voice area localizer run, where each volume acquisition was followed by a four times longer silent gap, and in this case slice-time correction is known to be more harmful than helpful, Friston et al., 2007), spatially normalized, and transformed into a common anatomical space, as defined by the SPM Montreal Neurological Institute (MNI) T1 template. Next, the functional EPI-BOLD images were spatially filtered by convolving the functional images with an isotropic 3-D Gaussian kernel (10 mm FWHM). The fMRI data were then statistically analyzed using a general linear model and statistical parametric mapping (Friston et al., 2007). For the relevant-task run, every single stimulus was modeled as a separate event. For the irrelevant-task run, seven consecutive stimuli, all representing the same voice-morph level, were modeled as a block. Conditions in the voice and face localizer runs were also modeled as blocks.

For the main analyses, condition regressors for the relevant and irrelevant-task tests were constructed per morph level. Sensitivity to voice-acoustic stimulus similarities was measured in a test contrasting continuum-central and continuum-peripheral stimuli, but controlling for category-membership properties by only including stimuli that were trained as identity boundaries. After voice20–60 training, these were morphs 20 and 60; after voice40–80 training these were morphs 40 and 80 (see Fig. 1A). Voice-identity sensitivity was tested in a contrast that had an identical stimulus load to that of the acoustic contrast, but those stimuli now also entailed a training-induced identity manipulation. Trained internal stimuli were compared to external stimuli (after voice20–60 training these were morphs 40 and 80 respectively, after voice40–80 training these were morphs 60 and 20 respectively; see Fig. 1A).

To determine the role of short-term stimulus similarity-based mechanisms in the relevant-task test, an additional analysis was performed. For that, critical condition regressors (corresponding to morphs 20, 40, 60, and 80) were split into more regressors, based on a one-back-distance measure, that is, the morph level distance of the actual trial from the preceding one (regressors of the new model: c10, c20, c30, c40, c50, p10, p20, p30, p40, p50, p60, p70; i10, i20, i30, i40, i50, e10, e20, e30, e40, e50, e60, e70 — where the number refers to the one-back-distance and c = acoustic-central from acoustic test, p = acoustic-peripheral from acoustic test, i = identity-internal, and e = identity-external). For example, the condition c10 involved acoustic-central stimuli as used in the acoustic test (so only identity-boundary cases are included) for which the preceding stimulus was 10 morph steps distant (e.g., after voice20–60 training, this would comprise those morph60 trials that come after morph50 or morph70). The effect of short-term similarity sensitivity was then measured by comparing trials with the minimal one-back distance to trials with the maximal one-back distance ($c10 + p10 + i10 + e10 < c50 + p50 + i50 + e50$; distances larger than 50 were not available for all critical conditions).

This split regressor model was also used in confirmatory follow-up tests that were aimed at distinguishing long-term from short-term effects. They did so by controlling for short-term biases in the main acoustic and identity tests. In those tests, low one-back distances were more frequent and thus overweighted among acoustic-central and identity-internal trials, while high one-back distances were more frequent and thus overweighted among acoustic-peripheral and identity-external trials. In the follow-up tests equal weights were therefore assigned to all one-back distances. The main acoustic analysis contrast $c < p$ was substituted with $c10 + c20 + c30 + c40 + c50 < p10 + p20 + p30 + p40 + p50$, and the main identity analysis contrast $i < e$ was substituted with $i10 + i20 + i30 + i40 + i50 < e10 + e20 + e30 + e40 + e50$.

Realignment regressors were also included for each run to model potential movement artefacts. A high-pass filter with a cycle-cutoff of 128 s was implemented in the design to remove low-frequency signals. Single-subject fixed effect analyses were followed by whole-brain random effects analyses on the group level. Significance levels were FDR-corrected.

Results

Behavioural results

The training was successful and had long-lasting effects: listeners learned that the voice category was located in the middle of the presented stimulus continua, and they shifted this category during re-learning on the second week (Fig. 1B). The learning effect found during training was present at the fMRI test as well (Fig. 1C). Repeated-measures ANOVAs on categorization responses during the training and then at the fMRI test examined the effect of condition (voice20–60 training or voice40–80 training) across nine morph levels (10, 20, …, 90; as described above, these levels for the training phase were created by binning data around these values). We found a main effect of morph level (training: $F(8, 184) = 257.89$, $p < 0.001$; test: $F(8, 184) = 70.21$, $p < 0.001$), no main effect of condition

(training: $F(1, 23) = 1.40$, $p = 0.250$; test: $F(1, 23) = 1.18$, $p = 0.289$), and a significant condition by morph level interaction (training: $F(8, 184) = 21.44$, $p < 0.001$; test: $F(8, 184) = 67.47$, $p < 0.001$). Moreover, the quadratic trend was highly significant for this interaction during training and at test (training: $F(1,23) = 643.86$, $p < 0.001$; test: $F(1,23) = 287.17$, $p < 0.001$). We also found a significant linear trend during training but not at test (training: $F(1,23) = 97.04$, $p < 0.001$; test: $F(1,23) < 1$). The presented degrees of freedom are uncorrected, but were Greenhouse–Geisser corrected for F score calculations.

Recognition performance accuracy during training was calculated for every listener (mean $d' = .85$, SD = .19). For the d primes, hit rates versus false alarm rates were calculated from responses to all stimuli with positive versus negative feedback respectively. These recognition accuracy scores were later compared to neural sensitivity scores in correlation analyses.

Decision difficulty affected both recognition accuracy and response times (see Table 1). The training stimuli corresponding to the boundary stimuli were categorized with lower recognition accuracy than those corresponding to internal and external stimuli. Response times during training were significantly longer for trials corresponding to boundary stimuli than for trials corresponding to internal/external stimuli. The same pattern was observed at test. Note that the stimulus load contributing to the easy and difficult conditions was identical.

fMRI results

Acoustic sensitivity

This test contrasted continuum-central and continuum-peripheral stimuli, including only identity-boundary trials in each condition (see Fig. 1A). Large regions were found in a whole-brain analysis (FDR-correction, $p < .05$). Clusters that showed response reduction for central compared to peripheral stimuli included anterior, middle and posterior parts of the bilateral superior temporal sulcus (STS; BA 21, 22), the bilateral orbitofrontal cortex extending to the anterior insula (BA 47, 11) and the bilateral posterior ventrolateral prefrontal cortex (VLPFC) along the inferior bank of the inferior frontal sulcus (BA 44, 45) (see Fig. 2 and Table 2). No clusters were found in the opposite test.

Identity sensitivity

Here we compared identity-internal to identity-external stimuli in a contrast that had an identical stimulus load to that of the acoustic contrast (see Fig. 1A). Reduced BOLD responses were found for identity-internal compared to identity-external stimuli in the bilateral middle and posterior STS (BA 21, 22) extending ventromedially to the middle temporal gyrus in the right hemisphere, and medially to the Heschl's gyrus in the left hemisphere (BA 41); the bilateral anterior temporal pole (BA 38); the left amygdala; and a left deep posterior STS region (BA 39) in the proximity of the angular gyrus and the intraparietal sulcus (see Fig. 2. and Table 2.). No regions were found in the reverse contrast.

There was a partial overlap of the posterior STS clusters found in the acoustic and the identity tests, in both hemispheres. There were no voxels in any other cortical areas that were significantly active in both the acoustic and the identity tests, not even at a more liberal threshold ($p < .001$, uncorrected).

Table 1
Voice recognition accuracy (d') and response times (RTs) at training and test.

	Boundary	Internal/external	$t(23)$
Training d'	.143 (+/−.136)	1.131 (+/−.324)	16.636*
Training RT (ms)	940 (+/−155)	924 (+/−156)	4.047*
Test RT (ms)	954 (+/−186)	931 (+/−182)	3.783*

The values refer to group mean and to standard deviations.
* Significant paired t-tests ($p < .001$).

Correlation analyses

To investigate the behavioural relevance of the variation in neural activity found in the acoustic contrast and identity contrast, these tests were followed up by correlation analyses. Recognition performance accuracy during training, characterized by d-prime scores for every subject, was compared to neural sensitivity, characterized by the size of significant response reductions in regions found in either contrast. Behavioural scores were added to both the acoustic and the identity contrast's group design matrix as a regressor. In the context of the GLM, carrying out a t-test on the coefficient of this regressor is equivalent to testing the corresponding correlation.

Small-volume correction analyses were performed for every activated cluster. Seven acoustic clusters and six identity clusters were investigated. Table 3 reports the local maxima and corrected p-values (corrected for the number of voxels within each cluster, but uncorrected for the number of tested clusters) for the behavioural regressor. Peaks with a significant correlation with recognition accuracy were found for identity clusters: the right middle/posterior STS (BA 21,22), the left deep posterior STS (BA 39), the right anterior temporal pole (BA 38), and the left amygdala (see Fig. 3). No significant positive correlations were found for acoustic clusters. No significant regions showed negative correlations between acoustic or identity sensitivity and behaviour.

Short-term effects

To determine whether the acoustic and identity effects could be caused by short-term perceptual similarity-based mechanisms, an additional analysis was performed. The short-term effect was measured in a contrast orthogonal to the acoustic and identity tests, by taking all critical conditions and comparing trials with the minimal distance between the stimulus and the immediately preceding stimulus (10 morph steps) to trials with the maximal distance between stimuli (50 morph steps). We expected that in regions sensitive to short-term stimulus similarities we would see an effect of one-back distance. Reported results were thresholded at the whole-brain level ($t > 4$, see Table 4, Figs. 2 and 4). Reduced BOLD responses were found for minimal-distance compared to maximal-distance stimuli in the bilateral middle/posterior STS (BA 21, 22), extending medially to the Heschl's gyrus (BA 42), and in the right hemisphere also ventromedially to the inferior temporal gyrus (BA 20). A further cluster was found in the right posterior ventrolateral prefrontal cortex (BA 44). The bilateral temporal clusters overlapped with the bilateral STS clusters of both the acoustic and the identity test. The right VLPFC cluster also overlapped with that found in the main acoustic test (see Table 6). This suggests that the STS and right VLPFC clusters detected in the main acoustic analyses and the STS clusters found in the main identity analyses are findings that can at least partially be explained by short-term adaptation effects. No regions were found in the reverse contrast.

Long-term effects

We have seen that some but not all of the acoustic and identity effects could be explained by short-term similarity-based mechanisms. To confirm that brain regions with acoustic or identity sensitivity but without a sensitivity to short-term similarities were indeed based on long-term mechanisms, we followed up on the acoustic and identity tests in a confirmatory analysis. ('Long-term' here and throughout the paper refers to a time interval that is longer than the distance between two consecutive trials.) We used contrasts that were parallel to the main acoustic and identity analysis contrasts, but we defined the contrasts with separate regressors for each distance (10, 20, 30, 40, 50 morph steps) from the preceding stimulus, to control for short-term stimulus similarity effects.

Results were thresholded at $t(23) > 3$ and small-volume corrected for each of the corresponding main analysis clusters (seven acoustic or six identity clusters, thresholded at $t(23) > 4$, see Fig. 4). Table 5

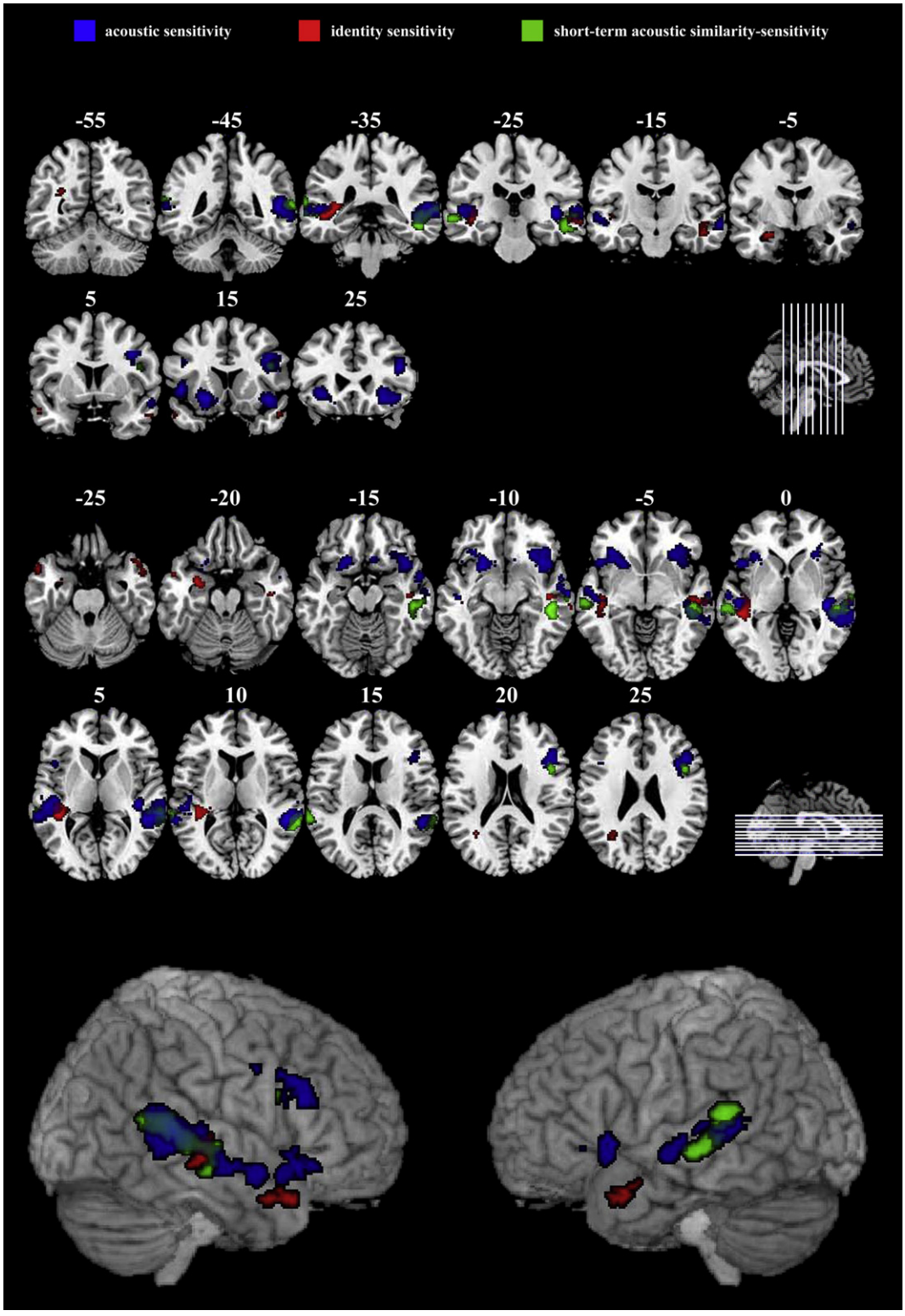

Fig. 2. Coronal and axial slices and sagittal views display significant acoustic sensitivity (blue), identity sensitivity (red) and short-term effects (green), thresholded at $t(23)>4$.

reports the local maxima and corrected *p*-values (corrected for the number of voxels within each cluster, but uncorrected for the number of tested clusters) for the long-term acoustic and identity sensitivity tests. Long-term acoustic sensitivity (response reduction to short-term controlled central compared to short-term controlled peripheral stimuli) was found in the right orbital/insular cortex (BA 47, 11); and in the posterior medial portion of the right STS cluster, close to the junction of BA 20, 37 and 41. No significant long-term acoustic sensitivity was found in the left STS cluster, the VLPFC clusters and the left orbital/insular cluster. Long-term identity sensitivity (response reduction to short-term controlled identity-internal compared to short-term controlled identity-external stimuli) was found in the bilateral anterior temporal pole (BA 38); in the left deep posterior STS region (BA 39) and in the left amygdala. No significant long-term

Table 2
List of regions found in the voice-acoustic and voice-identity sensitivity tests.

	BA	x	y	z	T	p (corr.)	mm³
Acoustic sensitivity							
R anterior/middle/posterior STS	21/22	58	−38	6	8.16	0.001	11312
L anterior/middle/posterior STS	21/22	−50	−32	4	6.11	0.002	5248
R orbitofrontal cortex/anterior insula	47	42	16	−12	5.67	0.003	5184
R medial orbitofrontal cortex	11	10	20	−14	5.88	0.003	248
L orbitofrontal cortex/anterior insula	47/11	−22	14	−12	6.03	0.003	4936
R posterior VLPFC	44/45	36	6	34	6.62	0.002	4672
L posterior VLPFC	45	−44	16	28	4.21	0.009	88
Identity sensitivity							
R middle/posterior STS	21/22	50	−20	−6	5.05	0.040	1416
L middle/posterior STS	21/22/41	−42	−38	4	6.01	0.037	3376
L deep posterior STS	39	−30	−58	22	5.40	0.037	304
R anterior temporal pole	21/38	48	18	−28	4.68	0.045	304
L anterior temporal pole	21/38	−54	10	−24	4.59	0.046	272
L amygdala	–	−30	−2	−20	4.86	0.041	464

A single peak per region is shown. Analyses were thresholded at $t(23)>4$, cluster size>10 voxels.

identity sensitivity was found in the middle/posterior STS clusters in either hemisphere. No clusters were found in the opposite tests. Although these confirmatory analyses are based on functionally non-independent small-volume corrections that can possibly result in false positives, they are nevertheless strict tests, since the largest STS clusters found in the main analyses did not survive them. These analyses thus suggest that activity in most of the brain regions that was found in the main acoustic and identity analyses, and that remained insensitive to short-term stimulus similarities, can indeed be explained by long-term mechanisms.

Voice- and face-sensitivity

Voice-sensitivity was measured with a functional localizer (Pernet et al., 2007) using a contrast of voice stimuli versus matched non-voice stimuli. Face-sensitivity was measured with another functional localizer using a contrast of faces versus matched scrambled objects. The localizer activities were thresholded at $t>4$ and narrowed down for the activated clusters of the acoustic and the identity test (Table 6). Among acoustic test clusters, a high proportion of voxels within the STS clusters showed voice-sensitivity, and the posterior part of the right STS also showed considerable face-sensitivity. Part of the right posterior VLPFC region from the acoustic test was also shown to be sensitive to voices but not to faces. In identity test clusters, the overwhelming majority of activated voxels in the bilateral middle/posterior STS and anterior temporal pole showed voice-sensitivity, but none showed face-sensitivity. On the contrary, the left amygdala as found in the identity test showed clear face-sensitivity but almost no voice-sensitivity. Interestingly, the left deep posterior STS region of the identity test which was also well correlated with recognition accuracy did not contain any voice- or face-sensitive voxels.

Table 3
Correlation of recognition accuracy and significant acoustic or identity sensitivity.

	BA	x	y	z	T	p (corr.)
Correlation with acoustic sensitivity	[No clusters contained suprathreshold voxels]					
Correlation with identity sensitivity						
R middle/posterior STS	21/22	46	−14	−20	3.86	**0.020**
L middle/posterior STS	21/22	−40	−42	10	3.16	0.357
L deep posterior STS	39	−32	−60	24	3.56	**0.015**
R anterior temporal pole	38	56	8	−28	3.39	**0.030**
L anterior temporal pole	[No suprathreshold voxels]					
L amygdala	–	−30	2	−22	3.81	**0.028**

Correlation contrasts were thresholded at $t(23)>3$. Small-volume correction was based on clusters from the corresponding main analyses, thresholded at $t(23)>4$. Significant probability values ($p<.05$) are highlighted in bold.

Lateralization

To directly compare hemispheric contributions to the two contrasts, lateralization indices were calculated from voxel values for the temporal lobes, where large clusters were found in both tests. Individual maps were thresholded at $p<.05$ uncorrected. Activity in the identity test was left-lateralized (mean(LI) = −.141, SD(LI) = .392), but in the acoustic test it was right-lateralized (mean(LI) = .182, SD(LI) = .406). There was a significant difference of individual lateralization indices in the temporal lobes between tests ($p=.025$, paired t-test).

The role of decision difficulty

To explore direct effects of decision difficulty on critical stimuli, a test comparing difficult and easy trials was performed. Difficult trials included the ambiguous identity-boundary stimuli, that is, all stimuli of the acoustic test. Stimulus-matched easy trials included the unambiguously trained identity-internal and identity-external stimuli, that is, all stimuli of the identity test. No significant voxels were found in either direction of the comparison (whole-brain analysis, FDR-correction, $p<.05$).

The role of the task

As noted in the Materials and methods, there was a test where listeners performed a word-repetition detection task instead of voice recognition, on the same stimuli. In an analysis of the fMRI data for this word-repetition task, no significantly active regions were found for the same acoustic and identity contrasts as were used in the main analysis.

Discussion

Voice-identity processing is separable from voice-acoustic processing

It has been proposed that the neural substrates for the recognition of voice identities are separable from general acoustic processing regions (see Belin et al., 2004 for a review). This view has been strengthened by reports on cortical regions that are differentially active in voice recognition tasks (Nakamura et al., 2001; von Kriegstein et al., 2003, 2005; Belin and Zatorre, 2003; Lattner et al., 2005; Stevens, 2004), and on selective deficits of voice-identity recognition abilities (Van Lancker et al., 1988; Garrido et al., 2009a,b). Nevertheless, until now there were few attempts to describe the neural mechanisms underlying voice-identity representations. We identified identity-sensitive regions that are both functionally and anatomically distinct from acoustic-sensitive regions. While temporal lobe activity in the acoustic contrast was right-lateralized, it was left-lateralized in the

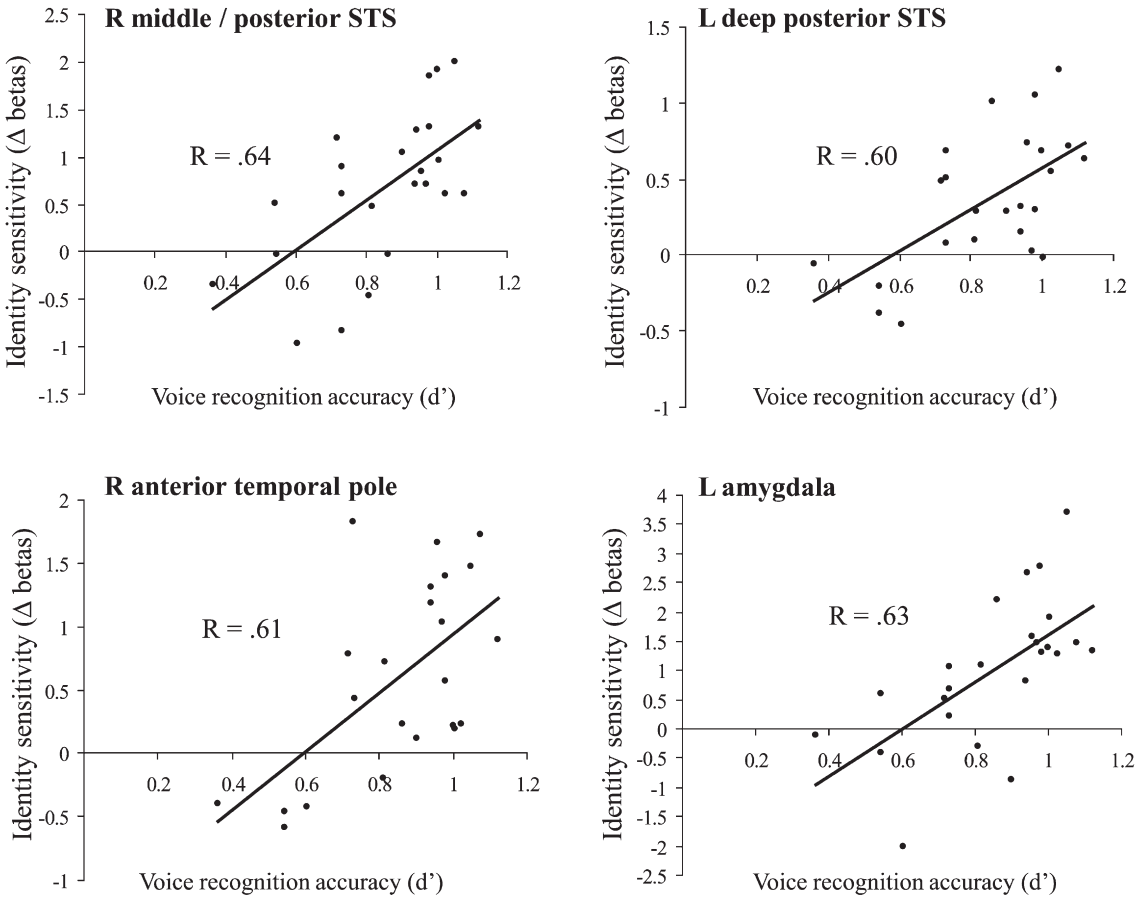

Fig. 3. Significant correlations between voice recognition accuracy scores and neural identity sensitivity for the peak coordinates defined in the correlation analyses (dots denote individuals).

identity contrast. This lateralization difference suggested that these stimulus-matched and task-matched contrasts indeed measure different functions. Identity-sensitive but not acoustically sensitive regions involved the voice-sensitive bilateral anterior temporal pole; the face-sensitive left amygdala; and a left deep posterior STS region which was not found in either of the functional localizer tests.

Voice-identity but not voice-acoustic sensitivity was found to covary with person identification performance. This covariation suggests that the identity sensitivity we described is indeed useful for voice recognition: listeners with a greater neural sensitivity for voice identities are more accurate at recognizing familiar voices. Covariation between significant identity sensitivity and behaviour was found for voice-sensitive regions (the middle/posterior STS and the anterior temporal pole) in the right but not in the left hemisphere. Right-hemisphere biases in voice recognition have been reported both in imaging (Nakamura et al., 2001; von Kriegstein et al., 2003; von Kriegstein and Giraud, 2004) and in clinical studies (Van Lancker and Kreiman, 1987; Ellis et al., 1989; Van Lancker et al., 1989; Gainotti et al., 2003). Covariation was also found between neural and behavioural identity sensitivity in regions that were not differentially sensitive to voices in the voice-localizer test, namely the amygdala and the deep posterior STS in the left hemisphere. These covariations not only validate our identity test but are also among the first demonstrations of the direct behavioural relevance of voice-identity representations. In addition, the fact that we did not find any significant covariation between neural sensitivity in acoustic regions and performance further strengthens our claim that identity processing is separable from acoustic processing.

Table 4
List of regions found in the short-term acoustic similarity sensitivity test.

	BA	x	y	z	T	p (corr.)	mm³
Short-term similarity sensitivity							
R middle/posterior STS, ITG	20/21/22/42	48	−32	−6	6.04	0.026	6256
R posterior VLPFC	44	46	14	22	5.23	0.026	640
L posterior STS	22/42	−66	−38	12	5.22	0.026	592
L middle/posterior STS	21/22	−62	−26	−4	5.02	0.026	1136

A single peak is shown per region. The analysis was thresholded at $t(23)>4$, cluster size>10 voxels.

Short-term similarity effects

Auditory stimuli that are similar to other, just presented stimuli are expected to elicit more reduced neural responses than dissimilar stimuli, in cortical regions that are sensitive to those auditory changes. This neural mechanism is known as the short-term carry-over effect (Aguirre, 2007), or, in its purest form in same versus different tests, as rapid fMR-adaptation (Grill-Spector and Malach, 2001). To reveal the possible contribution of short-term stimulus similarity-based mechanisms behind the sensitivities measured by our acoustic and identity tests, we separated the effect of stimulus similarity to the directly preceding voice stimulus from longer-lasting effects. Extensive regions were found in and around the bilateral middle/posterior STS (BA 21, 22) in both the acoustic and the identity tests. These were the only brain regions that were found to be differentially active in both main tests. Neural sensitivity in the right STS, as measured in the voice-identity test but not in the voice-acoustic test, was even found to covary with person identification performance. Furthermore, we

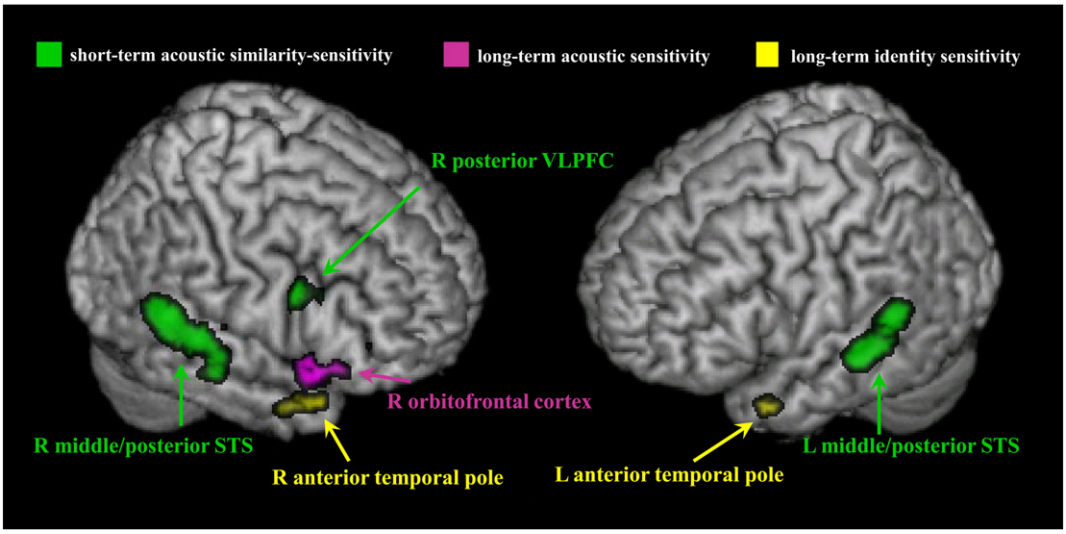

Fig. 4. Sagittal views display short-term effect (green), thresholded at $t>4$; long-term acoustic sensitivity effect (purple) and long-term identity sensitivity effect (yellow), thresholded at $t(23)>3$ and masked by the corresponding main analyses thresholded at $t(23)>4$.

demonstrated that these temporal regions were involved in short-term similarity processing. These regions are very similar to the temporal voice areas (Belin et al., 2000) that have been found to respond differentially to voice stimuli in healthy subjects but not in autism (Gervais et al., 2004). The present findings confirm short-term stimulus similarity sensitivity in the voice-tuned middle/posterior STS, and that better short-term sensitivity may lead to better voice recognition performance.

Only one further region, the right VLPFC, showed sensitivity to short-term stimulus similarity processing. This posterior ventrolateral prefrontal region on the inferior bank of the inferior frontal sulcus (BA 44, 45) was found bilaterally, but with strong right-hemisphere dominance in the main acoustic but not in the identity sensitivity test. The right ventrolateral prefrontal region, just as the bilateral STS, was also differentially sensitive to voice stimuli in general. This prefrontal region involves Broca's area in the left hemisphere and is known to be crucial for linguistic processing. Its right-hemisphere counterpart has been shown to be more active in nonverbal memory tasks with environmental sounds (Opitz et al., 2000). Additionally, right ventrolateral prefrontal regions have been proposed to be involved in voice analysis in both primates (Romanski et al., 2005) and humans (Fecteau et al., 2005). Our findings suggest that this right VLPFC region, similarly to the voice-tuned STS regions, participates in short-term voice-acoustic change detection.

Short-term sensitivity to acoustic similarities between voice stimuli in the middle/posterior STS and in the VLPFC confirms these areas' responsiveness to acoustic changes within the stimulus set. However, an area's involvement in a short-term cortical mechanism does not exclude its involvement in mechanisms based on long-term representations. The STS is a region that is highly heterogeneous functionally (e.g., Beauchamp et al., 2004), and the middle/posterior STS was proposed to be crucial for different stages of voice-identity processing (von Kriegstein and Giraud, 2004; Warren et al., 2006). Recent findings also suggested VLPFC involvement in the representation of long-term stored objects (Latinus et al., 2009). It was therefore somewhat surprising that in our confirmatory analyses we found no evidence suggesting that STS or VLPFC regions would mediate long-term voice memory (except for a small right posterior medial temporal region close to the junction of BA 20, 37 and 41). One explanation is that, contrary to these earlier claims, the neural substrates of long-lasting object space representations, including acoustic mean or category mean voice representations, are located elsewhere. Alternatively, it is possible that long-term effects were indeed present in the STS and VLPFC, but were masked by co-existing short-term effects in the present design. Further investigations are needed to resolve this issue.

Table 5
List of regions found in the long-term acoustic and identity sensitivity tests.

	BA	*x*	*y*	*z*	*T*	*p* (corr.)
Long-term acoustic sensitivity						
R posterior medial temporal cortex	21/22	40	−26	0	4.79	0.012
R orbitofrontal cortex/anterior insula	47	44	18	−16	4.79	0.002
R medial orbitofrontal cortex	11	8	18	−16	3.53	0.009
Long-term identity sensitivity						
L deep posterior STS	39	−30	−62	24	5.09	<0.001
R anterior temporal pole	21/38	48	18	−28	4.39	0.002
L anterior temporal pole	21/38	−52	14	−28	4.73	0.001
L amygdala	–	−20	−8	−18	3.02	0.034

A single peak is shown per region. Long-term sensitivity contrasts were thresholded at $t(23)>3$. Small-volume correction was based on clusters from the corresponding main analyses, thresholded at $t(23)>4$.

Table 6
Overlapping regions in main analyses and additional independent tests.

	Short%	Voice%	Face%
Acoustic sensitivity			
R anterior/middle/posterior STS	29	89	28
L anterior/middle/posterior STS	4	95	<1
R orbitofrontal cortex/anterior insula			
R medial orbitofrontal cortex			
L orbitofrontal cortex/anterior insula		3	
R posterior VLPFC	4	12	
L posterior VLPFC			
Identity sensitivity			
R middle/posterior STS	14	92	
L middle/posterior STS	<1	71	
L deep posterior STS			
R anterior temporal pole		100	
L anterior temporal pole		97	
L amygdala		2	90

The columns short%, voice% and face% show the proportion of voxels in each acoustically sensitive or identity-sensitive cluster that were also differentially active in the (1) short-term effect test (minimal distance<maximal distance), (2) voice area localizer (non-vocal stimuli<voices) and (3) face area localizer (scrambled objects<faces) respectively (thresholded at $t(23)<4$).

Voice-acoustics space representation

The acoustic sensitivity test contrasted acoustically central and peripheral stimuli. This contrast tested the hypothesis that during listening to stimuli from a voice-morph continuum, an implicit prototype-formation process takes place in the voice-acoustics space, resulting in the creation of a long-term stored 'acoustic mean voice' representation and hence in long-lasting neural sharpening for acoustically central stimuli. This hypothesis was confirmed. Although some regions found in this test, including the STS and the VLPFC, were shown to be biased by covarying short-term similarity, other regions, including the bilateral orbitofrontal cortex extending to the anterior insula (BA 47, 11) did not exhibit short-term stimulus similarity sensitivity. Furthermore, there was no difference in presentation frequency between central and peripheral stimuli at either training or test to motivate a long-term bias without an 'acoustic mean voice' representation. So the orbital/insular cortex activity in the acoustic sensitivity test can best be described as long-term stimulus similarity sensitivity. This claim was further supported by a confirmatory test looking for long-term acoustic space sensitivity: the bilateral orbital/insular cortex was found in this test but the STS and VLPFC regions were not (except for a small right posterior medial temporal region close to the junction of BA 20, 37 and 41). The anterior insula has been implicated in the processing of sound and more specifically speech information (Wong et al., 2004), and it has also been proposed to possibly play a role in processing vocal paralinguistic information such as vocal emotion or vocal identity (Remedios et al., 2009; Watson, 2009). Our findings do not confirm that the insula handles vocal identity information; instead, the response reduction for voice stimuli that were most versus least similar to the acoustic mean of all preceding stimuli suggests that 'acoustic mean voice' representations exist and that they may be created in the orbital/insular cortex. This acoustic mean voice seems to be created independently from any representation of trained voice identities. Our results thus show that a perceptual typicality-based organization arises automatically for voice representations, similarly to what has been reported for faces (Loffler et al., 2005).

Voice-identity space representation

We hypothesized that voice analysis at the stage of identity processing is also supported by neural representations of the stimulus space in which long-term stored typical values are coded more sparsely than atypical values. Our findings support this hypothesis. We found response reduction for identity-internal versus identity-external stimuli in regions (including the voice-tuned ATP, the amygdala and the deep posterior STS) that showed no response reduction for the same stimulus contrast when it was free from the identity manipulation. The response pattern of regions with an identity effect but no acoustic effect can be explained as a long-term neural sharpening effect induced by the explicit categorization feedback during training. These results and the finding of significant covariation between neural identity sensitivity and behavioural sensitivity in almost all identity-sensitive clusters (except for the left ATP) therefore argue for the existence of a neural voice-identity space and of 'trained category mean voice' representations. This explanation is further supported by our additional analyses that confirmed the presence of long-term identity representations but found no effects of short-term stimulus similarity sensitivity in the bilateral ATP, the left deep posterior STS and the left amygdala.

The finding of voice-identity representations in the anterior temporal pole confirms existing reports about the anterior temporal lobe's role in voice-identity processing (Nakamura et al., 2001; von Kriegstein et al., 2003, 2005; Belin and Zatorre, 2003; Lattner et al., 2005, Sokhi et al., 2005) and seems to support the idea that this region corresponds to the unimodal voice recognition module in the model proposed by Belin and colleagues (Belin et al., 2004; Campanella and Belin, 2007). The novelty of our ATP finding is that we demonstrated this voice-tuned region's involvement in the representation of a category mean-centered voice-identity space, and showed the effect of individual identity space sensitivity on voice recognition performance. Anterior temporal lobe regions, however, have also been shown to be involved in person identity recognition for different modalities (von Kriegstein and Giraud, 2006), in the multimodal integration of person information (for a review, see Olson et al., 2007; but see also Turk et al., 2005) and in the 'what' processing pathway (Scott and Johnsrude, 2003; Belin and Zatorre, 2003). Furthermore, clinical reports suggest that voice-identity recognition and supramodal person identity recognition can be selectively impaired after degeneration of the anterior temporal lobe (e.g., Hailstone et al., 2010). The location of anterior temporal lobe findings in the present study [48, 18, −28; −52, 14, −28] is in-between previously reported coordinates of supramodal person recognition in the temporal pole (slightly superior to e.g., [46, 16, −40; −44, 16, −40] in Sugiura et al., 2006) and those of unimodal voice recognition in the anterior STG/STS (slightly inferior and anterior to e.g., [57, 9, −21; 54,12,−15; 48, 6, −18] in von Kriegstein et al., 2003, or to [58, 2, −8] in Belin and Zatorre, 2003). We therefore cannot exclude the possibility that our anterior temporal pole findings correspond instead to a different stage in Belin and colleagues' model (Belin et al., 2004; Campanella and Belin, 2007), namely to the supramodal person identification stage. Note that other, non-neuroimaging research has also suggested that there may be distinct acoustic, unimodal and supramodal steps in person identification (Ellis et al., 1997; Neuner and Schweinberger, 2000). Further clarification of the distinction between unimodal and supramodal processing regions within the anterior temporal lobe will probably require a direct experimental comparison of these person identification steps. Furthermore, earlier studies have created some uncertainty with respect to whether voice-identity processing in ATP regions is restricted only to the right hemisphere or is present bilaterally. Our results, although remaining inconclusive, offer a better view on this issue: we found identity sensitivity in the ATP bilaterally, but voice recognition was shown to reflect only the right ATP sensitivity.

Voice-identity representations were also found in a left deep posterior STS region (BA 39) in our study. Our knowledge about the possible role in object recognition of the deep posterior STS region is very limited. Brodmann area 39 is often considered to be part of the Wernicke's area (Wise et al., 2001), an important centre for speech processing. Sensitivity to biological motion (Grossman et al., 2000) and audiovisual integration of voice and face information (Kreifelts et al., 2007) has been found for close but more lateral parts of the posterior superior temporal gyrus. Additionally, the left but not the right angular gyrus and medial parietal regions were found to be sensitive to voice familiarity in a prosopagnosic patient with bilateral damage (Arnott et al., 2008). Neighbouring, but more medial brain regions of the precuneus/retrosplenial cortex have shown sensitivity to person familiarity (Shah et al., 2001), and have been proposed as possible loci of cross-modal person identity nodes (Campanella and Belin, 2007). We suggest that this deep posterior STS region close to the angular gyrus and the intraparietal sulcus may contribute to a modality-nonspecific person identity representation.

We also found the identity effect in the amygdala, with significant covariation between neural and behavioural sensitivity. The amygdala activity persisted in our confirmatory long-term identity effect test. The amygdala has been suggested to be involved in the processing of socially relevant stimuli such as faces (Breiter et al., 1996; Morris et al., 1996; Whalen et al., 1998) and voices (Fecteau et al., 2007; Campanella and Belin, 2007), but the specific role of this region is debated. Belin et al. (2004, Campanella and Belin, 2007) proposed that during voice analysis distinct neural processing streams are responsible for the recognition of speech categories, emotions and identities,

and that the amygdala is responsible for vocal emotion processing. But recent findings suggest an important role for the amygdala also in the processing of emotionally neutral face stimuli both in monkeys (Gothard et al., 2007) and in humans (Kleinhans et al., 2009). Recently, Kleinhans et al. (2009) found reduced neural habituation in the amygdala for neutral facial stimuli in autism, a complex developmental disorder characterized by deficits in social interaction. It has also been proposed that there is a paralimbic network including both the amygdala and the anterior temporal pole which is specialized for person identification (Olson et al., 2007). The amygdala seems to be tuned to emotional stimuli more than to neutral stimuli, and to faces more than to voices, but our results indicate that it nevertheless participates in the representation of person identity given neutral voice stimuli. This finding is in line with psychophysical and electrophysiological evidence suggesting that voice analysis modules are not fully independent (Campanella and Belin, 2007), for example, speech perception has been shown to influence voice perception (Remez et al., 1997; Perrachione and Wong, 2007; Perrachione et al., 2010), and vocal emotions have been shown to modulate early sensory processing (Spreckelmeyer et al., 2009). A better understanding of the amygdala's role will clearly help to clarify the interplay of different voice analysis modules and the separability of neural substrates for different object types conveyed by voice and face stimuli.

Interestingly, no regions with identity sensitivity were found when, in an additional test, listeners had to perform a voice-irrelevant word repetition detection task. This indicates that identity sensitivity requires the presence of a relevant task, confirming earlier reports that specified similar brain regions responsible for voice-identity processing by manipulating task relevance but not stimuli (von Kriegstein et al., 2003, von Kriegstein and Giraud, 2004).

Flexibility in voice representation

Finally, this study demonstrates the dynamics of voice processing. Voices, although carrying information about an anatomically defined vocal tract, are modulated by less permanent factors such as language, dialect, speech style, emotions, volume, speed, health situation etc. that are known to influence talker identification (Nolan, 1997; Perrachione and Wong, 2007; Perrachione et al., 2010). Indeed, speakers dynamically tune their voices to the situation they find themselves in (e.g., in phonetic convergence, speakers tend to talk more like their interlocutors as a conversation progresses; Pardo, 2006). Therefore, the human perceptual ability to adapt flexibly to dynamic object changes (Kourtzi and DiCarlo, 2006; Jiang et al., 2007) is especially important for voice stimuli (cf. Schweinberger et al., 2008). Consequently, neural representations of voice identities need to be highly plastic to support voice recognition. Our findings demonstrate listeners' flexibility in learning and representing voice identities. On the first week of the experiment, listeners rapidly learned a new voice-identity and then, when a week later a different voice morph interval was associated with the same identity, they dynamically adapted their representations. Neural sharpening for a long-term stored 'category mean voice' followed the trained shift and therefore retuned the neural representation of the voice-identity space.

Conclusion

Our results are in line with the proposal that voice recognition is supported by a categorical level of processing that is anatomically separable from voice structural processing (Belin et al., 2004). Our findings also confirm that there exist dissociable neural mechanisms for short-interval versus long-interval fMRI repetition suppression (Epstein et al., 2008). More specifically, we have argued for the existence of dynamic, long-lasting 'mean voice' representations at both voice-acoustic and voice-identity stages of processing. In accordance with recent findings in behavioural studies of voice processing (Papcun et al., 1989; Mullennix et al., 2009, Bruckert et al., 2010) and with those in the face processing domain (Loffler et al., 2005), our demonstrations of neural 'mean voice' representations constitute the first neuroimaging evidence that voice representations are centered around prototypes in long-term memory.

Acknowledgments

This study was conducted as part of AA's PhD project, funded by the Max Planck Society. KMP was funded by a FCT grant IBB/CBME, LA, FEDER/POCI 2010. We thank two anonymous reviewers for their constructive comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.05.048.

References

Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. Neuroimage 35, 1480–1494.
Arnott, S.R., Heywood, C.A., Kentridge, R.W., Goodale, M.A., 2008. Voice recognition and the posterior cingulate: an fMRI study of prosopagnosia. J. Neuropsychol. 2 (1), 269–286.
Ashby, F.G., Maddox, W.T., 2005. Human category learning. Annu. Rev. Psychol. 56, 149–178.
Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. 7 (11), 1190–1192.
Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.
Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport 14, 2105–2109.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal lobe responses to vocal sounds. Cogn. Brain Res. 13, 17–26.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Belizaire, G., Fillion-Bilodeau, S., Chartrand, J.P., Bertrand-Gauvin, C., Belin, P., 2007. Cerebral response to 'voiceness': a functional magnetic resonance imaging study. NeuroReport 18, 29–33.
Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., 2007. Automatic speech recognition and speech variability: a review. Speech Commun. 49 (10–11), 763–786.
Boersma, P., Weenink, D., 2005. Praat: Doing Phonetics by Computer. http://www.praat.org/. Computer programme.
Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., Rosen, B.R., 1996. Response and habituation of the human amygdala during visual processing of facial expression. Neuron 17, 875–887.
Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G.A., Kawahara, H., Belin, P., 2010. Vocal attractiveness increases by averaging. Curr. Biol. 20, 116–120.
Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. Trends Cogn. Sci. 11, 535–543.
Ellis, A.W., Young, A.W., Critchley, E.M.R., 1989. Loss of memory for people following temporal lobe damage. Brain 112, 1469–1483.
Ellis, H.D., Jones, D.M., Mosdell, N., 1997. Intra- and inter-modal repetition priming of familiar faces and voices. Br. J. Psychol. 88 (1), 143–156.
Epstein, R.A., Parker, W.E., Feiler, A.M., 2008. Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. J. Neurophysiol. 99, 2877–2886.
Fecteau, S., Armony, J.L., Joanette, Y., Belin, P., 2005. Sensitivity to voice in human prefrontal cortex. J. Neurophysiol. 94, 2251–2254.
Fecteau, S., Belin, P., Joanette, Y., Armony, J.L., 2007. Amygdala responses to nonlinguistic emotional vocalizations. Neuroimage 36 (2), 480–487.
Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), 2007. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, London.
Gainotti, G., Barbier, A., Marra, C., 2003. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. Brain 126, 792–803.
Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., Duchaine, B., 2009a. Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 47, 123–131.
Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Friston, K.J., 2009b. Repetition suppression and plasticity in the human brain. Neuroimage 48 (1), 269–279.
Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthelemy, C., Brunelle, F., Samson, Y., Zilbovicius, M., 2004. Abnormal cortical voice processing in autism. Nat. Neurosci. 7 (8), 801–802.
Gothard, K.M., Battaglia, F.P., Eickson, C.A., Spitler, K.M., Amaral, D.G., 2007. Neural responses to facial expression and face identity in the monkey amygdala. J. Neurophysiol. 97, 1671–1683.
Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. 107, 293–321 (Amsterdam).

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., Blake, R., 2000. Brain areas involved in perception of biological motion. J. Cogn. Neurosci. 12 (5), 711–720.
Hailstone, J.C., Crutch, S.J., Vestergaard, M.D., Patterson, R.D., Warren, J.D., 2010. Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia 48 (4), 1104–1114.
Hoffman, K.L., Logothetis, N.K., 2009. Corical mechanisms of sensory learning and object recognition. Philos. Trans. R. Soc. B. 364, 321–329.
Jiang, X., Bradley, E., Rini, R.A., Zeffiro, T., VanMeter, J., Riesenhuber, M., 2007. Categorization training results in shape- and category-selective human neural plasticity. Neuron 53, 891–903.
Kawahara, H., 2006. STRAIGHT, exploration of the other aspect of VOCODER: perceptually isomorphic decomposition of speech sounds. Acoust. Sci. Technol. 27 (6), 349–353.
Kleinhans, N.M., Johnson, L.C., Richards, T., Mahurin, R., Greenson, J., Dawson, G., Aylward, E., 2009. Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. Am. J. Psychiatry 166, 467–475.
Kourtzi, Z., DiCarlo, J.J., 2006. Learning and neural plasticity in visual object recognition. Curr. Opin. Neurobiol. 16, 1–7.
Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. Neuroimage 37 (4), 1445–1456.
Latinus, M., Crabbe, F., Belin, P., 2009. fMRI investigations of voice identity perception. Organization for Human Brain Mapping 2009 Annual Meeting, July 2009: NeuroImage, 47(Supplement 1, p. S156.
Lattner, S., Meyer, M.E., Friederici, A.D., 2005. Voice perception: sex, pitch, and the right hemisphere. Hum. Brain Mapp. 24, 11–20.
Lloyd, R.L., Kling, A.S., 1988. Amygdaloid electrical activity in response to conspecific calls in squirrel monkey: influence of environmental setting cortical inputs and recording site. In: Newman, J.D. (Ed.), The Physiological Control of Mammalian Vocalization. Plenum Press, New York, pp. 137–151.
Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. FMRI evidence for the neural representation of faces. Nat. Neurosci. 8 (10), 1386–1390.
Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., Dolan, R.J., 1996. A differential neural response in the human amygdala to fearful and happy facial expressions. Nature 383, 812–815.
Mullennix, J.W., Ross, A., Smith, C., Kuykendall, K., Conard, J., Barb, S., 2009. Typicality effects on memory for voice: implications for earwitness testimony. Appl. Cogn. Psychol. doi:10.1002/acp.1635.
Myers, E.B., 2007. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: an fMRI investigation. Neuropsychologia 45, 1463–1473.
Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatan, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., Kojima, S., 2001. Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia 39, 1047–1054.
Neuner, F., Schweinberger, S.R., 2000. Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn. 44 (3), 342–366.
Nolan, F., 1997. Speaker recognition and forensic phonetics. In: Hardcastle, W., Laver, J. (Eds.), A Handbook of Phonetic Science. Blackwell, Oxford, pp. 744–766.
Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. Brain 130, 1718–1731.
Opitz, B., Mecklinger, A., Friederici, A.D., 2000. Functional asymmetry of human prefrontal cortex: encoding and retrieval of verbally and nonverbally coded information. Learn. Mem. 7, 85–96.
Papcun, G., Kreiman, J., Davis, A., 1989. Long-term memory for unfamiliar voices. J. Acoust. Soc. Am. 85, 913–925.
Pardo, J.S., 2006. On phonetic convergence during conversational interaction. J. Acoust. Soc. Am. 119 (4), 2382–2393.
Pernet, C., Charest, I., Bélizaire, G., Zatorre, R.J., Belin, P., 2007. The temporal voice areas: spatial characterization and variability. 13th International Conference on Functional Mapping of the Human Brain, Chicago, USA: NeuroImage, 36. Supp1.
Perrachione, T.K., Chiao, J.Y., Wong, P.C.M., 2010. Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. Cognition 114 (1), 42–55.
Perrachione, T.K., Wong, P.C.M., 2007. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. Neuropsychologia 45 (8), 1899–1910.
Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. Nat. Neurosci. 11 (3), 367–374.
Potter, R.K., Steinberg, J.C., 1950. Toward the specification of speech. J. Acoust. Soc. Am. 22, 807–820.
Remedios, R., Logothetis, N.K., Kayser, C., 2009. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J. Neurosci. 29, 1034–1045.
Remez, R.E., Fellowes, J.M., Rubin, P.E., 1997. Talker identification based on phonetic information. J. Exp. Psychol. Hum. Percept. Perform. 23, 651–666.
Romanski, L.M., Averbeck, B.B., Diltz, M., 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J. Neurophysiol. 93, 734–747.
Schweinberger, S.R., Casper, C., Hauthal, N., Kaufmann, J.M., Kawahara, H., Kloth, N., Robertson, D.M.C., Simpson, A.P., Zaske, R., 2008. Auditory adaptation in voice perception. Curr. Biol. 18 (9), 684–688.
Schweinberger, S.R., Herholz, A., Sommer, W., 1997. Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. J. Speech Lang. Hear. Res. 40, 453–463.
Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.
Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., Fink, G.R., 2001. The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 124 (4), 804–815.
Sokhi, D.S., Hunter, M.D., Wilkinson, I.D., Woodruff, P.W.R., 2005. Male and female voices activate distinct regions in the male brain. Neuroimage 27, 572–578.
Spreckelmeyer, K.N., Kutas, M., Urbach, T., Altenmuller, E., Munte, T.F., 2009. Neural processing of vocal emotion and identity. Brain Cogn. 69 (1), 121–126.
Stevens, A.A., 2004. Dissociating the cortical basis of memory for voices, words and tones. Cogn. Brain Res. 18, 162–171.
Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y., Maeda, Y., Matsue, Y., Fukuda, H., Kawashima, R., 2006. Cortical mechanisms of person representation: recognition of famous and personally familiar names. Neuroimage 31 (2), 853–860.
Turk, D.J., Rosenblum, A.C., Gazzaniga, M.S., Macrae, C.N., 2005. Seeing John Malkovich: the neural substrates of person categorization. Neuroimage 24, 1147–1153.
Van Lancker, D., Kreiman, J., 1987. Voice discrimination and recognition are separate abilities. Neuropsychologia 25, 829–834.
Van Lancker, D., Kreiman, J., Cummings, J., 1989. Voice perception deficits: neuroanatomical correlates of phonagnosia. J. Clin. Exp. Neuropsychol. 11, 665–674.
Van Lancker, D.R., Cummings, J.L., Kreiman, J., Dobkin, B.H., 1988. Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex 24, 195–209.
von Kriegstein, K., Eger, E., Kleinschmidt, A., Giraud, A.L., 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. Cogn. Brain Res. 17, 48–55.
von Kriegstein, K., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22, 948–955.
von Kriegstein, K., Giraud, A.L., 2006. Implicit multisensory associations influence voice recognition. PLoS Biol. 4 (10), e326.
von Kriegstein, K., Kleinschmidt, A., Sterzer, P., Giraud, A.L., 2005. Interaction of face and voice areas during speaker recognition. J. Cogn. Neurosci. 17, 367–376.
Warren, J., Scott, S., Price, C., Griffiths, T., 2006. Human brain mechanisms for the early analysis of voices. Neuroimage 31, 1389–1397.
Watson, R., 2009. Selectivity for conspecific vocalizations within the primate insular cortex. J. Neurosci. 29 (21), 6769–6770.
Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., Jenike, M.A., 1998. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. 18, 411–418.
Wise, R.J., Scott, S., Blank, S.C., Mummery, C.J., Murphey, K., Warburton, E.A., 2001. Separate neural subsystems within 'Wernicke's area'. Brain 13 (1), 83–95.
Wong, P.C., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J. Neurosci. 24, 9153–9160.