ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 32 (2006) 368 – 375

Processing the acoustic effect of size in speech sounds

K. von Kriegstein,[a,b,c] J.D. Warren,[a,d] D.T. Ives,[b] R.D. Patterson,[b] and T.D. Griffiths[a,b,c,*]

[a]*Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK*
[b]*Centre for the Neural Basis of Hearing, Physiology Department, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK*
[c]*Auditory Group, Medical School, Framlington Place, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE2 4HH, UK*
[d]*Dementia Research Centre, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK*

Received 19 September 2005; revised 21 February 2006; accepted 27 February 2006
Available online 27 April 2006

The length of a vocal tract is reflected in the sound it is producing. The length of the vocal tract is correlated with body size and humans are very good at making size judgments based on the acoustic effect of vocal tract length only. Here we investigate the underlying mechanism for processing this main auditory cue to size information in the human brain. Sensory encoding of the acoustic effect of vocal tract length (VTL) depends on a time-stabilized spectral scaling mechanism that is independent of glottal pulse rate (GPR, or voice pitch); we provide evidence that a potential neural correlate for such a mechanism exists in the medial geniculate body (MGB). The perception of the acoustic effect of speaker size is influenced by GPR suggesting an interaction between VTL and GPR processing; such an interaction occurs only at the level of non-primary auditory cortex in planum temporale and anterior superior temporal gyrus. Our findings support a two-stage model for the processing of size information in speech based on an initial stage of sensory analysis as early as MGB, and a neural correlate of the perception of source size in non-primary auditory cortex.

Keywords: Auditory; fMRI; Size; Pitch; Medial geniculate body; Voice

Determining the size of a sound source is a fundamental perceptual task for humans and other species (Fairchild, 1981; Reby et al., 2005; Smith et al., 2005) but the brain regions that analyze the size information in sounds have not been defined.

The size of a sound source is reflected in the sound it produces. In the case of human speech, the vocal tract acts as a filter to the sound produced by the vocal chords. The length of the vocal tract is highly correlated with speaker size (Fitch and Giedd, 1999) and so, the acoustic effect of vocal tract length (VTL) is an important sensory cue to the size of the speaker (Fitch and Giedd, 1999; Smith et al., 2005). The vocal tract filter introduces peaks in the magnitude spectrum at frequencies associated with the sizes of the resonances of the vocal tract, and it attenuates energy at other frequencies. The peaks are the so-called formants of speech, and the formant frequency ratios provide information about the shape of the vocal tract when producing a specific speech sound (i.e., an /a/ has different formant ratios than an /u/). For a given vocal tract configuration, there is a well-defined change in the sound (and its internal representation) when the length of the vocal tract changes (i.e., when a speaker with a different size says the same speech sound). Fig. 1 shows the time-stabilized, neural firing pattern, or auditory image (Patterson et al., 1995), of the vowel /a/ for a male and a female speaker (Irino and Patterson, 2002). The shorter vocal tract of the female speaker results in formants that have higher frequencies and faster decays. These changes are independent of the pitch produced by the vocal chords, which is the same for the two voices in this example. The sensory correlate of a change in the size of the source could be the formant frequency shift, the reduction in decay duration, or most likely both (Irino and Patterson, 2002).

The perception of size is also influenced by glottal pulse rate (GPR), which determines the pitch of the voice. GPR is not tightly correlated with body size (Kunzel, 1989) but men have longer and heavier vocal cords than women, resulting in a generally lower glottal pulse rate (GPR). Thus, although VTL is the main cue for processing size, a voice emanating from the same length of vocal tract can sound bigger if the glottal pulse rate is lower and smaller if the glottal pulse rate is higher (Smith and Patterson, 2005; Smith et al., 2005).

The analysis of size information is not only important in relation to the estimation of the size of the speaker; it also poses an invariance problem. When a specific vowel is produced by speakers with markedly different sizes, they are, nevertheless, still perceived as the same vowel despite the difference in the length of the vocal tract. This suggests that the auditory system can segregate information about the length of the vocal-tract from information about its shape (Irino and Patterson, 2002).

Abbreviations: GPR, glottal pulse rate; IC, inferior colliculus; MGB, medial geniculate body; PAC, primary auditory cortex; PT, planum temporale; STG, superior temporal gyrus; STS, superior temporal sulcus; VTL, acoustic effect of vocal tract length.

* Corresponding author. Auditory Group, University of Newcastle Medical School, Framlington Place, Newcastle-upon-Tyne NE2 4HH, UK. Fax: +44 191 222 6706.
E-mail address: t.d.griffiths@ncl.ac.uk (T.D. Griffiths).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.02.045

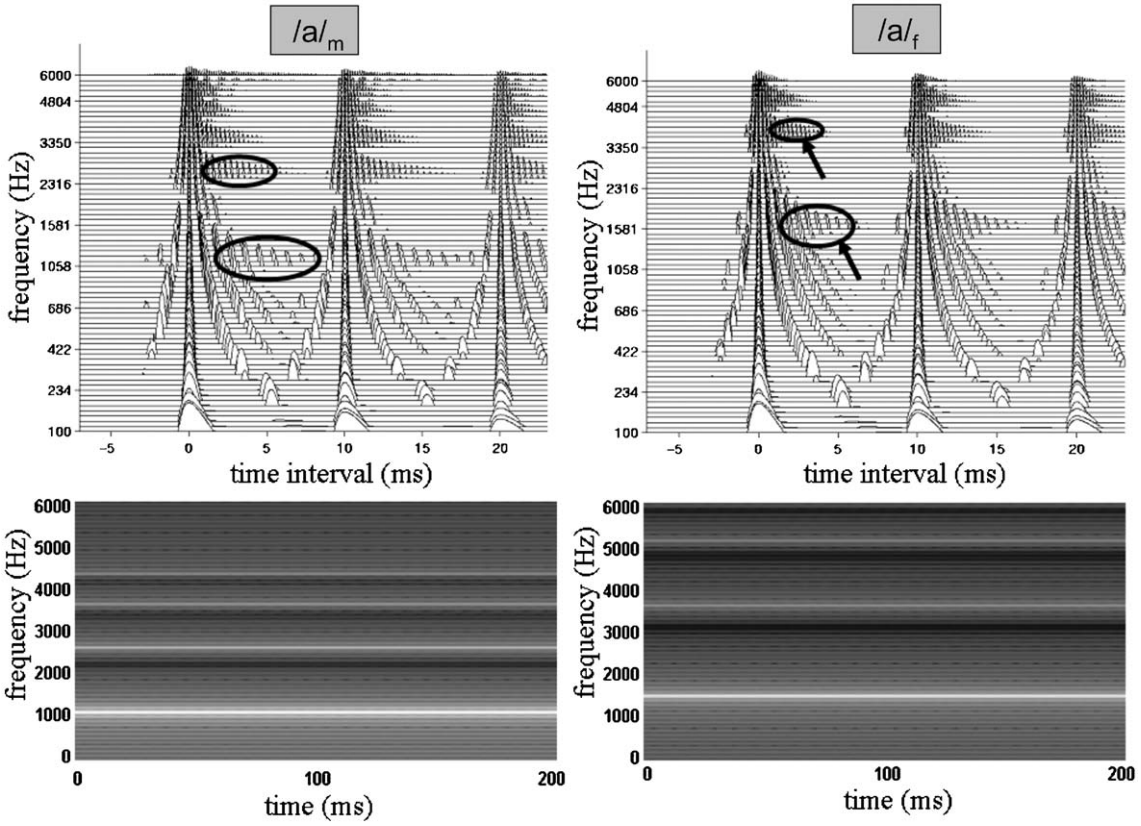

Fig. 1. Top row: auditory images for the vowel /a/ synthesized for male and female speakers. Both images are shown with the same glottal pulse rate (100 Hz). Note that the reduction in VTL causes an upward shift of the formant peaks and a decrease in the duration of the formants. Bottom row: spectrograms of the same vowels.

A priori, auditory size analysis is likely to involve processes in areas that support the sensory representation of acoustic cues to source size, and subsequent processes in areas that correspond to the neural correlate of the perceived size of auditory objects. For the analysis of pitch, human functional imaging studies (Griffiths, 2005; Griffiths et al., 2001) have identified a sensory representation of time- and frequency-domain properties in the ascending pathway to primary auditory cortex (PAC) (Griffiths et al., 2001), while a neural correlate of the percept of pitch has been demonstrated in non-primary auditory cortex (Patterson et al., 2002; Penagos et al., 2004). Here we investigate whether an operationally equivalent process can be identified for VTL analysis. By that we mean an area before PAC responding to the sensory information corresponding to VTL and an area after PAC responding to the percept of changes in VTL.

We manipulated the VTL and GPR of a set of syllables by using the signal processing package referred to as STRAIGHT (Kawahara and Irino, 2005; Kawahara et al., 1999). Although VTL within one speaker is relatively constant, speakers can modify their VTL a little by lip rounding and by raising or lowering the larynx. Both effects change the positions of the formant frequencies and so alter the perceived size of the speaker. The variability introduced by these factors reduces the correlation between formant frequency and height in adults, and when the range of heights is limited and/or the sample size is relatively small, the correlation can be unreliable (Gonzalez, 2004; Rendall et al., 2005). We therefore manipulated the stimuli within a broad range of VTL and GPR values, which were defined previously in an orthogonal space (Smith et al., 2005).

We hypothesized the existence of a neural substrate for the sensory representation of the acoustic effect of vocal tract length early in the auditory pathway, and subsequent cortical substrates that are candidate areas for the perceptual analysis of the size of the speaker's vocal tract. Analysis of VTL information requires a time-stabilized auditory image (Irino and Patterson, 2002), and therefore cannot be extracted prior to the analysis of pitch information. The extraction of pitch information is not completed before the inferior colliculus (IC) (Griffiths et al., 2001; Langner and Schreiner, 1988). Accordingly, we hypothesized that the earliest sensory representation of VTL could occur one step after the IC, at the Medial geniculate body (MGB). Previous studies have shown regions sensitive to the vocal aspects of speech in bilateral STS (Belin et al., 2000; von Kriegstein and Giraud, 2004). These studies have used natural voice stimuli, which vary across various dimensions, such as phonetic idiosyncrasies and talking speed, as well as GPR and VTL. Here we modulated only the two latter aspects of voice processing, and thus we expect to find that the regions activated are a subset of the voice selective regions previously described. The VTL information is represented in the spectral profile of the auditory image, and previous studies suggest that planum temporale (PT) and anterior superior temporal sulcus (aSTS) are involved in the analysis of spectral envelopes (Warren et al., 2005). Accordingly, we hypothesized that PT and aSTS are likely candidates for the processing of the VTL contribution to the perception of speaker size. Furthermore, as described above, our behavioral data (Smith and Patterson, 2005) show that GPR information influences the perception of speaker size through VTL information. Thus, VTL processing is not entirely independent of pitch processing. We

hypothesized therefore that activation in regions or subset of regions responsive to VTL information will be modulated by changes in GPR.

Material and methods

Subjects

Seventeen subjects (12 right handed, 5 left handed; 5 female, 12 male; aged 19–41 years) participated in the study. All subjects gave informed consent, and the experiment was carried out with the approval of the Institute of Neurology Ethics Committee, London. No subject had any history of audiological or neurological disorder, and all had normal structural MRI brain scans.

Stimuli

The stimuli were taken from a set of 180 syllables used in a previous study (Ives et al., 2005). There were 90 consonant-vowel and 90 vowel-consonant syllables recorded from one speaker with 16-bit resolution and a 48-kHz sample rate. The level of the syllables was normalized such that the RMS value of the vowel region was set to a common value; this approximately equates the loudness of the syllables. The position of the syllable within their file was also adjusted so that the perceptual-center (P-center) occurred at the same time relative to file onset. This P-center adjustment ensured that when a set of the syllables is played in a sequence, the syllables are perceived to proceed at a regular pace; an irregular rhythm produces an unwanted distraction. The algorithm for finding the P-centers focuses on vowel onsets; the details are described in Marcus (1981) and Scott (1993). After P-center correction, the syllable file was padded with silence out to 683 ms, the duration of the longest syllable. The original speaker had a GPR of between 112 Hz and 130 Hz which is typical for an average adult male. The height of the speaker was 173 cm which gives a VTL of about 15.2 cm (Fitch and Giedd, 1999).

Digital resynthesis was carried out using STRAIGHT (Kawahara and Irino, 2005; Kawahara et al., 1999) to produce stimuli with specified values of GPR and VTL. STRAIGHT analyses voiced speech into a sequence of glottal pulse times and a sequence of pitch-synchronous spectral frames, and then after independent manipulation of the GPR and VTL sequences, it resynthesizes the speech with the new values. Utterances recorded from a man can be transformed to sound like a women or a child. The combinations of GPR and VTL were all derived from the perceptual space characterized previously by Smith et al. (2005). The VTL ranged from 10.1 to 21.7 cm; the GPR ranged from 50 to 158 Hz. The classic data of Peterson and Barney (1952) for men, women, and children are characterized by a set of three ellipses centered at VTL and GPR values of: 15.2 ± 1.3 cm and 131 ± 19 Hz (men); 13.2 ± 1.2 cm and 223 ± 27 Hz (women); and 11.3 ± 1.3 cm and 264 ± 25 Hz (children). The ellipses are shown in Fig. 1 of Smith et al. (2005). Each of the stimulus files produces the perception of a person speaking a specific syllable, despite the fact that the range of VTLs included values that would be associated with speakers having sizes beyond the normal range in the human population. Extending the range of VTL and GPR beyond that normally experienced has very little effect on vowel recognition (Smith et al., 2005). This has been interpreted to mean that the sensory decoding mechanism includes a generic normalization process and that the mechanism does not depend on learning from experience. Examples of the stimuli are available on our webpage (http://www.mrc-cbupdn.cam.ac.uk/cnbh/).

Experimental design

The experiment was carried out in two sessions using stimuli from different regions of the GPR-VTL space. In one session, three high GPR values (120, 138, 158 Hz) were combined with three short VTLs (10.1, 11.3, 12.7 cm) to produce nine combinations of GPR and VTL. In the other session, three low GPR values (50, 57, 66 Hz) were combined with three long VTLs (16.9, 19.2 and 21.7 cm) to produce another nine combinations of GPR and VTL. The values of VTL were chosen to produce equal magnitude changes between the different values, as were the values of GPR. The two conditions in which VTL varied had the largest range of perceived size due to the higher influence of VTL on the size percept in comparison to GPR. Between these two conditions, there was no overall difference in perceived change in size because both VTL and GPR varied randomly and thus were on some trials congruent (e.g., GPR decreases as VTL increases; eliciting a larger range of size percept than if GPR were fixed) and on some trials not congruent (e.g., GPR increases as VTL increases; eliciting a smaller range of size percept than if GPR were fixed). During each session, subjects listened to sequences of syllables within which GPR and VTL either remained fixed or varied randomly, in a 2 × 2 factorial design. Stimuli were presented in triplets, i.e., the same stimulus was presented three times in a row. The changes between triplets occurred in a random order in all conditions.

A control condition with bursts of broadband noise was also included. The bursts had the same onsets and offsets as the syllables and they were matched for rms level. Thus, the five stimulus conditions were (i) fixed VTL, fixed GPR (VTLfGPRf); (ii) changing VTL, fixed GPR (VTLcGPRf); (iii) changing GPR, fixed VTL (VTLfGPRc); (iv) changing GPR, changing VTL (VTLcGPRc); (v) broadband noise bursts. Conditions were presented in randomized order. Subjects were asked to pay attention to the sound sequences. To help maintain alertness, they were required to make a single button press at the end of each sequence by using a button box positioned beneath the right hand. There was no active auditory discrimination task.

The stimuli were delivered using a custom electrostatic system at 70 dB SPL. Each sound sequence was presented for a period of 8.2 or 8.6 s, after which brain activity was estimated by the fMRI blood oxygen-level-dependent response at 1.5 T (Siemens Sonata, Erlangen, Germany) by using gradient echo planar imaging in a sparse acquisition protocol (Belin et al., 1999; Hall et al., 1999) (time to repeat/time to echo, TR/TE: 13,500/50 ms). 182 brain volumes were acquired for each subject in two sessions (36 volumes for each condition/session and two volumes preceded by silence at the beginning of the session). Each brain volume comprised 48 transverse slices with an in-plane resolution of 3 × 3 mm covering the entire brain.

Each subject was assessed immediately after scanning in a separate behavioral experiment. The purpose of the behavioral experiment was to screen the participants for their ability to detect the changes in GPR and VTL used in the imaging experiment. In a two-alternative forced choice psychophysical task, subjects were presented with syllable sequences in which either VTL or GPR was varied with the same values used during scanning. Subjects detected pitch change with a mean accuracy of 76% and VTL

change with a mean accuracy of 69%. The detection of the two types of change did not differ significantly ($t = -1.03$, $df = 14$, $P < 0.3$, two-tailed).

Image analysis

Imaging data were analyzed using statistical parametric mapping implemented in SPM2 software (http//:www.fil.ion.ucl.ac.uk/spm). Scans were realigned and spatially normalized (Friston et al., 1995a) to MNI standard stereotactic space (Evans et al., 1993) and spatially smoothed with an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Statistical parametric maps were generated by modeling the evoked hemodynamic response for the different stimuli as boxcars convolved with a synthetic hemodynamic response function in the context of the general linear model (Friston et al., 1995b). The first two scans of each session were discarded from the first analysis. In a supplementary analysis, they were modeled as a separate silent baseline to get a crude estimate of activation of auditory regions. Population-level inferences concerning BOLD signal changes between conditions of interest (syllable minus noise, changing VTL and/or GPR minus fixed GPR and VTL, changing minus fixed VTL, changing minus fixed GPR and the interaction between VTL and GPR, and in the supplementary analysis syllable minus silence) were based on a random effects model that estimated the second level t statistic at each voxel. For each contrast, responses were considered significant at $P < 0.001$, uncorrected, if the location of the activity was in accordance with prior hypothesis. The MGB was localized as described previously (Griffiths et al., 2001).

Results

Subjects listened to sequences of syllables within which GPR and VTL either remained fixed or varied randomly in a 2 × 2 factorial design. They were also presented with a baseline, noise condition. When the activity produced by the experimental stimuli was contrasted with the activity to the noise baseline, they elicited activity bilaterally in STS (0.05 FWE corrected; Left hemisphere maximum at −66, −28, 0, $Z = 6.49$, voxels activated in cluster 312; Right hemisphere maximum at 66, −10, −8 $Z = 6.37$, voxels activated in cluster 292). The main effects for the experimental conditions allowed explicit testing of the key a priori hypotheses about the existence of a sensory representation of VTL in the ascending pathway after IC, and a neural correlate of perceived size in cortex beyond PAC. Furthermore, we assessed the influence of GPR on VTL processing by means of an interaction. Statistical parametric maps for the contrasts of interest are presented in Fig. 2 (subcortical structures) and Fig. 3 (cortical structures) and coordinates of local activation maxima are presented in Table 1.

In the ascending auditory pathway, the main effect of changing VTL was increased activation in the left medial geniculate body (MGB) (Table 1, Fig. 2). The main effect of VTL change produced no significant differential activation in primary auditory cortex (PAC). There was no significant main effect of GPR change, and no significant interaction between VTL and GPR either in the MGB or in the PAC.

Both VTL and GPR influence the perception of speaker size, and isolated changes in either acoustic property can produce a change in the perceived size of the source if the change is sufficiently large. When the conditions in which there is a perception of speaker size differences (i.e., all conditions in which VTL and/or GPR were changing) were contrasted with the conditions in which both VTL and GPR were fixed ((VTLcGPRf + VTLcGPRc + VTLfGPRc) − (VTLfGPRf)), they elicited activation bilaterally along STS (Table 1). Changing VTL on its own and changing GPR on its own both produced main effects on the activation in different subsets of the activation located in anterior STS/STG (Table 1, Fig. 3). For the main effect of changing VTL ((VTLcGPRf + VTLcGPRc) − (VTLfGPRf + VTLfGPRc)), the activation was left lateralized, whereas for the main effect of changing GPR ((VTLcGPRc + VTLfGPRc) − (VTLcGPRf + VTLfGPRf)), the activation was bilateral. For the main effect of changing GPR, additional activation occurred in left posterior STG.

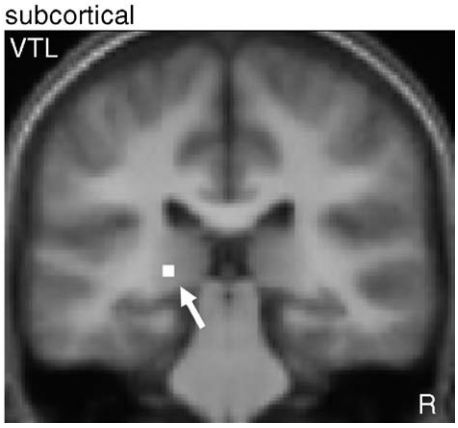

Fig. 2. Subcortical processing of the acoustic effect of vocal tract length. The group statistical parametric map for the contrast between changing and fixed VTL (thresholded at $P < 0.001$ uncorrected) has been rendered on a coronal section of the group mean normalized structural MRI volume. The arrow points to activation in the left medial geniculate body. VTL, vocal tract length. R, right hemisphere.

GPR interferes with the perception of VTL on a behavioral level (Smith and Patterson, 2005). In order to identify the cortical correlates where a concurrent change in GPR interferes with VTL processing, we analyzed the interaction of changing VTL and GPR. Specifically, we sought cortical areas that were activated in response to a change in VTL, more when GPR was fixed than when it was changing ((VTLcGPRf − VTLcGPRc) − (VTLfGPRf − VTLfGPRc)). The effect of a VTL change when GPR is fixed was greater in left anterior STS/STG and left posteromedial PT (Table 1, Fig. 3, top panel). The reverse interaction (i.e., cortical areas that are activated in response to a change in VTL, more when GPR was changing than when it was fixed ((VTLcGPRc − VTLcGPRf) − (VTLfGPRc − VTLfGPRf))), produced no significant activation.

Discussion

We have demonstrated anatomically distinct regions for the analysis of acoustic information that contributes to the perception of speaker size in subcortical and cortical structures in the human brain. The main effect of VTL change on activity, as early in the auditory pathway as the MGB, is in accordance with considerations of sensory signal processing: the extraction of VTL information from the sensory signal requires a time-stabilized representation of the stimulus like the auditory image. Although

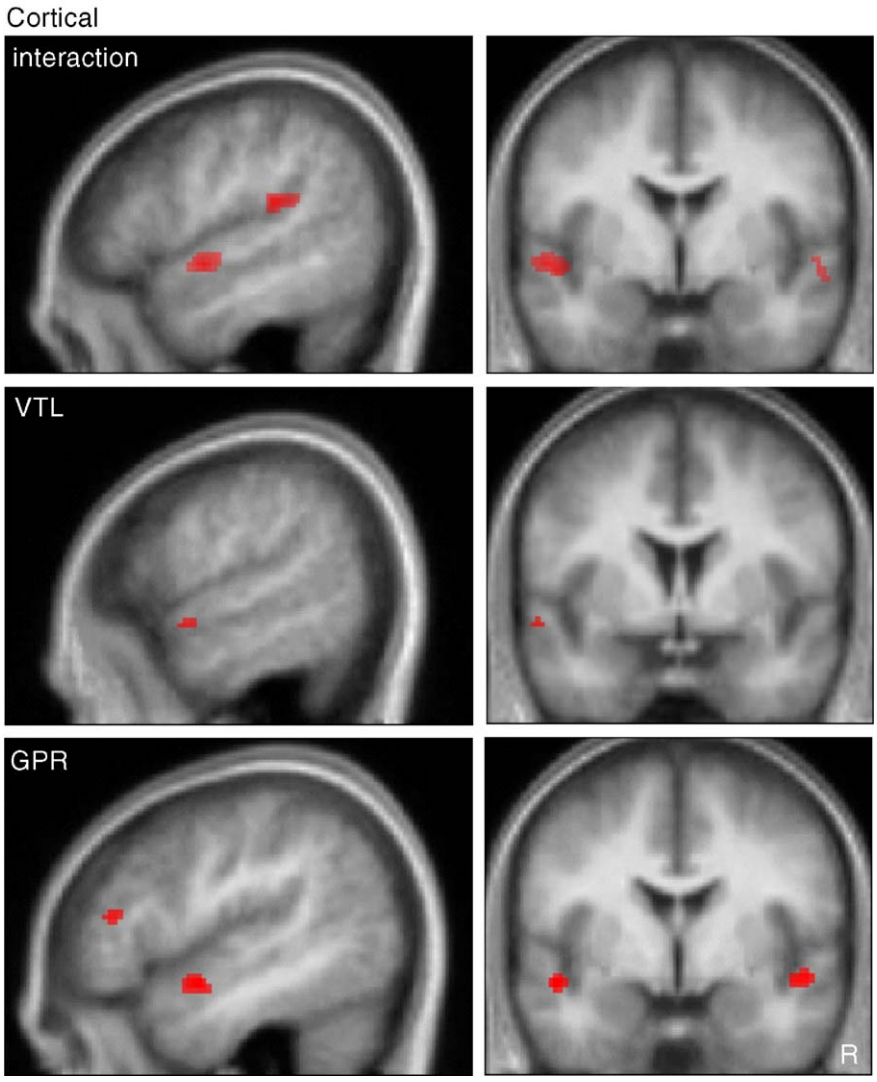

Fig. 3. Cortical correlates of processing VTL and GPR as components of auditory size. Group statistical parametric maps for the interaction between VTL and GPR and the main effects of VTL change and GPR change (thresholded at $P < 0.003$ for display purposes) have been rendered on sagittal (left) and coronal (right) sections of the group mean normalized structural MRI volume. VTL, vocal tract length. GPR, glottal pulse rate. R, right hemisphere.

earlier regions respond to timing information, it has been suggested that time stabilization is completed at IC (Griffiths et al., 2001; Langner and Schreiner, 1988). The MGB is the logical substrate for analysis of VTL differences as it is the next processing stage in the auditory pathway. The data are consistent with a primary mechanism in the human auditory system for representation of a sensory cue for auditory size that does not require other auditory associations, such as pitch or visual information about body size (Irino and Patterson, 2002). Our study covered a broad range of vocal tract lengths, extending from a vocal size corresponding to a height of 1.1 m to a vocal size corresponding to a height of 2.4 m. The activity in MGB across this broad range supports the view that the initial stage of size processing does not depend on learned associations, but rather on an early and generic mechanism for processing the scale of the sound (Smith et al., 2005). Activity in MGB is independent of any concurrent change in GPR, suggesting that the representation of VTL information is not influenced by GPR perception at this early stage of processing. This is consistent with the hypothesis that MGB is engaged in the analysis of the sensory effect of VTL changes rather than their eventual effect on perceived size. We do not assume that the MGB is specifically active in response to the size aspect of the signal, but that it responds to the specific physical properties of the sound. Thus, it would respond to an acoustic effect of size change even if there were no perceptual correlate of size change.

We cannot exclude the possibility that there is activity in response to VTL changes occurring earlier than MGB because our method was not optimized for detecting activity changes below MGB; that is, we did not use cardiac gating to reduce signal dropout due to brainstem movement. Accordingly, a supplementary analysis, using the first two scans during which no sounds were presented, as a silent baseline, showed that even with such reduced power, contrasting all conditions against a silence baseline revealed significant activity in primary auditory cortex ($P < 0.001$ uncorrected), and a trend to significance in the left MGB ($P < 0.005$ uncorrected, right MGB $P < 0.03$, uncorrected). In contrast to that, the IC did not show any differential activation against the silent baseline ($P > 0.05$ uncorrected), which may relate to the absence of cardiac gating in addition to the limited power for this contrast.

Table 1
Local activation maxima for the main effects of VTL change and GPR change ((VTLcGPRc + VTLcGPRf + VTLfGPRc) − (VTLfGPRf)), of VTL change alone ((VTLcGPRc + VTLcGPRf) − (VTLfGPRc + VTLfGPRf)), of GPR change alone ((VTLcGPRc + VTLfGPRc) − (VTLcGPRf + VTLfGPRf)), and the interaction of VTL and GPR ((VTLcGPRf − VTLcGPRc) − (VTLfGPRf − VTLfGPRc)) in the auditory pathway and auditory cortical regions

		VTL and GPR					VTL					GPR					Interaction				
		x	*y*	*z*	*Z*	*cl*	*x*	*y*	*z*	*Z*	*cl*	*x*	*y*	*z*	*Z*	*cl*	*x*	*y*	*z*	*Z*	*cl*
Subcortical																					
MGB	Left	–	–	–	–	–	−20	−30	2	3.53	4	–	–	–	–	–	–	–	–	–	–
Cortical																					
PT	Left	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−52	−32	18	4.61	57
Heschl's	Left	−40	−26	2	3.23	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Right	44	−18	2	3.27	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
STG/STS																					
Anterior/ middle	Left	−58	−4	−4	3.96	94	−56	4	−10	3.29	3	−50	−2	−14	3.93	33	−50	−4	−8	5.05	118
		−46	0	−12	3.5	scl															
	Right	54	2	−10	4.71	200	–	–	–	–	–	46	2	−14	3.76	68	52	4	−8	3.55	10
		62	2	−12	3.65	scl															
		60	−10	−2	3.64	scl															
Middle/ posterior	Left	−60	−22	−4	4.14	41	–	–	–	–	–	−58	−20	0	3.21	2	−62	−44	12	3.67	10
	Right	64	−16	6	4.14	12	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
		56	−44	8	3.26	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

x, *y*, *z* are the MNI coordinates of the local maxima (in millimeters). Statistical threshold is $P < 0.001$ uncorrected. MGB, medial geniculate body; STG, superior temporal gyrus; PT, planum temporale; *Z*, *Z* score; scl, subcluster; cl, activated voxels in the cluster.

Processing of VTL information in MGB was left-lateralized in this experiment. Lateralization at this level has been found in a previous fMRI experiment in which the right (but not the left MGB) was shown to respond to temporal regularity in regular-interval noise (Griffiths et al., 2001). In the current experiment, this lateralization was not an explicit a priori prediction but was shown to be a robust effect in a random effects analysis, suggesting that it is a property of the population from which the group is drawn. Left-lateralized activation of MGB could be driven by corticofugal connections from left-lateralized cortical areas that process syllables. MGB and cortex are highly interconnected and both regions are active simultaneously during most of their response (Miller et al., 2002). On the other hand, lateralization of auditory responses to complex (speech) sounds can also occur as early as MGB in other species (for example, anaesthetized guinea pig (King et al., 1999)): such findings are difficult to explain on the basis of top-down influences from centers concerned with speech.

Detailed behavioral studies using the same stimuli as those used here demonstrate that size judgments are influenced by GPR (Smith and Patterson, 2005). We take these results as an indication that VTL processing is not entirely independent of pitch processing.

Using fMRI, we have demonstrated an interaction between the brain activation produced by changing VTL and GPR in left PT and anterior STS. The effect of VTL change is greater when GPR is fixed, rather than when both parameters are changing. VTL is a specific aspect of vocal timbre perception. Although it is not impossible to perceive timbre differences independently from each other, these two factors have been shown to interact (Krumhansl and Iverson, 1992; Marozeau et al., 2003; Melara and Marks, 1990). Important to the interpretation of our results are the experiments done by Krumhansl and Iverson (1992) on the dependency of timbre processing with sequences of tones of changing or fixed pitches and timbres. In experiment three of their study, they find that variations in timbre do not prevent pitch from being coded relative to surrounding pitches of the sequence while variations in pitch prevent timbre from being coded relatively to the surrounding timbres. They conclude that within a sequence of changing timbres timbre is perceived relatively only if pitch is fixed in this sequence. The effect we find here for sequences changing in VTL (voice timbre) only when GPR (pitch) is fixed might be a manifestation of the different coding for timbre in fixed pitch sequences.

Manipulations of the effect of vocal tract length are by its nature always associated with a shift in the spectral centroid of the spectral envelope, i.e., the average frequency weighted by amplitude is different for differently sized vocal tracts. In the human brain, there are several regions around and near Heschl's gyrus organized in a frequency-specific fashion, i.e., different frequencies are coded within different areas of one region. Although the sensitivity of fMRI is so low that to show this so-called tonotopy usually sounds need to differ in frequencies by several octaves (Talavage et al., 2000, 2004), changes in VTL could theoretically lead to responsiveness in these regions. Not only changes in VTL, however, but also different speech sounds lead to changes in the spectral centroid. An /a/ spoken by the same speaker has a different spectral centroid than an /i/. By always contrasting conditions with varying speech sounds, we have thus controlled for regions which are responsive to frequency shifts only. We assume that regions responsive to VTL are sensitive to the change in spectral envelope as a whole brought about by the formant frequency shift and/or the reduction in decay duration.

We have argued previously that PT is involved in the computation of spectro-temporal 'templates' before the routing of auditory information into different cortical processing streams (Griffiths and Warren, 2002). It may be that PT is also involved in the normalization of the time-stabilized tonotopic representation of the sound. The implication is that PT may be the center that segregates the acoustic information concerning source size and pitch from syllable type; that is, the center makes speech recognition robust to changes in GPR and VTL.

There was activation as a function of the interaction of VTL and GPR in PT and anterior STS, in accordance with our a priori anatomical hypothesis. A previous study demonstrated bilateral PT

activation and right-lateralized mid-STS activation as a substrate for the abstraction of spectral envelope shape, a generic property of acoustic sources (Warren et al., 2005). The changes in spectrum in the current experiment are more subtle than in the previous work (Warren et al., 2005) and require stabilization in time for accurate analysis. The lateralization may be restricted to speech stimuli; it may be that cortical processes for the perception of auditory size per se are not lateralized. It will be of considerable future interest to determine whether the analysis of resonator-size information is lateralized in the perception of non-speech sounds like musical instruments and animal calls.

Our findings are consistent with a hierarchical scheme for the analysis of size information in natural sounds such as voices. The initial normalization of the sensory signal could occur at the level of MGB in the ascending auditory pathway, while the subsequent cortical analysis of VTL information engages a network including posteromedial PT and anterior STS. The interaction with GPR at the cortical level suggests that these areas are the neural correlate of perceived changes in VTL.

We suggest three directions for future studies. (1) It will be of interest to determine whether the processing stages for size information in speech are generic for processing of resonator scale, and thus extend to the analysis of size of other auditory objects like animals and musical instruments. (2) Resonance scale has more than one effect on sounds and it will require more experiments to break down the different effects on the acoustic stimulus, e.g., shifts in the spectral envelope as well as decay duration. (3) Further work is required to demonstrate the neural correlate of the perception of size.

Acknowledgments

This work was supported by the VW Stiftung (I/79 783), the Wellcome Trust and the UK MRC (G9900362).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2006.02.045.

References

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. NeuroImage 10, 417–429.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L., Phinney, R.E., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. 3, 1813–1817.

Fairchild, L., 1981. Mate selection and behavioral thermoregulation in Fowler's toads. Science 212, 950–951.

Fitch, W.T., Giedd, J., 1999. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J. Acoust. Soc. Am. 106, 1511–1522.

Friston, K.J., Ashburner, j., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalisation of images. Hum. Brain Mapp. 2, 165–189.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Gonzalez, J., 2004. Formant frequencies and body size of speaker: a weak relationship in adult humans. J. Phon. 32, 277–287.

Griffiths, T.D., 2005. Functional imaging of pitch processing. In: Plack, C.J., Oxenham, A.J., Fay, R.R., Popper, A.N. (Eds.), Pitch: Neural Coding and Perception. Springer Verlag.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.

Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., Patterson, R.D., 2001. Encoding of the temporal regularity of sound in the human brainstem. Nat. Neurosci. 4, 633–637.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Irino, T., Patterson, R.D., 2002. Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-Mellin transform. Speech Commun. 36, 181–203.

Ives, D.T., Smith, D.R.R., Patterson, R.D., 2005. Discrimination of speaker sizes from syllable phrases. J. Acoust. Soc. Am. 118, 3816–3822.

Kawahara, H., Irino, T., 2005. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi, P. (Ed.), Speech Separation by Humans and Machines. Kluer Academic, Massachusetts, pp. 167–180.

Kawahara, H., Masuda-Kasuse, I., de Cheveigne, A., 1999. Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: possible role of repetitive structure in sounds. Speech Commun. 27, 187–207.

King, C., Nicol, T., McGee, T., Kraus, N., 1999. Thalamic asymmetry is related to acoustic signal complexity. Neurosci. Lett. 267, 89–92.

Krumhansl, C.L., Iverson, P., 1992. Perceptual interactions between musical pitch and timbre. J. Exp. Psychol. Hum. Percept. Perform. 18, 739–751.

Kunzel, H., 1989. How well does average fundamental frequency correlate with speaker height and weight? Phonetica 46, 117–125.

Langner, G., Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms. J. Neurophysiol. 60, 1799–1822.

Marcus, S.M., 1981. Acoustic determinants of perceptual center (P-center) location. Percept. Psychophys. 30, 247–256.

Marozeau, J., de Cheveigne, A., McAdams, S., insberg, S., 2003. The dependency of timbre on fundamental frequency. J. Acoust. Soc. Am. 114, 2946–2957.

Melara, R.D., Marks, L.E., 1990. Interaction among auditory dimensions: timbre, pitch, and loudness. Percept. Psychophys. 48, 169–178.

Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., 2002. Spectro-temporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 87, 516–527.

Patterson, R.D., Allerhand, M.H., Giguere, C., 1995. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. 98, 1890–1894.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.

Penagos, H., Melcher, J.R., Oxenham, A.J., 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 24, 6810–6815.

Peterson, G.E., Barney, H.I., 1952. Control methods used in the study of vowels. J. Acoust. Soc. Am. 24, 75–184.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W.T., Clutton-Brock, T., 2005. Red deer stags use formants as assessment cues during intrasexual agonistic interactions. Proc. Biol. Sci. 272, 941–947.

Rendall, D., Vokey, J.R., Nemeth, C., Ney, C., 2005. Reliable but weak voice-formant cues to body size in men but not women. J. Acoust. Soc. Am., 117.

Scott, S. K., 1993. P-centres in speech: an acoustic analysis. PhD Thesis, University College London.

Smith, D.R.R., Patterson, R.D., 2005. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex and age. J. Acoust. Soc. Am. 118, 3177–3186.

Smith, D.R.R., Patterson, R.D., Turner, R., Kawahara, H., Irino, T., 2005. The processing and perception of size information in speech sounds. J. Acoust. Soc. Am. 117, 305–318.

Talavage, T.M., Ledden, P.J., Benson, R.R., Rosen, B.R., Melcher, J.R., 2000. Frequency-dependent responses exhibited by multiple regions in human auditory cortex. Hear Res. 150, 225–244.

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R., Dale, A.M., 2004. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J. Neurophysiol. 91, 1282–1296.

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.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. NeuroImage 24, 1052–1057.