Auditory-Motor Expertise Alters "Speech Selectivity" in Professional Musicians and Actors

Frederic Dick[1], Hwee Ling Lee[2], Howard Nusbaum[3] and Cathy J. Price[4]

[1]Department of Psychological Sciences, Birkbeck/University College London (UCL) Centre for NeuroImaging, Birkbeck College, University of London, London, WC1E 7HX, UK, [2]Max-Planck-Institute for Biological Cybernetics, 72076 Tübingen, Germany, [3]Department of Psychology, University of Chicago, Chicago, IL 60637, USA and [4]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, WC1N 3BG London, UK

Address correspondence to Frederic Dick, Birkbeck/UCL Centre for NeuroImaging (BUCNI), Birkbeck College, Malet Street, London, WC1E 7HX, UK. Email: f.dick@bbk.ac.uk.

Several perisylvian brain regions show preferential activation for spoken language above and beyond other complex sounds. These "speech-selective" effects might be driven by regions' intrinsic biases for processing the acoustical or informational properties of speech. Alternatively, such speech selectivity might emerge through extensive experience in perceiving and producing speech sounds. This functional magnetic resonance imaging (fMRI) study disambiguated such audiomotor expertise from speech selectivity by comparing activation for listening to speech and music in female professional violinists and actors. Audiomotor expertise effects were identified in several right and left superior temporal regions that responded to speech in all participants and music in violinists more than actresses. Regions associated with the acoustic/information content of speech were identified along the entire length of the superior temporal sulci bilaterally where activation was greater for speech than music in all participants. Finally, an effect of performing arts training was identified in bilateral premotor regions commonly activated by finger and mouth movements as well as in right hemisphere "language regions." These results distinguish the seemingly speech-specific neural responses that can be abolished and even reversed by long-term audiomotor experience.

Keywords: auditory, expertise, fMRI, language, plasticity

Introduction

Human speech comprehension has been argued to have evolved as a special and unique form of perceptual expertise. The "specialness" of speech processing derives in part from the demands that the speech signal puts on the auditory system but also in the way that speech is produced. It is almost impossible to find other human skills that parallel those of spoken language in the frequency and breadth of exposure, the manner of production, the complexity and hierarchy of its form, and above all the inextricable integration of speech processing with meaningful language itself.

Given this, it is perhaps unsurprising that previous studies (e.g., Vouloumanos et al. 2001) report that listening to speech evokes much stronger neural responses in certain brain regions than do other stimuli—as has also been observed for "special" classes of visual stimuli such as faces (Kanwisher et al. 1997). In addition, different aspects of meaningful auditory speech tend to be preferentially processed by different parts of the temporal lobe (Scott and Johnsrude 2003). Thus, speech is an interesting case for understanding how and why some "higher-level" brain regions come to show seemingly stimulus class–specific response properties.

The aim of the present study was to investigate the role of expertise in tuning such speech-selective regions. How does sensorimotor experience in producing and perceiving speech signals contribute to this selective response? One type of sensorimotor expertise that parallels (but is certainly not equivalent to) the experience we have in producing and perceiving speech is found in professional musicians who both produce and perceive music. Specifically, we asked to what extent professional violinists listening to violin music activate brain regions that are typically speech selective. By early adulthood, expert violinists' experience approximates several important characteristics of spoken language use, including 1) early exposure to and production of musical sounds, often resulting in up to 10 000+ h of musical production by early adulthood (Krampe and Ericsson 1996); 2) the use of a basic vocabulary of scales and keys to create infinite numbers of unique musical utterances; 3) a nonlinear and exquisitely timed mapping of smaller and larger sound units onto embedded motor schemas; 4) integration of auditory information across multiple time frames, with dependencies within and across time frames; 5) synchronization and turn-taking within musical ensembles; and 6) a detailed internal representation of the sound of an instrument, for example, a kind of musical "inner voice."

By analogy to research on visual expertise showing recruitment of "face-selective" fusiform cortex after intensive training on nonface stimuli (Palmeri et al. 2004; Harley et al. 2009), violinists listening to violin music may recruit regions that are typically speech selective, despite the large acoustical and informational differences between speech and music (see Discussion). This would identify which speech-selective regions are driven by "audiomotor expertise" than by speech content per se. Indeed, a study by Ohnishi et al. (2001) demonstrated that a mixed group of musicians listening to a single repeated piece of keyboard music (Bach's Italian Concerto) showed more activation than naive nonmusicians in several superior temporal and frontal regions often associated with language processing. However, this study did not directly compare music to speech processing and therefore did not distinguish regions that were modulated by auditory experience from those that were speech-selective despite musical expertise. Moreover, a musician's experience differs from that of a nonmusician in several fundamental ways. Perhaps most importantly, expert musicians are by definition not drawn from the typical population. Because of self-selection or years of intensive and attentionally demanding practice and tutelage, such musicians may show more general differences in brain

organization as a result of performing arts training (Gaab and Schlaug 2003; Gaser and Schlaug 2003; Bermudez et al. 2009).

Our experimental design therefore needed to distinguish the effect of audiomotor expertise within speech-selective regions from the effects of performing arts training. To do this, we sought nonmusicians who had performing arts training in producing and perceiving speech itself. Professional actors are ideal in this regard because, like musicians, they have spent thousands of hours perfecting the physical and psychological techniques of sound production with a small "canonical" repertoire of works that are learned through explicit models and teaching. Actors must also plan their sound production on multiple timescales and often couch their performance in a particular historical style. The experience and training that actors have in reproducing speech is therefore over and above that experienced through natural language acquisition in violinists who—like any person who speaks a language—are also "speech experts." Here, we define performance arts training as the history of deliberate sound production, planning, and attention to the class of sounds that the performers were trained in. We then investigate neural activation associated with performance arts training as that which was greater for listening to 1) dramatic speech versus violin music in actors and 2) violin music versus dramatic speech in violinists.

We might also expect that expertise-related upregulation in activation would be modulated by the relative familiarity of the specific piece or excerpt being listened to. Indeed, within the musical domain, Margulis et al. (2009) showed shared instrument-specific upregulations in left premotor and posterior planum temporale (PT)/superior temporal sulcus (STS) activation when flautists and violinists listened to their own instruments. Leaver et al. (2009) also observed musical familiarity–related changes in premotor activation. Finally, Lahav et al. (2007) found that novice pianists showed more activation in bilateral inferior and prefrontal regions when listening to melodies they were familiar with playing, versus those that they had simply passively listened to, suggesting that mere exposure to sound sequences is insufficient to engage motor systems during perception. With reference to the present study, we hypothesize that greater activation for familiar compared with unfamiliar stimuli would suggest expertise at the level of specific sequences of musical and motor events, whereas common effects for familiar and unfamiliar stimuli would suggest expertise at the level of well-established and stereotypical patterns that generalize across different pieces of music.

In summary, we compared professional female violinists and "actors" neural responses to short (i.e., 10 s) excerpts from violin pieces and dramatic monologues from female roles. Only females were included in the study because males and females will have different experience with gender-specific acting roles—a factor that has been shown to affect activation in male and female professional dancers (Calvo-Merino et al. 2006). In order to disentangle effects of expertise and familiarity, excerpts were split by familiarity, so that violin excerpts were familiar or unfamiliar to female violinists, and monologue excerpts were familiar or unfamiliar to female actors. Responses to music and speech stimuli were compared with phase-scrambled versions of the stimuli (Thierry et al. 2003); these phase-scrambled stimuli (henceforth the auditory baseline condition) were also used to assure that there were no differences in low-level auditory activation between the 2 expert groups. In the scanner, participants simply listened to the stimuli presented in "miniblocks" of 2 excerpts, interspersed with silence (Fig. 1). In order to monitor attention while minimizing task-related confounds, participants were asked to push a button after each miniblock to indicate their level of alertness while listening to the sounds.

Our experimental design allowed us to tease apart 3 different patterns of speech-related effects. First, we looked for speech-selective regions that could be attributable to "general audiomotor expertise"; such regions would show speech activation in both groups and significantly more activation for violin music in violinists than in actors. We predicted that this effect would be observed in the left posterior STS (pSTS) region where we have previously reported increased activation after short-term training categorizing nonspeech sounds (Leech et al. 2009), as well as in the right homologue of this region. Second, we looked for "speech-selective" regions that could be related to the acoustical or informational content of speech; such regions would show more activation for speech than violin music in both actors and violinists. We predicted that this effect might also be observed in the left middle STS regions that Specht et al. (2009) found to be more sensitive to acoustical complexity manipulations in speech than music and that Narain et al. (2003) found to be more sensitive to intelligible speech. Third, we looked for effects of "performing arts training"; in such regions, violinists should show more activation for music than speech, whereas actors should show more activation for speech than for music. We predicted that these effects would be in the frontal and parietal regions associated with performance arts training in the visual domain when the participants were highly skilled dancers observing the specific dance moves they themselves make (Calvo-Merino et al. 2005, 2006). Finally, we looked for
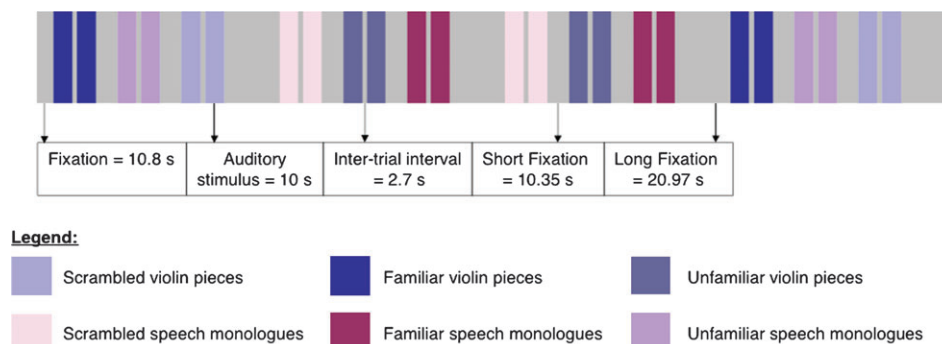

Figure 1. Schematic representation of the sequence of events in a single fMRI run (session).

stimulus "familiarity" effects and predicted that these effects would be observed in the premotor and frontal regions that Leaver et al. (2009) and Lahav et al. (2007) associated with musical familiarity. Auditory activation that was common to dramatic speech and violin music in both groups (i.e., not speech selective) is reported in the Supplementary Materials.

Materials and Methods

Participants
Fifteen violinists (mean age 24.5 years, range 18-31) and 15 actors (mean age 26.7 years, range 18-37), all female, participated in the study. As noted in the Introduction, it was important to restrict participants to one gender due to actors' differential experience in studying and performing male and female roles. All participants were right-handed speakers of American or British English with no known neurological or physical problems. They were either professionals or performance students at London conservatories or drama schools. Violinists began taking private lessons at an average age of 5.9 years (range 3-8) and actors at age 10.8 (range 4-18). For the year before scanning, violinists estimated that they individually practiced an average of 4.2 h daily (standard deviation [SD] 1.71); actors estimated that they individually practiced an average of 3.2 h daily (SD 3.1). Actors were experts in reproducing speech but representative of the general population in their ability to reproduce music. In contrast, violinists were experts in reproducing music but representative of the general population in their ability to reproduce speech. The study was approved by the ethics committee at the National Hospital for Neurology and Neurosurgery, and all participants gave written consent before beginning the study.

Stimuli
Stimuli were short excerpts from the violin literature and from dramatic monologues for female characters; the violinist and actor who recorded the excerpts are established artists in the UK. Both violin and voice recordings were made in a sound-attenuated chamber (IAC) using a Shure SM57 instrument microphone suspended on a sound boom approximately 300 mm from the bridge of the violin or above and in front of the actor's mouth. The microphone was plugged via insulated leads into a Tascam Digital Audio Tape (DAT) recorder, using 44.125 kHz sampling with 16-bit quantization. The original DAT recording was read into lossless Audio Interchange File Format files using ProTools Mbox via optical cable at original resolution. Additional sound editing was performed in SoundStudio 2.1.1 and Praat 4.3.12.

Each recording was spliced into 10-s intervals, with splice points chosen on the basis of phrasing, intonation, and amplitude cues. For both speech and violin stimuli, ~200-ms threshold ramps were introduced at the beginning and end of each file to avoid pops. All stimuli were scaled in Praat to average 75 dB intensity; previous to scaling, segments of some of the speech stimuli were dynamically compressed to preclude clipping during the intensity scaling.

The violin pieces and dramatic monologues were chosen to be very familiar or quite unfamiliar to the performers. Familiar violin pieces were chosen such that any conservatory student would have been required to play these pieces during training, whereas unfamiliar violin pieces were chosen as being infrequently played, yet matched to the familiar pieces as much as possible in terms of style, dynamics, tempo, and historical period. With respect to the speech monologues, the familiar pieces were chosen from standard auditory repertoire; unfamiliar pieces were chosen to match their familiar counterparts in style and period but are infrequently encountered on stage and in collections of audition works. Familiarity groupings for violin and speech excerpts were confirmed in a postscanning questionnaire; note that familiarity only applies within the expert domain, for example, actors did not differentiate between familiar and unfamiliar violin pieces.

Our stimuli varied in their emotional tone. Musical excerpts were drawn from minor and major keys at a number of tempi (Adagio to Allegro) and musical styles; dramatic excerpts were drawn from tragedies and comedies that like the musical excerpts varied in

emotional tone. However, we did not investigate how our results were affected by emotional tone. Nor was it possible to match cross-domain emotional arousal within or across participant groups because this is highly subjective and dependent on the individual. A list of excerpts and example recordings are available in Supplementary Materials.

Auditory baseline stimuli were also created by phase scrambling a subset of the music and dramatic speech excerpts; this process retains the overall spectrum of the original excerpt but removes longer-term temporal changes (Thierry et al. 2003). For example, see Supplementary Materials.

Experimental Paradigm
There were 2 within-subject factors, Domain (Music/Speech) and Familiarity (Familiar/Unfamiliar/Auditory Baseline) and thus 6 stimulus types: 1) violin pieces familiar to violinists, 2) violin pieces unfamiliar to all participants, 3) speech monologues familiar to actors, 4) speech monologues unfamiliar to all participants, 5) scrambled violin pieces (the auditory baseline for music), and 6) scrambled speech monologues (the auditory baseline for speech. All stimuli were presented in pairs, with each stimulus lasting 10 s (±0.3 s), 2.7 s between stimuli from the same pair and 10.35 or 20.97 s fixation between pairs of stimuli (see Fig. 1 for a graphical representation of a single run). All 6 conditions were presented within a single run of 129 scans with each subject participating in 6 runs (total scanning time = 46.44 min). Four of these volumes were acquired before beginning a run to allow for magnetization to equilibrate. Conditions were counterbalanced within and between runs. Participants were instructed to listen to the stimuli while keeping their eyes open and watched a black fixation cross presented on a white background; participants pressed the key button after each pair of auditory stimuli to indicate if they were awake (left finger) or getting sleepy (right finger). In addition to the finger-press response, an eye monitor was used to ensure that the participants kept their eyes open and did not doze off to sleep during the experiment.

Magnetic Resonance Imaging Methodology and Analysis
Magnetic resonance images were acquired using a 1.5-T Siemens Sonata MRI scanner (Siemens Medical). A gradient-echo planar image sequence was used to acquire functional images (time repetition [TR] 3600 ms; time echo [TE] 50 ms; field of view 192 × 192 mm; 64 × 64 matrix). Forty oblique axial slices of 2 mm thickness (1 mm gap), tilted approximately 20 degrees, were acquired. A high-resolution anatomical reference image was acquired using a T_1-weighted 3D Modified Driven Equilibrium Fourier Transform (MDEFT) sequence (TR 12.24 ms; TE 3.56 ms; field of view 256 × 256 mm; voxel size 1 × 1 × 1 mm). Auditory stimuli were presented using KOSS headphones (KOSS Corporation; modified for use with magnetic resonance imaging by the MRC Institute of Hearing Research); stimuli were presented at a comfortable volume that remained constant over participants and that minimized interference from acoustical scanner noise.

Functional image analysis was performed using Statistical Parametric Mapping software (SPM5; Wellcome Department of Imaging Neuroscience). The first 4 volumes of each functional magnetic resonance imaging (fMRI) session were discarded because of the nonsteady condition of magnetization. Scans were realigned, unwrapped, and spatially normalized to the Montreal Neurological Institute space using an echo-planar imaging template. Functional images were then spatially smoothed (full-width at half-maximum of 6 mm) to improve the signal-to-noise ratio. The functional data were modeled in an event-related fashion with regressors entered into the design matrix after convolving each event-related stick function with a canonical hemodynamic response function. The model consisted of four active conditions: 1) familiar music, 2) familiar speech, 3) unfamiliar music, and 4) unfamiliar speech, and two auditory control conditions: 5) scrambled music and 6) scrambled speech. Condition-specific effects (relative to fixation) were estimated for each subject according to the general linear model (Friston et al. 1995). These parameter estimates were passed to a second-level analysis of variance (ANOVA) that modeled 12 different conditions (6 per subject) as a 2 × 6 design, with group as the between-subject variable and stimulus as the within-subject variable. The 6 levels of stimulus were familiar, unfamiliar, and scrambled music, and familiar,

unfamiliar, and scrambled speech. Note that this is not a fully balanced 2 × 3 × 2 factorial design because familiar violin excerpts were unfamiliar to actors and familiar dramatic monologues were unfamiliar to violinists. A correction for nonsphericity was included. Unless otherwise stated, results are reported at $P < 0.05$ at the cluster and/or peak level using a height threshold of $P < 0.001$ (uncorrected), Family-Wise Error (FWE) corrected for multiple comparisons across the whole brain. Probabilistic mapping was determined by comparing the overlap between the projected activations on the cortical surface in FreeSurfer and the probabilistic cytoarchitectonic atlases provided in the FreeSurfer distribution (Dale et al. 1999).

Preliminary investigation of the effects confirmed that there were no between-group differences in activation for scrambled speech or scrambled violin (the auditory baseline conditions). We then searched for our effects of interest as follows:

1) Speech-selective activations related to audiomotor expertise.

Speech-selective regions related to auditory expertise were those activated by speech in both groups and by music in violinists. We identified regions where activation for music was greater in violinists than actors ($P < 0.05$ following FWE correction for multiple comparisons). To focus on speech-processing regions, we used the inclusive masking option in SPM to limit the statistical map to voxels that were also activated ($P < 0.001$ uncorrected) by [speech > music in actors]; [speech > auditory baseline in actors]; [speech > auditory baseline in violinists]; and [music > auditory baseline in violinists].

2) Speech-selective and music-selective activations related to informational or acoustical content.

Speech-selective regions related to informational or acoustical content were defined as the set of regions showing more activation for speech than music in violinists and in actors. We used the conjunction analysis option in SPM5 to identify voxels where both groups showed significantly more activation for speech than music; activation was thresholded at FWE-corrected $P < 0.05$. The conjunction analysis only identified areas that were more activated ($P < 0.001$ uncorrected) for [speech > music in actors]; [speech > music in violinists]; [speech > auditory baseline in actors]; and [speech > auditory baseline in violinists]. The opposite set of contrasts was used to identify music-selective regions.

3) The effect of performing arts training.

The effect of performing arts training was identified as that which was greatest for music compared with dramatic speech in violinists and for dramatic speech compared with music in actresses (i.e., an interaction between group and stimuli). To focus on effects that were consistent across groups, we used the inclusive masking option in SPM to limit the statistical map to voxels that were also activated ($P < 0.001$ uncorrected) by [music > speech in violinists]; [speech > music in actors]; [speech > auditory baseline in actors]; and [music > auditory baseline in violinists].

4) The effect of familiarity.

In each of the regions identified in effects 1–3 above, we investigated whether there was more activation at $P < 0.001$ uncorrected for 1) familiar versus unfamiliar music in violinists and 2) familiar versus unfamiliar speech in actors. No significant effects were observed, and there were also no significant effects of familiarity at the whole brain level ($P < 0.05$ following FWE correction for multiple comparisons).

Behavioral data from 3 participants (1 violinist) were not available due to technical problems. Behavioral percent response and reaction time data were analyzed using nonparametric (Wilcoxon) signed-rank test or rank-sum Bonferroni-corrected comparisons, as behavioral data were not normally distributed.

Results

Behavioral responses in the scanner indicated that all participants were consistently alert and awake during all listening conditions. There were no significant differences in any condition between violinists and actors in their ratings of alertness, their reaction times in indicating alertness, or in the number of nonresponses (very low in both groups). Across groups, alertness ratings were significantly higher for speech and music relative to their scrambled versions (Bonferroni-corrected pairwise Wilcoxon sign-rank comparisons all $P < 0.05$, no other comparisons significant).

All fMRI analyses were based on a second (group)-level 2 × 6 ANOVA, with Group (Violinists/Actors) as a between-subject factor and Stimulus as a within-subject factor. The 6 levels of stimulus were familiar, unfamiliar, and scrambled music, and familiar, unfamiliar, and scrambled speech—see Materials and Methods for further explanation. As noted above, this is not a fully balanced factorial design as familiar violin excerpts were unfamiliar to actors and familiar dramatic monologues were unfamiliar to violinists.

Activations Related to Audiomotor Expertise in ''Speech Regions''

These were regions activated by speech in both groups and by music in violinists. They were defined as regions where activation for music was greater in violinists than actors (violin-expertise effect, $P < 0.05$ corrected), but we only report these effects when activation was also greater ($P < 0.001$ uncorrected) for [speech > music in actors]; [speech > auditory baseline in actors]; [speech > auditory baseline in violinists]; and [music > auditory baseline in violinists]. The results are shown in Figure 2 and Table 1.

Three regions met our criteria. The most extensive effect was observed in the right STS, inferior to and extending into the PT. There were also 2 separate regions identified in the left temporal lobe, one along the anterior superior temporal gyrus (aSTG) and the other in the PT. In each of these regions, activation was not significantly different for music and speech in violinists who have substantial expertise with both speech and music. This pattern of response differs to that observed in the left STS (i.e., the homologue of the right STS region), where violinists and actors showed more activation for speech than music. Finally, we note that the effects of audiomotor expertise were observed for both familiar and unfamiliar stimuli, and the difference between familiar and unfamiliar stimuli did not survive a statistical threshold of $P < 0.001$ uncorrected.

Speech-Selective and Music-Selective Activations Related to Informational or Acoustical Content

Speech-selective regions related to acoustical/information content were those activated in the conjunction of speech more than music in both actors and violinists. We only report this conjoint effect when activation was also greater ($P < 0.001$ uncorrected) for [speech > music in actors]; [speech > music in violinists]; [speech > auditory baseline in actors]; and [speech > auditory baseline in violinists]. This identified "speech-selective" activation along almost the entire length of the left STS, with corresponding effects in the right hemisphere in the anterior portion of the STS and upper bank of the STS (see Fig. 3 and Table 2). In contrast, the parallel set of analyses did not reveal any "music-selective" activation shared by both actors and violinists. To facilitate comparison with other studies on music perception (e.g., Levitin and Menon 2003), we include tables of activations for violinists and actors for the intact versus scrambled violin music contrast as in Supplementary Materials.

The Effect of Performing Arts Training

The effect of performing arts training was identified as that which was greater for music compared with speech in violinists
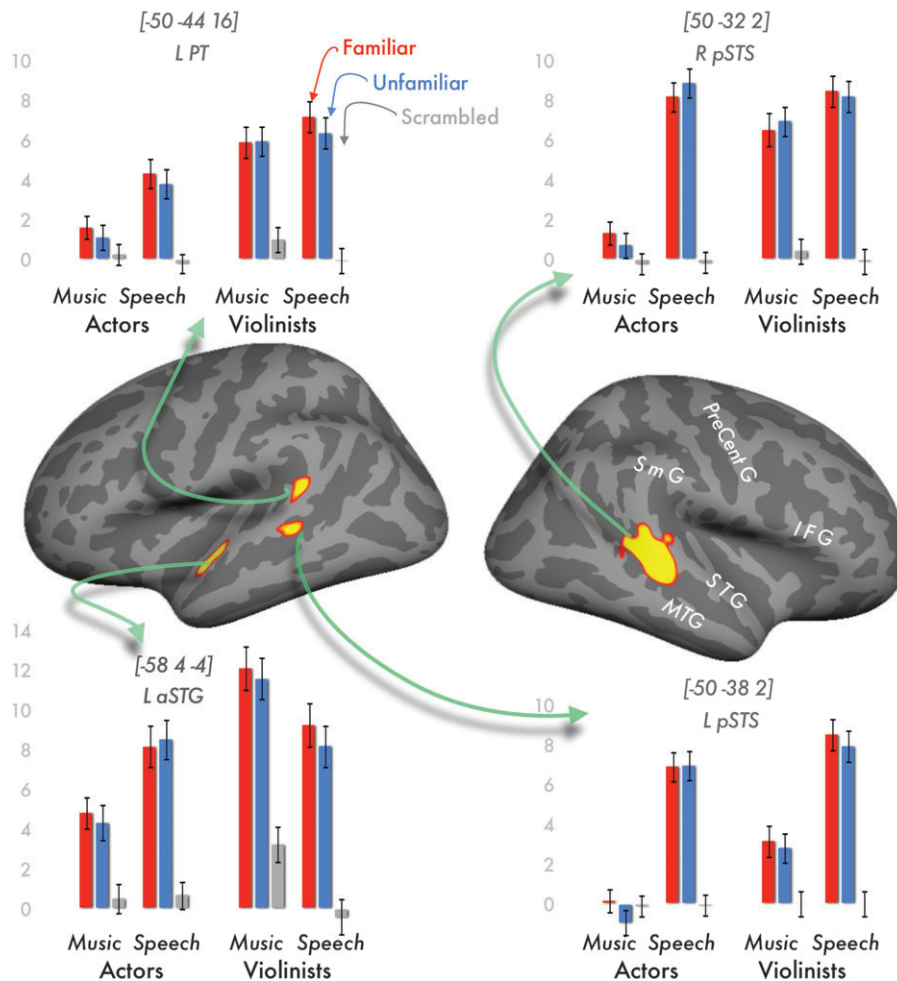
Figure 2. Effects of audiomotor expertise, where violinists show greater activation than actors for violin music in regions that are "speech selective" in actors. Thresholded activation maps were registered to and displayed on an average cortical surface in FreeSurfer. Bar graphs show parameter estimates of each condition, separated by familiarity, stimulus, and group; error bars show ±1 standard error. Bracketed Talairach coordinates refer to the peak voxel in each activation cluster. Abbreviations: L PT = left planum temporale; L or R pSTS = left or right posterior superior temporal sulcus; L aSTG = left anterior superior temporal gyrus.

Table 1

Audiomotor expertise: regions where violinists showed more activation for violin music than actors (FWE-corrected P < 0.05) and actors showed greater activation for dramatic speech than music (P < 0.001 voxelwise).

Region	Hemi	x	y	z	Z-score	Clust size	Vln-V	Vln-A	Spch-V	Spch-A
Mid-pSTS	R	50	−32	2	6.37	597	6.6	n.s.	Inf	Inf
	R	62	−28	2	5.78		7.1	5.2	Inf	Inf
(LH homologue to above)	L	−50	−38	2	3.72	38	3.7	n.s.	Inf	Inf
Anterior STG	L	−58	4	−4	5.81	107	6.7	4.4	Inf	7.1
Posterior STG/SMG	L	−50	−44	16	5.05	122	5.9	n.s.	7.7	5.2

Regions within the performing arts training network were excluded (see Table 3). Z-score is for peak voxel within cluster. Cluster size is FWE-corrected P < 0.05, with a height threshold of P < 0.001. Last 4 columns show Z-scores of main effects of stimulus for each group with corresponding baseline condition subtracted; n.s. = not significant at voxelwise P < 0.001. Left hemisphere (LH) homologue to right mid-STS activation is significant only when uncorrected for multiple comparisons (P = 0.0001). Abbreviations: Hemi = hemisphere; Clust size = cluster size; Vln = violin music; Spch = dramatic speech; V = violinists; A = actors.

and for speech compared with music in actresses (i.e., an interaction between group and stimuli). We only report those regions for which activation was greater (*P* < 0.001 un-corrected) for [music > speech in violinists]; [speech > music in actors]; [speech > music in actors]; [music > speech in violinists]; [speech > auditory baseline in actors]; and [music > auditory baseline in violinists]. The results are shown in Figure 4 and Table 3.

The results revealed increased activation for sounds that performers are experts in reproducing and perceiving, relative to all other conditions, in bilateral dorsal premotor regions (with activation falling within probabilistically defined Brodmann's Area 6 in FreeSurfer—see Materials and Methods), the right pars opercularis (bridging probabilistically defined BA44 and BA45), the right PT, and the right superior posterior cerebellum. These effects were observed for both familiar and unfamiliar stimuli, and the difference between familiar and unfamiliar stimuli did not survive a statistical threshold of *P* < 0.001 uncorrected.

Discussion

Our results parcellate speech-selective regions into those related to auditory expertise and those related to acoustical/
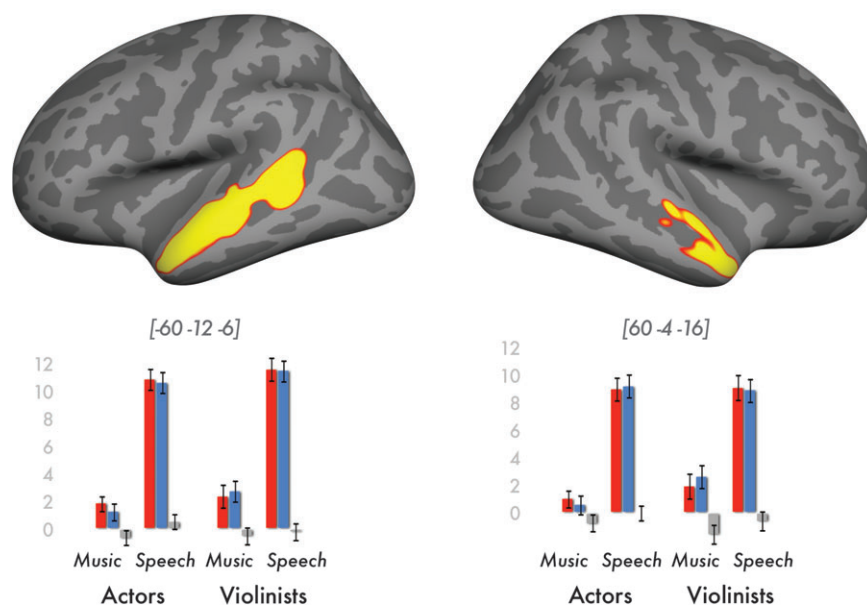
Effects of Speech Selectivity

Figure 3. Effects of speech selectivity, where both violinists and actors show greater activation for dramatic speech than for violin music.

Region	Hemi	x	y	z	Z-score	Clust size	Vln-V	Vln-A	Spch-V	Spch-A
Length of STG/STS	L	−60	−10	−6	Inf	1566	3.9	n.s.	Inf	Inf
		−58	2	−20	Inf		n.s.	n.s.	Inf	Inf
		−62	−24	−6			4.5	3.6	Inf	Inf
Anterior STG/STS	R	50	12	−30	7.05	431	n.s.	n.s.	7.9	6.2
		60	−4	−18	6.9		3.9	n.s.	Inf	Inf
		56	2	−22	6.64		n.s.	n.s.	7.3	7.3

Other information as in Table 1.

information content. To distinguish the effect of auditory expertise from the influence of performing arts training, we also differentiated activation that is selective to the sounds that performers have been trained to produce. Speech-selective regions were defined as those activated by speech relative to auditory baseline in all participants and by speech relative to music in actors. Within these regions, the effect of audiomotor expertise was identified where activation for listening to music was greater for violinists than actors and the effect of information content was identified where activation was greater for speech than music in both actors and violinists. In contrast, the defining feature of regions associated with performing arts training was that stimulus selectivity was literally "flipped" depending upon an individual's expert domain. Moreover, regions associated with performing arts training were not necessarily activated by speech in violinists. Below, we discuss each of these effects. We acknowledge that, as we only tested females, our results may be less generalizable to males. However, we think this is highly unlikely because there is no reliable evidence to suggest that the neural processing of speech or music is gender specific. Moreover, previous studies have shown that any detectable gender

differences in language activation are 1) inconsistent across studies and 2) negligible compared with other sources of variance; see Kherif et al. (2009) for a previous discussion of this issue.

Activations Related to Audiomotor Expertise in "Speech Regions"

Our finding that speech-sensitive regions were also recruited by expert violinists listening to violin music is consistent with a training-related change in response selectivity but inconsistent with response selectivity that is driven purely by acoustical or informational properties unique to speech sounds or language more generally. Four regions were identified. The largest was in the right superior temporal cortex, with 3 smaller regions in the left superior temporal cortex (including the homologue of the right hemisphere region). Previous studies have shown that these same regions are selectively upregulated for finer-grained aspects of speech. Below, we discuss the response properties of each region in turn with the aim of understanding what processes might be enhanced by expert listening.

In the right pSTS region, previous studies have shown greater activation for human voice versus mixed-animal and other nonvocal sounds (Belin et al. 2000), for complex speech versus nonspeech stimuli (Vouloumanos et al. 2001), and for stimuli that are "speech-like" (Heinrich et al. 2008)—see Supplementary Table 1 for comparison of peak voxel coordinates across studies discussed in this section). Perhaps, most strikingly, the peak coordinates of this right pSTS region were within the radius of one functional voxel to the region found to be "speech selective" relative to carefully matched musical stimuli in a study of typical adults (Tervaniemi et al. 2006). Such a strong response preference for speech over music (exactly that found in actors in the present study) is entirely absent in our group of violinists who showed equivalently robust activation for music as well as speech. Thus, violinists show
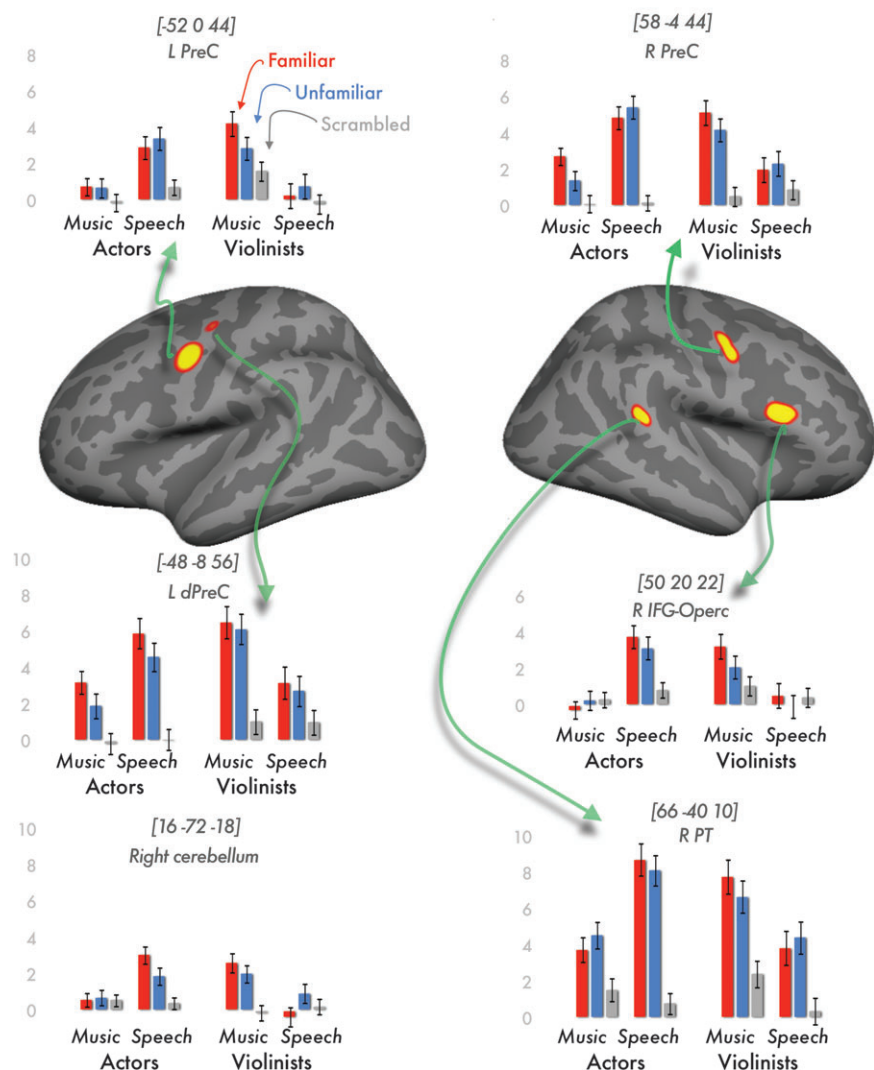
Figure 4. Effects of performing arts training, where violinists showed more activation for violin music than dramatic speech and actors showed more activation for dramatic speech than music. Abbreviations: L/R PreC = left or right precentral gyrus; L dPreC = left dorsal precentral gyrus; R IFG-Operc = right inferior frontal gyrus, opercular part; R PT = right planum temporale.

Table 3

Performing arts training: regions showing effects of performing arts training, where there was an interaction of Group and Stimulus Type (FWE-corrected P < 0.05), inclusively masked by violinists showing more activation for violin than dramatic speech (P < 0.001 voxelwise) and actors showing more activation for dramatic speech than for violin (P < 0.001 voxelwise).

Region	Hemi	x	y	z	Z-score	Clust size	Vln-V	Vln-A	Spch-V	Spch-A
Dorsal precentral	R	58	−4	44	5.75	48	5.5	3.5	n.s.	6.5
gyrus and sulcus	R	50	−4	54	4.97		5.7	3.7	n.s.	5.6
	L	−52	0	44	5.61	86	3.1	n.s.	n.s.	3.7
	L	−48	−8	56	5.05	19	5.2	3.4	n.s.	5.3
Right cerebellum	R	16	−72	−18	5.43	19	5.2	n.s.	n.s.	3.4
Right pars	R	50	20	22	6.16	154	3.2	n.s.	n.s.	3.8
opercularis										
Right pSTG	R	66	−40	10	5.74	17	5.2	3.8	4	7.8

Other information as in Table 1.

increased activation for music in regions associated with fine-grained analysis of speech and voices.

The left hemisphere analogue of this pSTS patch was also upregulated for violinists listening to music (relative to actors),

although, unlike the right pSTS, activation was not equivalent to speech. Again, this patch of cortex has been shown to exhibit sensitivity to higher-level speech effects. Here (within 1 functional voxel of the peak left STS voxel in the current study), Desai et al. (2008) showed that a cluster of voxels was correlated with the degree to which subjects categorically perceived sine-wave speech–generated consonant–vowel combinations. Möttönen et al. (2006) also found that subjects who perceived sine-wave speech as speech showed increased activation just laterally to the current focus (1 cm). Interestingly, Leech et al. (2009) found that subjects who were successful in learning to categorize nonspeech sounds after short-term naturalistic training also showed increases in this same region, although the extent of increase was not to the same extent as was observed in the present study for the violinists listening to violin. In tandem with the findings of Leech et al., the results of the current study hint that this left pSTS region is involved in the analysis and categorization of behaviorally relevant fine-grained acoustical detail. Thus, we

find that violinists show increased activation for violin music in regions that are involved in abstracting categories from meaningful auditory input. Importantly, activation in the same left and right pSTS coordinates was reported by Ohnishi et al. (2001) for listening to piano music in a mixed group of musicians compared with musically naive controls. In contrast, however, Baumann et al. (2007) found no expertise-related differences in temporal regions in a similar comparison of pianists and nonmusicians. Margulis et al. (2009) also found that violinists and flutists showed more activation in the same left pSTS region when listening to a Bach partita written for and performed by their own instrument. The fact that we found no effect of excerpt familiarity in this region suggests that the Margulis et al. effect may have been driven by instrumental expertise (as they suggest) and not by familiarity with the particular piece.

The left PT region activated by music in violinists and speech in all participants is primarily considered an auditory processing region because of its adjacency and strong connectivity to the primary auditory cortex (Upadhyay et al. 2008). Many previous neuroimaging studies have documented its response to both auditory speech (e.g., Zatorre et al. 2002) and music (e.g., Zatorre et al. 1994), with activation increasing with the auditory working memory demands of the task for both speech and nonspeech stimuli (Zatorre et al. 1994; Gaab et al. 2003; Buchsbaum and D'esposito 2009; Koelsch et al. 2009). Although left PT activation is not typically speech selective (Binder et al. 1996), it increases when speech comprehension is made more difficult at the perceptual or conceptual level (Price 2010). This may relate to increased demands on auditory working memory or top-down influences from regions involved in speech or music production (Zatorre et al. 2007). Left PT activation in the current study may therefore reflect increased auditory working memory or top-down influences from motor regions when participants are used to producing the sounds presented (speech in both groups and music in violinists).

Finally, activation in the left aSTG region has previously been associated with increasing speech intelligibility (Scott et al. 2004; Friederici et al. 2010; Obleser and Kotz 2010). This region is not speech specific, however, because it is also activated by melody and pitch changes and has been associated with the integration of sound sequences that take place over long versus short timescales (Price et al. 2005). The aSTG has also been implicated in syntactic processing of musical sequences (Koelsch 2005). Indeed, one of the effects of auditory expertise will be to enhance the recognition of auditory sequences over longer timescales.

In summary, the regions that we associate with audiomotor expertise are associated with fine-grained auditory analysis (right pSTS), speech categorization (left pSTS), audiomotor integration (left PT), and the integration of sound sequences over long timescales (left aSTG). The response of these regions to music in violinists suggests that the emergence of response selectivity may depend on long-term experience in perceiving and producing a given class of sounds.

Speech-Selective Activation Related to Information Content

Although we have focused above on regions that are sensitive to audiomotor experience, the majority of speech-selective regions (in bilateral middle and anterior STS, superior temporal gyrus [STG], and middle temporal gyrus) showed no significant upregulation in violinists listening to violin relative to violinists listening to speech. These speech- or language-selective regions can therefore be generalized to nonexpert populations because they were not dependent on the type of audiomotor expertise. The most likely explanation of these effects is that they reflect differences in acoustical and informational content of speech and music. For instance, English-language speech processing relies heavily on relatively quick changes in the prominence and direction of change in energy bands (formants), as well as the presence and timing of noise bursts and relative silence. In contrast, the acoustical signal from the violin is spectrally quite dense and stable over time (see Supplementary Fig. 1 for spectrogram comparing speech and music stimuli), with important acoustical information conveyed through changes in stepwise pitch, metrical rhythm, timbre, and rate and amplitude of frequency modulation (vibrato); see Carey et al. (1999) for a discussion of acoustical differences between speech and music. Thus, speech-selective activation may reflect in part this cross-domain difference in acoustical processing, as suggested by the results of Binder et al. (2000), who reported a similar region of cortex with increased activation for words, pseudowords, or reversed words versus tones, as well as those of Specht et al. (2009), who showed selective responses in this region when sounds were morphed into speech-like, but not music-like, stimuli. Alternatively, it may reflect higher-level language-specific processes (such as language comprehension, e.g., Narain et al., 2003) as well as more general differences in language versus musical processing—for a discussion of this point, see Steinbeis and Koelsch (2008). However, it is important to note that increased activation for speech compared with music, irrespective of auditory experience, does not imply that the same regions are not activated by other types of auditory or linguistic stimuli (Price et al. 2005).

The Effect of Performing Arts Training

Whereas both violinists and actors showed strong activation for speech in the regions discussed above, another shared network of primarily motor-related regions showed strongly selective responses for listening to the "expert" stimulus alone. Indeed, in some cases, activation for the nonexpert stimulus did not reach significance relative to the low-level scrambled baseline stimuli. It is quite remarkable that speech-related responses were enhanced for actors relative to violinists, given that violinists—like almost all humans—are also "speech experts," and therefore might not be expected to show less activation for speech than actors. This result suggests that expert performers listen fundamentally differently than normal listeners. In particular, actors listening to language may often need to respond by linking language comprehension to action in a more specific way than average listeners. Indeed, when expert hockey players (who talk about and play hockey all the time) understand hockey-action sentences, they show increased activation in premotor effector–related regions relative to people without hockey expertise (Beilock et al. 2008). "Working memory" (Koelsch et al. 2009) and sustained selective attention may also be differentially engaged in performing artists and other experts (see also Palmeri et al. 2004).

The effect of performing arts training was observed primarily in regions associated with motor control (e.g., left and right

dorsal premotor regions, the right pars opercularis, and right cerebellum), with an additional expertise-related activation in the right pSTG, in the vicinity of the right PT. Interestingly, in their study of novice pianists trained only for 5 days, Lahav et al. (2007) found increased activation in such bilateral premotor and right inferior frontal regions during passive auditory perception of excerpts from melodies that participants had learned to play, suggesting that some performing arts training effects may emerge quite early in the acquisition of expertise. The dorsal premotor and right cerebellum activations that were enhanced in musicians (or actors) are located in regions previously reported to be commonly activated by finger tapping and articulation (Meister et al. 2009). As fingers and articulation have distinct motor effectors, Meister et al. (2009) suggest a role in action selection and planning within the context of arbitrary stimulus–response mapping tasks. This explanation can explain the shared motor network that we observed for actors and violinists, given that different effectors are involved in the sounds they are expert in (mouth and hands/arms).

The observation that regions associated with motor planning were activated during purely perceptual tasks is consistent with many previous studies—see Scott et al. (2009) for review. For instance, Wilson et al. (2004) reported bilateral premotor activation for both producing and perceiving meaningless monosyllables in typical adults (see Supplementary Table 2 for peak coordinates) and interpreted this activation as potentially reflecting an auditory-to-articulatory motor mapping. Such motor involvement in passive perception has also been linked to theories of analysis-through-synthesis or "mirror" networks (e.g., Buccino et al. 2004). Our results extend these observations of motor activation during perceptual tasks from speech stimuli to violin music.

In a pair of experiments that could be considered the visual analogues to the current study, Calvo-Merino et al. (2005) showed that dancers and capoeira artists selectively increased activation in the same left (but not right) dorsal premotor region when watching the type of action that they themselves practiced. Calvo-Merino et al. (2006) further showed that this upregulation in motor activation was not simply a product of exposure or familiarity with seeing particular movements. Here, male and female ballet dancers viewed videos of dance moves that are typically performed by only one gender but are equally visually familiar to both genders. Again, the same left premotor region (but not right) showed greater activation for the movement that the subject had experience in performing and not just observing.

Given these findings, it is intriguing that the effect of performing arts training in actors and violinists was observed irrespective of whether the stimuli were familiar or unfamiliar. In this respect, it is important to point out that the familiar music excerpts were not only highly familiar but also had been played by almost all violinists, whereas the unfamiliar excerpts were almost completely unknown to them. The similarity in the effect sizes for familiar and unfamiliar stimuli suggests that motor activation in response to the expert sounds was not a reenactment of the whole sequence of motor plans but was more likely to relate to familiarity with well-established and stereotyped patterns that are common to familiar and un-familiar sequences. A speculative interpretation is that activations reflect the prediction and integration of upcoming events in time. In this vein, Chen et al. (2008) showed the same

bilateral premotor activation when nonmusicians anticipated or tapped different rhythms.

In addition to the motor planning regions discussed above, performing arts training also increased activation in the right pars opercularis and the right STG, on the border with the right PT. These effects were not observed in the left hemisphere homologues that were strongly activated by speech in violinists as well as actresses. The activation of right hemisphere language regions for sounds that the performers were experts in reproducing is frankly an unexpected finding and one that bears further investigation. For example, future studies could investigate whether right pars opercularis and right STG activations increase for nonexpert sounds when the task requires a motor response (e.g., an auditory repetition task). Such an outcome would suggest a role for these regions in sensorimotor processing that might be activated implicitly (irrespective of task demands) for stimuli that the performer is used to producing.

Conclusions

We have disambiguated 3 different patterns of auditory speech processing. The first set of regions was speech selective but also activated by music in violinists. These regions were associated with expertise in fine-grained auditory analysis, speech categorization, audiomotor integration, and the audi-tory sequence processing. The second set of regions was associated with the many acoustic and linguistic properties that are selective for speech relative to music in violinists as well as actors. The third set of regions was selective to the sound that the performers were expert in [speech > music in actors and music > speech in violinists]. This performing arts training effect was observed in regions associated with motor planning and right hemisphere "language regions." We tenta-tively suggest that such activation may play a role in predicting and integrating upcoming events at a finer resolution than in "normal" listeners.

Together, these results illustrate that speech-selective neural response preferences in multiple perisylvian regions are changed considerably by experience and indeed sometimes even reversed. These effects were not driven by individuals' familiarity with particular stimuli but were instead associated with the entire stimulus category. Such results should make us wary of ascribing innate response preferences to certain brain regions that show consistent "category-specific" response preferences in most individuals, in that these may simply be the result of experience and task demands.

References

Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L. 2007. A network for audiomotor coordination in skilled pianists and non-musicians. Brain Res. 1161:65-78.

Beilock SL, Lyons IM, Mattarella-Micke A, Nusbaum HC, Small SL. 2008. Sports experience changes the neural processing of action language. Proc Natl Acad Sci U S A. 105:13269-13273.

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

Bermudez P, Lerch JP, Evans AC, Zatorre RJ. 2009. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cereb Cortex. 19:1583-1596.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 10:512-528.

Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW. 1996. Function of the left planum temporale in auditory and linguistic processing. Brain. 119:1239-1247.

Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund HJ, Rizzolatti G. 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. Neuron. 42:323-334.

Buchsbaum B, D'Esposito M. 2009. Repetition suppression and reactivation in auditory-verbal short-term recognition memory. Cereb Cortex. 19(6):1474-1485.

Calvo-Merino B, Glaser DE, Grèzes J, Passingham RE, Haggard P. 2005. Action observation and acquired motor skills: an FMRI study with expert dancers. Cereb Cortex. 15:1243-1249.

Calvo-Merino B, Grezes J, Glaser D, Passingham R, Haggard P. 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. Curr Biol. 16:1905-1910.

Carey MJ, Parris ES, Lloyd-Thomas H. 1999. A comparison of features for speech, music discrimination. ICASSP, IEEE International Conference on Accoustics, Speech and Signal Processing. 1: 149-152.

Chen JL, Penhune VB, Zatorre RJ. 2008. Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex. 18:2844-2854.

Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 9:179-194.

Desai R, Liebenthal E, Waldron E, Binder JR. 2008. Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci. 20:1174-1188.

Friederici AD, Kotz SA, Scott SK, Obleser J. 2010. Disentangling syntax and intelligibility in auditory language comprehension. Hum Brain Mapp. 31(3):448-457.

Friston K, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS. 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp. 2:189-210.

Gaab N, Gaser C, Zaehle T, Jancke L, Schlaug G. 2003. Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. Neuroimage. 19:1417-1426.

Gaab N, Schlaug G. 2003. Musicians differ from nonmusicians in brain activation despite performance matching. Ann N Y Acad Sci. 999:385-388.

Gaser C, Schlaug G. 2003. Brain structures differ between musicians and non-musicians. J Neurosci. 23:9240-9245.

Harley E, Pope W, Villablanca J, Mumford J, Suh R, Mazziotta J, Enzmann D, Engel S. 2009. Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. Cereb Cortex. 19:2746-2754.

Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. 2008. Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. J Cogn Neurosci. 20:1737-1752.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 17:4302-4311.

Kherif F, Josse G, Seghier SL, Price CJ. 2009. The main sources of intersubject variability in neuronal activation for reading aloud. J Cogn Neurosci. 21:654-668.

Koelsch S. 2005. Neural substrates of processing syntax and semantics in music. Curr Opin Neurobiol. 15:207-212.

Koelsch S, Schulze K, Sammler D, Fritz T, Müller K, Gruber O. 2009. Functional architecture of verbal and tonal working memory: an FMRI study. Hum Brain Mapp. 30:859-873.

Krampe RT, Ericsson KA. 1996. Maintaining excellence: deliberate practice and elite performance in young and older pianists. J Exp Psychol Gen. 125:331-359.

Lahav A, Saltzman E, Schlaug G. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci. 27:308-314.

Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP. 2009. Brain activation during anticipation of sound sequences. J Neurosci. 29:2477-2485.

Leech R, Holt L, Devlin J, Dick F. 2009. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci. 29:5234-5239.

Levitin DJ, Menon V. 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. Neuroimage. 20:2142-2152.

Margulis EH, Mlsna LM, Uppunda AK, Parrish TB, Wong PC. 2009. Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum Brain Mapp. 30:267-275.

Meister IG, Buelte D, Staedtgen M, Boroojerdi B, Sparing R. 2009. The dorsal premotor cortex orchestrates concurrent speech and fingertapping movements. Eur J Neurosci. 29:2074-2082.

Möttönen R, Calvert GA, Jääskeläinen IP, Matthews PM, Thesen T, Tuomainen J, Sams M. 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage. 30:563-569.

Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen S, Matthews PM. 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex. 13:1362-1368.

Obleser J, Kotz SA. 2010. Expectancy constraints in degraded speech modulate the language comprehension network. Cereb Cortex. 20:633-640

Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, Katoh A, Imabayashi E. 2001. Functional anatomy of musical perception in musicians. Cereb Cortex. 11:754-760.

Palmeri TJ, Wong AC, Gauthier I. 2004. Computational approaches to the development of perceptual expertise. Trends Cogn Sci. 8:378-386.

Price C, Thierry G, Griffiths T. 2005. Speech-specific auditory processing: where is it? Trends Cogn Sci. 9:271-276.

Price CJ. 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci. 1191:62-88.

Scott SK, Johnsrude IS. 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26:100-107.

Scott SK, McGettigan C, Eisner F. 2009. A little more conversation, a little less action—candidate roles for motor cortex in speech perception. Nat Rev Neurosci. 10:295-302.

Scott SK, Rosen S, Wickham L, Wise RJ. 2004. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. J Acoust Soc Am. 115:813-821.

Specht K, Osnes B, Hugdahl K. 2009. Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic "sound morphing" technique. Hum Brain Mapp. 30:3436-3444.

Steinbeis N, Koelsch S. 2008. Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. PLoS One. 3(5):e2226.

Tervaniemi M, Szameitat AJ, Kruck S, Schröger E, Alter K, De Baene W, Friederici AD. 2006. From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. J Neurosci. 26:8647-8652.

Thierry G, Giraud AL, Price C. 2003. Hemispheric dissociation in access to the human semantic system. Neuron. 38:499-506.

Upadhyay J, Silver A, Knaus TA, Lindgren KA, Ducros M, Kim DS, Tager-Flusberg H. 2008. Effective and structural connectivity in the human auditory cortex. J Neurosci. 28:3341-3349.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci. 13:994-1005.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. 2004. Listening to speech activates motor areas involved in speech production. Nat Neurosci. 7:701-702.

Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. Trends Cogn Sci. 6:37-46.

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci. 8:547-558.

Zatorre RJ, Evans AC, Meyer E. 1994. Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci. 14:1908-1919.