Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects

Stefan Elmer[1], Martin Meyer[1,2] and Lutz Jäncke[1,2,3]

[1]Division of Neuropsychology, Institute of Psychology, University of Zurich, CH-8050 Zurich, Switzerland, [2]Center for Integrative Human Physiology, 8057 Zurich, Switzerland and [3]International Normal Aging and Plasticity Imaging Center (INAPIC), CH-8006 Zürich, Switzerland

Address correspondence to Stefan Elmer, Division of Neuropsychology, Institute of Psychology, University of Zurich, Binzmühlestrasse 14/25, CH-8050 Zurich, Switzerland. Email: s.elmer@psychologie.uzh.ch.

The perception of rapidly changing verbal and nonverbal auditory patterns is a fundamental prerequisite for speech and music processing. Previously, the left planum temporale (PT) has been consistently shown to support the discrimination of fast changing verbal and nonverbal sounds. Furthermore, it has been repeatedly shown that the functional and structural architecture of this supratemporal brain region differs as a function of musical training. In the present study, we used the functional magnetic resonance imaging technique, in a sample of professional musicians and nonmusicians, in order to examine the functional contribution of the left PT to the categorization of consonant–vowel syllables and their reduced-spectrum analogues. In line with our hypothesis, the musicians showed enhanced brain responses in the left PT and superior discrimination abilities in the reduced-spectrum condition. Moreover, we found a positive correlation between the responsiveness of the left PT and the performance in the reduced-spectrum condition across all subjects irrespective of musical expertise. These results have implications for our understanding of musical expertise in relation to segmental speech processing.

Keywords: fMRI, musicians, speech processing, transfer effects

Introduction

The impact of intensive instrumental training on auditory processing has been well documented in cross-sectional (Pantev et al. 2001; Baumann et al. 2008) and longitudinal studies (Hyde et al. 2009; Moreno et al. 2009). In fact, to date, there is an abundance of evidence that shows better discrimination skills in professional musicians in the domain of music processing. These skills are indeed associated with functional and structural adaptations throughout the auditory system (Schlaug et al. 1995; Munte et al. 2002; Jancke 2009). In particular, a functional and structural asymmetry of the planum temporale (PT) has been increasingly accepted as a neural substrate for musical expertise (Ohnishi et al. 2001; Bermudez et al. 2009), especially for absolute pitch possessors (Schlaug et al. 1995; Keenan et al. 2001; Wilson et al. 2009). In this context, Ohnishi et al. (2001) reported an increased leftward PT asymmetry among musicians while they passively listened to piano melodies. A similar lateralization of music processing in professional musicians was also found in behavioral and electrophysiological studies (Hirshkowitz et al. 1978; Bever and Chiarello 2009); thereby, suggesting that the functional capacity of the left supratemporal plane, as well as the PT, may increase with musical training.

In the domain of speech processing, one of the most important cues that conveys phonetic information is the so-called voice-onset-time (VOT); this is a fast temporal cue, which can be defined as the time between the release of a stop consonant and the onset of vocal folds vibrations. For example, based on the discrimination of the VOT, a German listener is able to distinguish the consonant–vowel (CV) syllable /da/ from /ba/. These fast temporal cues are not only specific to speech processing but they occur likewise during music perception in the form of short, sharp, or impulsive tones. Meanwhile, there is mounting evidence illustrating that among other brain regions, the left posterior supratemporal plane that accommodates auditory-related cortex and in particular the PT, contain neurons that are preferentially driven by transient acoustic features in CV syllables and their nonspeech analogues (Jancke et al. 2002; Zaehle et al. 2008). One should consider that even though the PT was originally considered a speech-specific brain region (Galaburda et al. 1978), researchers have been recognizing that the human PT is not a dedicated language processor. Rather, it is engaged in the analysis of many types of complex sounds. In fact, Griffiths and Warren (2002) proposed a model of the PT as a computational engine for the segregation and matching of spectrotemporal patterns. The relevance of this brain region for spectrotemporal processing becomes particularly evident in clinical and developmental contexts (Heiervang et al. 2000). In fact, it has been shown that subjects affected by acquired brain lesions and aphasia (Efron 1963; Swisher and Hirsh 1972), children with general language-learning disabilities (Tallal and Stark 1981), and children and adults with dyslexia (Tallal 1980), have difficulties with categorizing rapidly changing auditory patterns, such as phonemes. For example, individuals affected by a variety of lesions covering the left PT have been shown to demonstrate a number of auditory discrimination and speech comprehension deficits mainly concerning phonetic processing (Caplan et al. 1995; Shapleske et al. 1999). Also in the context of developmental dyslexia, a condition which is often associated with a difficulty in manipulating sounds and segmenting words into syllables and syllables into phonemes (Lundberg and Hoien 1989; Rumsey 1992), the left PT was proposed to be critically altered (Shapleske et al. 1999).

Notably, even though speech and music are perceptually distinct, they share many acoustic commonalities. In fact, both signals convey acoustic information by means of timing, pitch, and timbre cues (Kraus and Chandrasekaran 2010). Exactly these acoustic commonalities between speech and music have aroused increasing interest in the scientific community; thus, it is our aim to understand whether long-term instrumental training may favor different aspects of sublexical speech processing. Whereas transfer effects from musical training to domains that are not necessarily directly linked to music have

been addressed in children (Schlaug et al. 2005; Magne et al. 2006) and adults (Schlaug et al. 2005), there exists to date only a handful of studies that have investigated transfer effects excerpted by musical training on speech processing (Schon et al. 2004; Marques et al. 2007; Oechslin et al. 2010). These previous research findings (Schon et al. 2004; Marques et al. 2007) consistently revealed that musicians are more skilled than nonmusicians at processing pitch contour information in speech stimuli but only when the pitch variations are difficult to detect. However, a question that has not yet been investigated by means of the functional magnetic resonance imaging (fMRI) technique, is whether long-term instrumental training and the fine-graded auditory skills associated with it have an influence on phonetic and temporal processing. To our knowledge, only 2 studies addressed a similar issue by using a behavioral approach (Hillenbrand et al. 1990) or the electroencephalography method (Marie et al. 2011). Hillenbrand et al. (1990) found evidence for the notion that musicians are not faster or even better in discriminating synthetic speech sounds. Furthermore, Marie et al. (2011), who examined the influence of musical expertise on metric aspects of speech processing, could show that musical expertise has an influence on the processing of metric structures of words in the time range between 150 and 700 ms. In particular, the authors could show that musicianship was associated with 1) an enhanced P200 amplitude reflecting the automatic detection of the syllable temporal structure, 2) with a reduced N400 response reflecting the integration of metric structure and its influence on word comprehension, and 3) with an enlarged P600 response allegorizing the reanalysis of metric violations. The findings of Marie et al. (2011) are of noticeable relevance because the authors could show that musicians process the syllabic structure of words in a different manner than nonmusicians. This previous finding motivated us to investigate the categorization of syllables and their reduced-spectrum analogues in the context on an fMRI design.

The present fMRI study was specifically designed in order to investigate whether the previously reported PT asymmetry in professional musicians may favor the acoustic analysis of fast temporal cues in CV syllables and their reduced-spectrum analogues. With this purpose in mind, we measured professional musicians and nonmusicians during a phonetic categorization task that consisted of assigning 2 CV syllables and their white noise analogues to the categories /ka/ or /da/. Based on previous studies, which adopted similar stimulus material (Jancke et al. 2002; Zaehle et al. 2008), we hypothesized that professional musical training is associated with enhanced responses in the left PT reflecting musical expertise. Furthermore, in reference to the enhanced fine-graded auditory skills often observed in professional musicians, we expected that the experts would perform better than the laymen on the phonetic categorization task.

Materials and Methods

Subjects
Twelve professional musicians without absolute pitch (four females/eight males; mean age = 24.9 years, standard deviation [SD] = 5.5; mean practice hours since childhood = 15677.3, SD = 7055; mean age of practice commencement = 5.9, SD = 1.3) and 13 control subjects without formal musical education (five females/eight males; mean age = 25.3 years, SD = 3.5) participated in this study. All musicians we measured commenced their musical training before the age of 7

years, and none of them was a professional vocalist. The musician group consisted of 4 subjects playing flute, 4 pianists, 3 violinists, and one cellist (primary musical instruments). Nonmusicians were selected on the basis that they never participated in musical practice with the exception of flute lessons at school. The subjects reported no past or current neurological, psychiatric, or neuropsychological problems and denied consumption of drugs or illegal medication. Subjects were paid for participation. The local ethics committee (local ethic committee of the canton of Zürich, Switzerland) approved the study and written informed consent was obtained from all participants. All subjects were consistent right-handers as revealed by the Annett Handedness Inventory (Annett 1970).

Behavioral Data

History of Musical Training
The history of musical training was assessed by an in-house questionnaire. This questionnaire was used to evaluate the age of commencement, instruments played, and the estimated number of training hours across the life span. In particular, the subjects estimated the total number of training hours they performed per week in the following periods of life (age): 0-7, 8-10, 11-13, 14-16, etc. Based on these values, we extrapolated the total amount of training hours across life span for all subjects.

Music Aptitudes
All subjects performed an auditory test, in order to examine their developmental and stabilized music aptitudes, as well as their music achievement (Gordon 1989). This test consisted of 30 successive trials in which subjects compared pairs of piano melodies and then decided whether the melodies were equivalent (i.e., the exactly same acoustic pattern), rhythmically different, or tonally different. Due to technical difficulties, one subject of the control group could not be tested.

Cognitive Capability
In order to rule out differences in intelligence between groups, we adopted the KAI test (http://www.testzentrale.ch/). This procedure permits to measure the actual cognitive capability (fluid intelligence) and is based on working memory and speed of information processing.

Stimulus Material
All subjects heard 2 different classes of auditory stimuli (speech and reduced-spectrum conditions) in the context of a phonetic categorization task. The speech condition consisted of the German CV syllables /ka/ (voiceless initial consonant) and /da/ (voiced initial consonant) which were digitally recorded by a trained phonetician at a sampling rate of 44.1 kHz/sampling depth of 16-bit (Jancke et al. 2002). The onset, duration, intensity, and fundamental frequency of the 2 CV syllables were edited and synchronized by means of a speech editor (Adobe Audition). The criterion for temporal alignment of the syllables was the onset of articulatory release. The total duration of the syllables was of about 350 ms (for both syllables). Furthermore, the VOT for /da/ was approximately of 13 ms and that for /ka/ of about 53 ms. The VOT was defined as the interval between the noise burst produced at consonant release and the onset of the waveform periodicity associated with vocal cord vibrations (Lisker and Abramson 1967). Evidently, the 2 CV syllables /ka/ and /da/ did not vary only in VOT but also in place of articulation; whereas /k/ is a voiceless velar plosive, /d/ is a voiced alveolar plosive. In natural speech conditions, different places of articulation are known to be associated with different VOTs (Lisker and Abramson 1967).

The reduce-spectrum condition consisted of white noise analogues (white noise /ka/ = /wnka/; white noise /da/ = /wnda/), which were synthesized from the 2 CV syllables used in the natural speech condition. For this purpose, we used a variation of the procedure described by Shannon et al. (1995). In particular, spectral information was removed from the CV syllables by replacing the frequency-specific information in a broad frequency region with band-limited white noise (band 1: 500–1500 Hz, band 2: 2500–3500 Hz). Amplitude and temporal cues were preserved in each spectral band, resulting in double-band-pass filtered noise with temporal CV-amplitude dynamics.

The Experimental Procedure

Before applying the fMRI protocol, the 4 auditory stimuli (/ka/, /da/, /wnka/, and /wnda/) were presented to every subject outside the scanner, in order to familiarize the participants with the different sounds. The 4 stimuli were presented in a randomized order across 3 runs, each of 12-min duration. In the context of a phonetic categorization task, the subjects were instructed to decide whether the presented stimuli belong either to the category /ka/ or /da/ by pressing 1 of 2 different buttons. The subjects were instructed to place emphasis on a good performance rather than to respond as quickly as possible. The 3 runs were counterbalanced across subjects and groups. Each of the 3 runs included a series of 60 stimuli (15 /ka/, 15 /da/, 15 /wnka/, and 15 /wnda/), and 177 empty trials were used to define baseline and to avoid expectancy. The auditory stimuli were jittered with an ISI corresponding to 2–5 repetition times (TRs). Auditory stimuli presentation and collection of behavioral responses were controlled by the "presentation" software (Neurobehavioral Systems, Version 0.70). The entire experimental procedure was accomplished only inside the scanner without collecting behavioral data in a completely quiet environment outside the scanner.

FMRI Data Acquisition and Processing

During scanning, the participants were instructed to keep their eyes open and to focus on the fixation cross presented on the screen. Binaural auditory stimuli were presented by a digital playback system and included a high-frequency shielded transducer system. The acoustic transmission system included a piezoelectric loudspeaker enabling the transmission of strong sound pressure levels (105 dB) with excellent attenuation characteristics (Jancke et al. 2002). The approximate delivered intensity level in the scanner was about 90 dB. These loudspeakers are embedded in tightly occlusive headphones allowing unimpeded conduction of the stimulus with good suppression of ambient scanner noise by about 20 dB. The headphones we used for the experiment had a frequency response ranging from 100 Hz to 16 KHz. Additionally, noise-protection ear plugs within the loudspeakers provided an additional noise attenuation of about 15–20 dB, resulting in a total noise attenuation of 35–40 dB. The acoustic transmission system allows stimulation of acoustic stimuli with relatively few distortions.

A Philips Intera 3-T whole-body MR unit (Philips Medical Systems, Best, Netherlands) that is equipped with an 8-channel Philips SENSE head coil was used to acquire fMRIs at the University Hospital, Zurich. Functional data were obtained from 242 whole-head scans per run using a Sensitivity Encoded (SENSE) single-shot echo-planar imaging (EPI) technique (TR = 3000 ms, acquisition time [TA] = 2000 ms, time echo = 35 ms, flip angle = 78°, field of view = 220 mm, acquisition matrix = 80 × 80, 33 transverse slices, voxel size = 1.72 × 1.72 × 4.00 mm). We adopted a TA of 2 s and a TR of 3 s, in order to avoid an overlap of scanner noise and stimulus presentation (Jancke et al. 2002).

MRI data analysis was performed using MATLAB 2010 (Mathworks Inc., Natick, Massachusetts) and the SPM8 software package (Institute of Neurology, London, UK). All images were realigned to the first image of each run, spatially normalized into standard stereotactic MNI space (EPI template provided by the Montreal Neurological Institute), interpolated to a voxel size of 2.00 × 2.00 × 2.00 mm, and spatially smoothed using a 8-mm full-with at half-maximum Gaussian kernel.

Statistical Analysis

Statistical analysis was based on the general linear model (GLM). Due to the experimental design, an event-related analysis was conducted. The standardized canonical hemodynamic response was applied to model the blood oxygen level–dependent response to each of the 4 auditory stimuli (/ka/, /da/, /wnka/, and /wnda/). In addition, the behavioral responses were modeled as events. On the first-level analysis, 4 comparisons of interest (long vs. short VOT and short vs. long VOT) were implemented as linear contrasts: 1) /ka/ versus /da/, 2) /da/ versus /ka/, 3) /wnka/ versus /wnda/, and 4) /wnda/ versus /wnka/. These specific contrasts not only permit us to investigate the processing associated with short and long VOTs (or noise-onset-time) but are also more straight forward than making comparisons with empty trials. The resulting set of voxel values for each contrast constitutes a statistical parametric map of the T-statistic. For further group-level analysis, we specified the SPM8 factorial design built up by 2 independent variables and this resulted in a 2 × 4 ANOVA (2 groups × 4 contrasts). In our report and discussion of the ANOVA results (SPM8), only significant clusters of activation were considered (family-wise error corrected [FWE]-corrected α-level 0.05, $k \geq 150$ voxels). In order to elucidate the significant results of the ANOVA, we performed region of interest (ROI) analyses using spheres of 7-mm radius centered at maximal local F values. The software Marsbar (http://marsbar.source-forge.net) was used to define 7-mm radius spherical ROIs (ROI PT-LEFT left PT [−56, −36, 22]; ROI PMC-LEFT left premotor cortex [−22, 24, 54]). These 2 ROIs were also mirrored to the right hemisphere, in order to elucidate hemispheric asymmetries (ROI PT-RIGHT [56, −36, 22]; ROI PMC-RIGHT [22, 24, 54]). Mean BETA values were read out by in-house programmed MATLAB (http://www.mathworks.com/) scripts and were further analyzed by means of t-tests (SPSS, http://www.spss.com/). All statistical tests performed with the data extracted from the ROIs were corrected for multiple comparisons by using the Bonferroni procedure.

ROI Analyses and BETA Values

In the present work, ROI analyses were performed by extracting mean BETA values from the post hoc-defined ROIs. The meaning of BETA values can be extrapolated by considering that the fMRI analyses were based on the GLM, which can be allegorized in a simplified manner by the following regression equation: $Y = BX + E$; were Y is the measured signal, X is the design matrix, E is error, and B is a vector of BETA weights, which estimate the fit of the model (design matrix) to the measured signal (sums of squares differences between the predicted model and the measured signal). Therefore, large positive (or negative) BETA weights typically indicate that a particular voxel exhibits strong activation (or deactivation) during the modeled experimental condition relative to baseline. In the present work, BETA values were used to evaluate 1) the relationship between the total number of hours of musical training and PT responsiveness, 2) hemispheric asymmetries, and 3) the relationship between the total number of correct responses and PT responsiveness. Since the whole-head ANOVA analysis yielded a main effect of group, for each person, the BETA values associated with the 4 contrasts (/ka/ vs. /da/, /da/ vs. /ka/, /wnka/ vs. /wnda/, and /wnda/ vs. /wnka/) were extracted from the respective ROIs and an average was computed. This average value was used for all post hoc ROI analyses (i.e., t-tests and correlations).

Results

Biographical Data and Musical Aptitude

The 2 groups (musicians and nonmusicians) did not differ in age and general cognitive capability (age $t_{25} = -0.209$, $P = 0.83$; cognitive capability $t_{25} = 1.75$, $P = 0.09$; t-tests for independent samples). Notably, the musicians performed significantly better in the tonal ($t_{24} = 4.33$, $P = 0.001$) and in the rhythmical ($t_{24} = 3.19$, $P = 0.004$) parts of the test for musical aptitude (t-tests for independent samples). These results show that the 2 groups were comparable in age and cognitive capability; however, they differed significantly in terms of musical expertise.

In-Scanner Behavioral Results

In a first statistical analysis, we tested whether the total number of correct responses for each group and condition significantly differed from chance by using one-sample t-tests (one-tailed) and a test value of 22.5 (by a total number of 45 presented stimuli this value corresponds to 50%). The results of this statistical procedure indicate that the performance of the 2 groups significantly differed from chance (musicians: /ka/, $t_{11} = 112.85$, $P = 0.001$; /da/, $t_{11} = 42.90$, $P = 0.001$, /wnka/, $t_{11} = 5.585$, $P = 0.001$, /wnda/, $t_{11} = 2.931$, $P = 0.014$; controls: /ka/, $t_{12} = 66.929$, $P = 0.001$, /da/, $t_{12} = 93.602$,

$P = 0.001$, /wnda/, $t_{12} = -3.003$, $P = 0.011$) or at least evidenced a statistical trend (controls: /wnka/, $t_{12} = -1.592$, $P = 0.068$).

The number of correct responses in the scanner was evaluated by a 2 × 4 ANOVA (2 groups, 4 conditions; repeated measures). This statistical procedure lead to significant group ($F_{1,11} = 30.58$, $P = 0.001$), condition ($F_{3,11} = 66.73$, $P = 0.001$), and group × condition effects ($F_{3,11} = 29.36$, $P = 0.001$). To further explore the main group effect and the group × condition interaction, we compared the performance of the 2 groups during the 4 conditions by performing t-tests for independent samples (two-tailed, Bonferroni-corrected). In line with the information derived from the descriptive statistic (Fig. 1), the performance of the 2 groups significantly differed during the reduced-spectrum condition (/wnka/, $t_{23} = 4.75$, $P = 0.001$; /wnda/, $t_{23} = 4.13$, $P = 0.001$) but did not differ during the "speech condition" (/ka/, $t_{23} = -0.053$, $P = 0.958$; /da/, $t_{23} = -0.967$, $P = 0.344$). Furthermore, as shown in Figure 1, the main effect of the condition could be attributed to a lower number of correct responses in both groups during the "reduced-spectrum condition" (/ka/ vs. /wnka/, $t_{24} = 6.489$, $P = 0.001$; /ka/ vs. /wnda/, $t_{24} = 7.589$, $P = 0.001$; /da/ vs. /wnka/, $t_{24} = 6.980$, $P = 0.001$; /da/ vs. /wnda/, $t_{24} = 8.069$, $P = 0.001$; t-tests for paired samples, Bonferroni-corrected). A further analysis of the reduced-spectrum condition revealed a statistical trend in the musician group only ($t_{11} = 2.19$, $P = 0.051$); thus, suggesting that the musicians experienced more difficulty when categorizing the /wnda/, as opposed to the /wnka/ stimuli. In summary, the in-scanner behavioral data clearly illustrate that 1) even though both groups had more difficulty to categorize the stimuli related to the reduced-spectrum condition, 2) the musicians performed better than the controls.

These results are in line with previous studies, which show improved fine-graded auditory skills in professional musicians (Kraus and Chandrasekaran 2010). The behavioral results are shown in Figure 1 (upper part).

fMRI Results

In order to examine brain responses in regions previously shown to be involved in phonetic perception and temporal categorization, the functional activation maps of professional musicians and control subjects were compared. With this purpose in mind, we bi-directionally contrasted the auditory signals with shorter and longer VOT (or noise-onset-time for the reduced-spectrum stimuli) and implemented these contrasts in a full factorial analysis (ANOVA). In line with our hypothesis, the factorial design applied in this study revealed a main effect of group in the left PT (peak in Montreal Neurological Institute coordinates: $x = -56$, $y = -36$, $z = 22$; $F = 45.94$) and in the left premotor cortex (PMC, peak in Montreal Neurological Institute coordinates: $x = -22$, $y = 24$, $z = 54$; $F = 44.58$). An accurate visual inspection attested that the peak we revealed in the left supratemporal plane was indeed situated on the lower bank of the sylvian fissure and can, therefore, be collocated to the PT. The 2 main group effects also remain after a more conservative FWE-corrected threshold of $P < 0.001$. Neither the main effect of contrast nor the group × contrast interaction reached significance, not even at an uncorrected threshold of $P < 0.001$. Furthermore, similar results were obtained by contrasting the 4 auditory stimuli with empty trials and by implementing these contrasts to a second-level ANOVA analysis (FWE-corrected, cluster extent of 90 voxels, $P < 0.05$). In particular, the modulation of the left PT as a function of
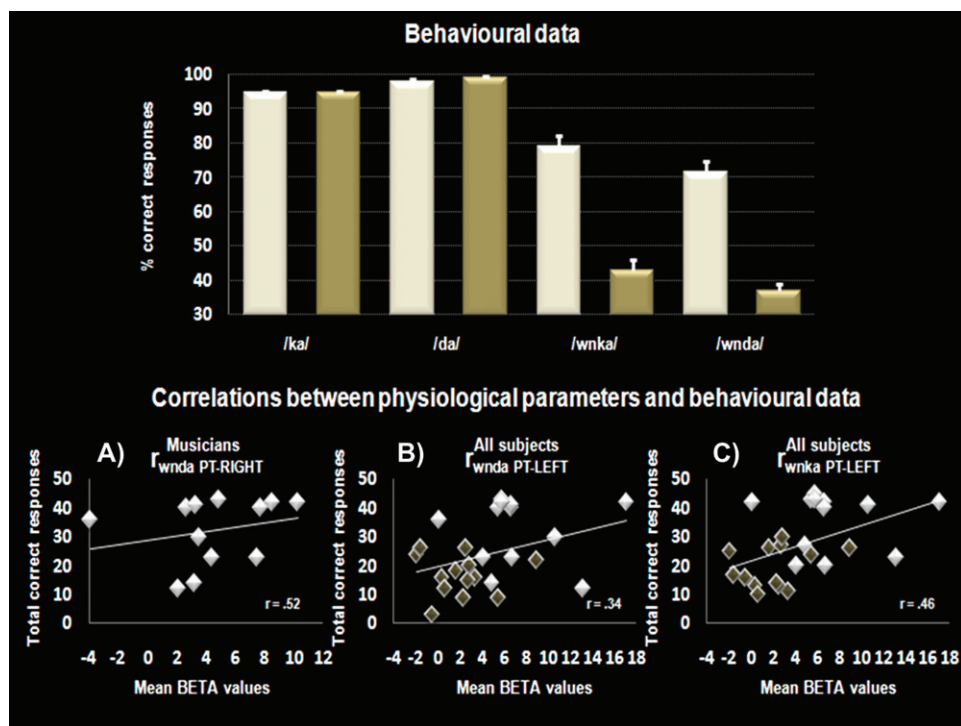
Figure 1. The upper plot displays the percentual number of correct responses for both groups and each stimulus (i.e., /ka/, /da/, /wnka/ and /wnda/). The lower part of the figure shows significant correlations between the total number of correct responses and PT activity. (*A*) Musicians: positive correlation between right PT activity and number of correct /wnda/ responses. (*B*) All subjects: positive correlation between left PT activity and number of correct /wnda/ responses. (*C*) All subjects: positive correlation between left PT activity and number of correct /wnka/ responses. White = musicians, gold = controls. The bars depict standard error.

expertise (main effect of group) was still present and situated at the same anatomical coordinates as reported above, that is (–54, –36, 22). These coordinates correspond to the PT according to the probability maps of anatomical landmarks provided by Westbury et al. (1999), with a probability range of 46–65%. Figure 2 indicates that the main effect of group in the left PT was associated with enhanced responses in the musician group in comparison to the controls. The opposite pattern was observed in the left PMC. In fact, in this latter region, the musicians showed reduced activity in contrast to the controls.

Hemispheric Asymmetries

In order to investigate functional asymmetries of the PT as a function of musical training, we statistically compared the mean BETA values extracted from ROI PT-LEFT and ROI PT-RIGHT within the 2 groups. Whereas the musician group clearly showed a leftward asymmetry (t_{11} = 2.23, P = 0.047; t-test for dependent samples, two-tailed), the controls were characterized by a rightward asymmetry (t_{12} = –2.96, P = 0.012, t-test for dependent samples, two-tailed). The same pattern of brain responses, as depicted in Figure 2, was also visible in each of the 4 contrasts implemented in the full factorial design (data not shown here). Since the whole-head ANOVA did not reveal group differences in the PT-RIGHT, the reduced responsiveness observed in the PT-LEFT in the control subjects cannot be explained by a right-sided enhancement of activity. Figure 2 shows an overview of the results.

Due to the fact that the ROI analysis shown in Figure 2 points to an inverse relationship between the activity of the left PT and the left PMC, we correlated the mean BETA values from ROI PT and PMC in the whole sample of subjects. As expected, this analysis revealed a negative nonparametric correlation

(according to Spearman's rho) between the 2 left hemispheric ROIs (r = –0.392; P = 0.026, one-tailed). This result suggests a reciprocal modulation between the PT and the PMC irrespective of group. A similar left-hemispheric relationship was not found within the 2 groups. We also did not reveal a significant relationship between the right-hemispheric ROIs within the 2 groups or in the whole sample of subjects.

Correlations between Physiological Parameters and Biographical/Behavioral Data

In order to examine whether the brain responses in the left auditory cortex were related to the number of hours of musical training, we nonparametrically correlated (employing Spearman's rho, two-tailed) the mean BETA values extracted from ROI PT-LEFT with the total number of hours of music practice since childhood as estimated by the subjects. This statistical analysis resulted in a significant negative correlation (r = –0.589; P = 0.044, Fig. 1) and leads us to suggest that even though the musicians showed overall enhanced responses of the auditory association cortex, this activity was accessorily modulated by the total number of hours of musical training. The correlation between age of commencement and the BETA values obtained from PT-LEFT did not reach significance (r = –0.237, P = 0.459). This is most likely the case since the musicians in this study began their music lessons before the age of 7 years; thus, the variability of age of practice onset was too small.

A second main question that we addressed by correlative post hoc analyses was whether the number of correct responses in the 2 groups during the reduced-spectrum condition (i.e., /wnka/ and /wnda/) was related to the BETA values obtained from PT-LEFT and PT-RIGHT. To explore this
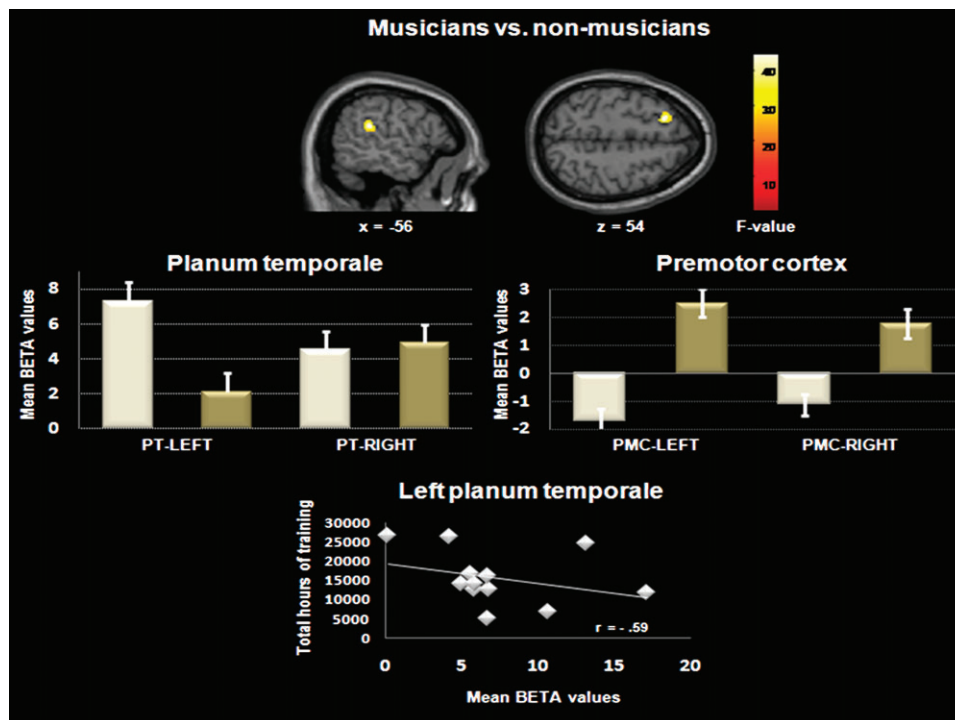
Figure 2. This figure depicts the main group effect (musicians vs. nonmusicians) in the left PT and in the left PMC with the related mean BETA values. The lower regression indicates the significant correlation between the total hours of musical training across life span and the activity in the left PT. White = musicians, gold = controls. The bars depict standard error.

question, we computed nonparametric correlations (Spearman's rho, two-tailed) within the 2 groups, as well as for the whole sample (musicians and nonmusicians). This procedure was applied because even though both groups were less proficient at categorizing the 2 stimuli of the reduced-spectrum condition, the musicians performed better on the task than the nonmusicians. These correlations reached significance, or at least evidenced a statistical trend, in the musician group ($r_{wnda\ PT-RIGHT}$ = 0.527, P = 0.078) and in the whole sample of subjects ($r_{wnka\ PT-LEFT}$ = 0.463, P = 0.020; $r_{wnda\ PT-LEFT}$ = 0.342, P = 0.094). Nevertheless, these correlations were not significant within the nonmusician group. In summary, these results primarily indicate that across the whole sample of subjects, the behavioral performance in the reduced-spectrum condition was positively related to the brain responses in the PT-LEFT. Furthermore, the statistical trend we found within the musician group between PT-RIGHT activity and number of correct /wnda/ responses, leads to suppose that the musicians relied on the functional capacity of the right PT when the task became more difficult. This is supported by the notion that the musicians tended to be better at recognizing /wnka/ than /wnda/ (please consider the behavioral results). The correlations between the behavioral and physiological data are shown in Figure 1.

Discussion

General Discussion

The present study was specifically designed to investigate whether musicianship fosters phonetic and fast temporal processing. Twelve professional musicians and 13 nonmusicians were measured by means of the fMRI technique during an auditory task based on the discrimination of CV syllables and their reduced-spectrum analogues. All acoustic stimuli were characterized by fast changing acoustic properties and required an estimation of the timing between the offset of activity evoked by the leading element and the onset of activity mediating the trailing element (Zaehle et al. 2008). Exactly these fine-graded temporal analyses were previously shown to recruit brain regions residing in the posterior supratemporal plane, with a strong bias to the left hemisphere (Jancke et al. 2002; Hickok and Poeppel 2007; Zaehle et al. 2008, 2009).

The goal of this study was to investigate whether the frequently observed functional and structural PT asymmetry in professional musicians favors the encoding and analysis of fast temporal changing cues in an acoustic environment. In line with previous results showing the involvement of the left posterior supratemporal plane in phonetic perception and in the processing of fast changing verbal and nonverbal cues (Zaehle et al. 2004, 2008), we suggest that the responsiveness of the left PT was enhanced in musicians as a function of musical training. Furthermore, we found evidence for the notion that the activity in the left PT was predictive for the performance in the reduced-spectrum categorization task in the entire sample of subjects. These results are of fundamental relevance since they enable a deeper understanding of the impact of professional musical training on phonetic processing. Furthermore, they provide a starting point for the predictive modeling of audition. Next, we shall place the results of the present study in a broader context by integrating the biographical, behavioral, and physiological data.

Behavioral Data

In the present study, subjects were instructed to assign the 4 auditory stimuli /ka/, /da/, /wnka/, and /wnda/ to the category /ka/ or /da/ by pressing the respective response button. While the performance in the natural speech condition did not differ between the 2 groups, the musicians behaviorally outperformed the nonmusicians in the reduced-spectrum condition. These results are not surprising and reflect ceiling effects, as well as enhanced auditory skills in musicians. In fact, the discrimination between /ka/ and /da/ can be considered an over-learned task in which all subjects, independent of musical expertise, reached almost a maximal hit rate. The enhanced auditory skills often reported in professional musicians (Parbery-Clark et al. 2009; Kraus and Chandrasekaran 2010) probably account for their better performance during the reduced-spectrum condition. This training-related enhanced auditory acuity was also reflected in the test for musical aptitudes (Gordon test). In this context, the musicians were consistently better, than the nonmusicians, at detecting tonal and rhythmical discrepancies in an acoustic musical environment. Since we did not reveal significant group differences in the test for general cognitive abilities (KAI test), we rule out any influence of this effect.

Planum Temporale

In line with our hypothesis, we revealed a main effect of group in the posterior part of the left PT. This effect was characterized by generally enhanced brain responses in the musicians irrespective of whether the stimuli consisted of CV syllables or of their reduced-spectrum analogues. The negative correlation between the total number of hours of instrumental training and PT activity points to a profound influence of musical training on auditory brain responses especially within the PT. Moreover, we found evidence that the responsiveness of the left PT in the whole sample of subjects was generally associated with their performance in the reduced-spectrum condition. Thus, these results indicate that professional musical training favors functional adaptations in the left PT and that the activity within this brain region is predictive of the behavioral performance on the more demanding reduced-spectrum condition, that is, irrespective of expertise. These results are comparable with those of previous work, which adopted CV syllables and reduced-spectrum analogues to investigate the functional properties of the left supratemporal plane during phonetic and temporal processing (Zaehle et al. 2004, 2008; Giraud et al. 2007; Hickok and Poeppel 2007). We provide first evidence for the notion that intensive musical training facilitates the categorization of speech-like temporal cues by enhancing the functional capacity of the left PT. Given that the perception of rapidly changing auditory patterns is a fundamental prerequisite for both speech and music processing, it is not surprising that intense music training has the potential to enhance the functional capacity of a brain region which is particularly sensitive for this kind of acoustic analyses (Zaehle et al. 2004, 2008; Hickok and Poeppel 2007). It should be kept in mind, however, that hearing CV syllables or structured noise stimuli is by no means the same as hearing music or instrumental tones. Given the fundamental difference between these 2 acoustic environments, our data rather bolster the notion that the left posterior supratemporal region supports a more general function in processing rapidly changing auditory cues (Meyer et al. 2005).

A particular previous study which combined the structural MRI and magnetoencephalography methods to investigate a large sample of professional musicians could show that those musicians who played an instrument producing short, sharp, or impulsive tones (e.g., drums, guitar, piano, trumpet, or flute) were associated with both larger gray matter volume and enhanced P50m activity in the left auditory cortex, which is sensitive to rapid temporal processing (Schneider et al. 2005). Otherwise, the musicians who played melodic instruments that produce rather sustained tones with characteristic changes in timbre (e.g., bassoon, saxophone, French horn, violoncello, or organ) exhibited a more dominant right auditory-related cortex, which is known to be sensitive to slower temporal and richer spectral processing. In our study, 8 out of 12 musicians played an instrument producing short, sharp, or impulsive tones (i.e., flute and piano). Therefore, taking into account, the results reported by Schneider et al. (2005), we may speculate whether the enhanced responsiveness of the left PT we revealed in the musician groups was mainly driven by the primary instrument played by the musicians rather than by musicianship in general. Further studies performed with large samples of subjects would be useful to clarify this issue. In other words, such studies would be beneficial to definitively answer the question whether the enhanced functional capacity of the left PT in professional musicians during speech processing may be associated with the acoustic properties of the primary instrument played.

Functional Hemispheric Dominance

In a further exploratory (nonindependent) ROI analysis, we addressed the question of whether intensive musical training has an influence on between-hemispheric differences. Statistical comparisons indicated that although the musician group's PT activity was lateralized to the left hemisphere, its right counterpart was more strongly responsive in the control subjects. This was the case even though the 2 groups did not differ in terms of right-sided activity. In order to investigate whether an inverse relationship between left premotor and temporal regions (as shown in Fig. 2) may account for this group difference in hemispheric asymmetry, we correlated the mean BETA values extracted from ROI PT-LEFT with those extracted from ROI PMC-LEFT in the whole sample of subjects. As expected, this post hoc analysis yielded a negative relationship between PT-LEFT and PMC-LEFT activity irrespective of group. The evidence that the 2 groups did not differ in terms of activity within the right PT (as revealed by the whole-head ANOVA), as well as the negative relationship we found between PT-LEFT and PMC-LEFT, leads us to suggest that the different hemispheric dominance between the 2 groups is not related to processing asymmetries per se. In fact, our data militate in favor of the view that these between-hemispheric differences are actually a by-product of reciprocal fronto-temporal modulation patterns.

An interplay between frontal and temporal brain regions has been previously proposed in the context of sensory gating mechanisms in healthy subjects (Weisser et al. 2001) and in patients with positive schizophrenic symptoms (Farzan et al. 2010). A reciprocal modulation between temporal and frontal regions was also reported in a single-case study of a patient with a destroyed auditory cortex in the left- and a frontal lesion in the right hemisphere (Griffiths et al. 2000). Since the right-sided auditory cortex disconnection from frontal areas has been associated with an inability of the patient to detect sound fine structure at the level of tens and hundreds of milliseconds (even though the patient showed a normal pure tone audiometry), it may be suggested that frontal areas are also involved in the processing of temporal auditory patterns. Even though the functional interplay between frontal and temporal regions that we observed is in line with these previous observations, further research is necessary to better comprehend how it varies as a function of musical expertise.

Premotor Cortex

As a second main result, we revealed a main group effect in the left PMC. Figure 2 shows that this effect was characterized by a significant reduced activity in the musicians compared with the nonmusicians. By considering the nature of the experimental task, our data speaks in favor of the view that this brain region is engaged in supporting phonological and temporal processing. Previously, a seminal model of cortical speech processing proposed that a left dorsal stream, which extends from posterior perisylvian regions to the frontal operculum, is strongly involved in auditory-to-motor transformations and in articulation processes. Thus, the left dorsal stream seems to be important for mapping sounds to articulation (Wise et al. 2001; Hickok and Poeppel 2007). Furthermore, there is some evidence that exactly this dorsal processing stream supports language learning via refined speech-motor coding (Scott and Wise 2004). Previous work has shown that the stronger the activity in the left premotor cortex, the better the categorization of phonetic patterns in both healthy (Wong et al. 2007; Dufor et al. 2009) and dyslexic (Dufor et al. 2009) subjects. Callan et al. (2003) have found additional evidence for the contribution of the left premotor cortex to phonetic processing after conducting an extensive perceptual identification training with native Japanese speakers who were learning the English /r-l/ phonetic contrast. These previous observations lead us to speculate that the left premotor cortex contributes to optimizing phoneme categorization via refined speech-motor coding mechanisms. Our data support this notion as the music experts in this study principally performed the task by engaging sensory functions; this is in contrast to the control subjects who reverted to auditory–articulatory mapping mechanisms. In fact, the differential activation patterns we found in the left premotor cortex, which are a function of musical training, are also consistent with anecdotal reports from the control subjects who credited vocal rehearsal as an important factor that aided their completion of the task, at least during the more difficult reduced-spectrum condition. Finally, one should consider that a variety of regions residing in the frontal lobe are involved in the control of executive functions, such as, working memory, attention, and inhibition. Since the phonetic categorization task applied in this experiment involved attention and working memory functions, we cannot completely exclude that some of these cognitive processes were differently engaged between the 2 groups.

Limitations

In the present study, the phonetic categorization task was conducted inside the scanner without collecting behavioral data from the same subjects in a completely silent environment outside the scanner. This means that we did not compare the

results between different acoustic environments because it was not a major issue of the study. However, we are aware of the fact that the fMRI scanner noise can be regarded as a kind of background noise, which increases the difficulty to identify the test stimuli. Thus, this stimulation condition is different to condition during which no additional background noise is present. Even though in the present experiment, the auditory stimuli were presented in silent periods of 1 s (TA = 2 s, TR = 3 s), we cannot completely exclude that the acoustically adverse conditions inside the scanner may have influenced the data in some directions. Furthermore, it is possible that the task we applied inside the scanner may have been influenced by the robustness of hearing of the subjects as a function of expertise (i.e., the ability to hear out subtle differences under adverse conditions) rather than measuring phonetic categorization per se. Finally, we want to emphasize that in the present study, we presented to the subjects 4 acoustic stimuli with a high repetition rate. This specific experimental setting complicates the generalization of our results to natural listening conditions where syllables or reduced-spectrum analogues are not presented at such a high rate.

Conclusions

In the present work, we addressed a novel research question, namely, whether long-term musical training may facilitate phonetic and temporal processing. In this context, our findings shed light on some important issues. First, our results suggest that musical training favors functional adaptations in the left PT, a brain region which was previously associated with spectrotemporal processing in general and phonetic processing in particular. Second, we were able to demonstrate that the activity of this specific brain region was predictive of the performance during the reduced-spectrum condition, independent of expertise. Third, we found support for the notion that hemispheric dominance of auditory-related processing was not related to processing asymmetries per se but rather associated with reciprocal fronto-temporal modulation patterns.

Funding

Notes

References

Annett M. 1970. A classification of hand preference by association analysis. Br J Psychol. 61:303–321.

Baumann S, Meyer M, Jancke L. 2008. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. J Cogn Neurosci. 20:2238–2249.

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.

Bever TG, Chiarello RJ. 2009. Cerebral dominance in musicians and nonmusicians. J Neuropsychiatry Clin Neurosci. 21:94–96.

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage. 19:113–124.

Caplan D, Gow D, Makris N. 1995. Analysis of lesions by MRI in stroke patients with acoustic-phonetic deficits. Neurology. 45:293–298.

Dufor O, Serniclaes W, Sprenger-Charolles L, Demonet JF. 2009. Left premotor cortex and allophonic speech perception in dyslexia: a PET study. Neuroimage. 46:241–248.

Efron R. 1963. Temporal perception, aphasia and deja vu. Brain. 86:403–424.

Farzan F, Barr MS, Levinson AJ, Chen R, Wong W, Fitzgerald PB, Daskalakis ZJ. 2010. Evidence for gamma inhibition deficits in the dorsolateral prefrontal cortex of patients with schizophrenia. Brain. 133:1505–1514.

Galaburda AM, Sanides F, Geschwind N. 1978. Human brain. Cytoarch-itectonic left-right asymmetries in the temporal speech region. Arch Neurol. 35:812–817.

Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 56:1127–1134.

Gordon EE. 1989. Manual for the advanced measures of music education. Chicago (IL): G.I.A. Publications, Inc.

Griffiths TD, Penhune V, Peretz I, Dean JL, Patterson RD, Green GGR. 2000. Frontal processing and auditory perception. Neuroreport. 11:919–922.

Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. Trends Neurosci. 25:348–353.

Heiervang E, Hugdahl K, Steinmetz H, Smievoll AI, Stevenson J, Lund A, Ersland L, Lundervold A. 2000. Planum temporale, planum parietale and dichotic listening in dyslexia. Neuropsychologia. 38:1704–1713.

Hickok G, Poeppel D. 2007. Opinion—the cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.

Hillenbrand J, Canter GJ, Smith BL. 1990. Perception of intraphonemic differences by phoneticians, musicians, and inexperienced listeners. J Acoust Soc Am. 88:655–662.

Hirshkowitz M, Earle J, Paley B. 1978. EEG alpha asymmetry in musicians and non-musicians: a study of hemispheric specialization. Neuropsychologia. 16:125–128.

Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, Schlaug G. 2009. The effects of musical training on structural brain development: a longitudinal study. Ann N Y Acad Sci. 1169:182–186.

Jancke L. 2009. The plastic human brain. Restor Neurol Neuros. 27:521–538.

Jancke L, Wustenberg T, Scheich H, Heinze HJ. 2002. Phonetic perception and the temporal cortex. Neuroimage. 15:733–746.

Keenan JP, Thangaraj V, Halpern AR, Schlaug G. 2001. Absolute pitch and planum temporale. Neuroimage. 14:1402–1408.

Kraus N, Chandrasekaran B. 2010. Music training for the development of auditory skills. Nat Rev Neurosci. 11:599–605.

Lisker L, Abramson AS. 1967. Some effects of context on voice onset time in English stops. Lang Speech. 10:1–28.

Lundberg I, Hoien T. 1989. Phonemic deficits; a core symptom of developmental dyslexia. Irish J Psychol. 10:579–592.

Magne C, Schon D, Besson M. 2006. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. J Cogn Neurosci. 18:199–211.

Marie C, Magne C, Besson M. 2011. Musicians and the metric structure of words. J Cogn Neurosci. 23:294–305.

Marques C, Moreno S, Castro SL, Besson M. 2007. Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. J Cogn Neurosci. 19:1453–1463.

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A. 2005. Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport. 16:1985–1989.

Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M. 2009. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cereb Cortex. 19:712-723.

Munte TF, Altenmuller E, Jancke L. 2002. The musician's brain as a model of neuroplasticity. Nat Rev Neurosci. 3:473-478.

Oechslin MS, Meyer M, Jäncke L. 2010. Absolute pitch-functional evidence of speech-relevant auditory acuity. Cer Cortex. 20:447-455.

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.

Pantev C, Roberts LE, Schulz M, Engelien A, Ross B. 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport. 12:169-174.

Parbery-Clark A, Skoe E, Lam C, Kraus N. 2009. Musician enhancement for speech-in-noise. Ear Hear. 30:653-661.

Rumsey JM. 1992. The biology of developmental dyslexia. JAMA. 268:912-915.

Schlaug G, Jancke L, Huang YX, Steinmetz H. 1995. In-vivo evidence of structural brain asymmetry in musicians. Science. 267:699-701.

Schlaug G, Norton A, Overy K, Winner E. 2005. Effects of music training on the child's brain and cognitive development. Ann N Y Acad Sci. 1060:219-230.

Schneider P, Sluming V, Roberts N, Bleeck S, Rupp A. 2005. Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. Ann N Y Acad Sci. 1060:387-394.

Schon D, Magne C, Besson M. 2004. The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology. 41:341-349.

Scott SK, Wise RJS. 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 92:13-45.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. Science. 270:303-304.

Shapleske J, Rossell SL, Woodruff PWR, David AS. 1999. The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. Brain Res Brain Res Rev. 29:26-49.

Swisher L, Hirsh IJ. 1972. Brain damage and ordering of two temporally successive stimuli. Neuropsychologia. 10:137-152.

Tallal P. 1980. Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang. 9:182-198.

Tallal P, Stark RE. 1981. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J Acoust Soc Am. 69:568-574.

Weisser R, Weisbrod M, Roehrig M, Rupp A, Schroeder J, Scherg M. 2001. Is frontal lobe involved in the generation of auditory evoked P50? Neuroreport. 12:3303-3307.

Westbury CF, Zatorre RJ, Evans AC. 1999. Quantifying variability in the planum temporale: a probability map. Cereb Cortex. 9:392-405.

Wilson SJ, Lusher D, Wan CY, Dudgeon P, Reutens DC. 2009. The neurocognitive components of pitch processing: insights from absolute pitch. Cereb Cortex. 19:724-732.

Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. 2001. Separate neural subsystems within 'Wernicke's area'. Brain. 124:83-95.

Wong PCM, Perrachione TK, Parrish TB. 2007. Neural characteristics of successful and less successful speech and word learning in adults. Hum Brain Mapp. 28:995-1006.

Zaehle T, Geiser E, Alter K, Jancke L, Meyer M. 2008. Segmental processing in the human auditory dorsal stream. Brain Res. 1220:179-190.

Zaehle T, Jancke L, Herrmann CS, Meyer M. 2009. Pre-attentive spectro-temporal feature processing in the human auditory system. Brain Topogr. 22:97-108.

Zaehle T, Wustenberg T, Meyer M, Jancke L. 2004. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur J Neurosci. 20:2447-2456.