Sadness is unique: neural processing of emotions in speech prosody in musicians and non-musicians

*Mona Park[1,2,3†], Evgeny Gutyrchik[1,2,3†], Lorenz Welker[2,4], Petra Carl[1,2], Ernst Pöppel[1,2,3,5,6], Yuliya Zaytseva[1,2,3,7,8], Thomas Meindl[9], Janusch Blautzik[9], Maximilian Reiser[9] and Yan Bao[1,2,3,5] ***

[1] Institute of Medical Psychology, Ludwig-Maximilians-Universität, Munich, Germany
[2] Human Science Center, Ludwig-Maximilians-Universität, Munich, Germany
[3] Parmenides Center for Art and Science, Pullach, Germany
[4] Institute of Musicology, Ludwig-Maximilians-Universität, Munich, Germany
[5] Department of Psychology and Key Laboratory of Machine Perception (MoE), Peking University, Beijing, China
[6] Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[7] Moscow Research Institute of Psychiatry, Moscow, Russia
[8] Prague Psychiatric Centre, 3rd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic
[9] Institute of Clinical Radiology, Ludwig-Maximilians-Universität, Munich, Germany

Musical training has been shown to have positive effects on several aspects of speech processing, however, the effects of musical training on the neural processing of speech prosody conveying distinct emotions are yet to be better understood. We used functional magnetic resonance imaging (fMRI) to investigate whether the neural responses to speech prosody conveying happiness, sadness, and fear differ between musicians and non-musicians. Differences in processing of emotional speech prosody between the two groups were only observed when sadness was expressed. Musicians showed increased activation in the middle frontal gyrus, the anterior medial prefrontal cortex, the posterior cingulate cortex and the retrosplenial cortex. Our results suggest an increased sensitivity of emotional processing in musicians with respect to sadness expressed in speech, possibly reflecting empathic processes.

Keywords: functional magnetic resonance imaging, language processing, prosody, basic emotions, musical training, temporal processing

INTRODUCTION

Musical training is associated with changes in cognitive and affective processing (Barrett et al., 2013). Musicians exhibit different expressions of musical emotion (Juslin and Laukka, 2003), and show stronger emotional experience in response to music (Blood and Zatorre, 2001). Musicians possess higher skills for the recognition of emotions expressed in music (e.g., Bhatara et al., 2011), and they differ from non-musicians in the processing of the sadness and fear conveyed in music (Park et al., 2014). However, the effects of musical training are not limited to the musical domain, and in particular certain aspects of speech processing have been shown to benefit from musical training (Thompson et al., 2004; Hyde et al., 2009; Lima and Castro, 2011; Patel, 2011, 2014). Musicians show improved performance in the encoding of speech sounds (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009a,b), in detecting speech in noise (Strait and Kraus, 2011a), in extracting rhythmical patterns in auditory sequences (Su and Pöppel, 2012), and in processing pitch in speech (Moreno and Besson, 2005; Magne et al., 2006; Besson et al., 2007; Musacchia et al., 2007; Chandrasekaran and Kraus, 2010). Moreover, musicians seem to possess advantages in processing speech prosody (Thompson et al., 2004; Lima and Castro, 2011) and extra-linguistic properties such as

the emotional content of speech (Nilsonne and Sundberg, 1985; Schön et al., 2004; Chartrand and Belin, 2006; Magne et al., 2006).

The advantages musicians exhibit in both music and speech processing have been explained by enhanced acoustic skills that musicians acquire through continuous training (Patel, 2003; Chartrand et al., 2008). The transfer effect from musical training to speech processing is assumed to be due to acoustic and rhythmic similarities between the two functional domains (Besson et al., 2011; Strait and Kraus, 2011b; Jäncke, 2012). Specifically in the communication of affect, music and speech share strong similarities, which has motivated the proposition of a shared "emotional protolanguage" of music and speech (Thompson et al., 2012). In order to express emotions, both music and speech make use of the same or similar acoustic elements such as timbre or pitch (Patel, 2003; Besson et al., 2007; Chartrand et al., 2008). Similarities between music and speech are also observed in the temporal domain as musical and verbal expressions use "temporal windows" of a few seconds within which musical motives or speech utterances are represented (Pöppel, 1989, 2009).

These strong associations between music and speech have also been observed on the neural level. Similarities have been found in brain networks active during processing of both music and language (Maess et al., 2001; Levitin and Menon, 2003; Brown

et al., 2004; Koelsch et al., 2004; Abrams et al., 2011; Zatorre and Schönwiesner, 2011; Escoffier et al., 2013; Frühholz et al., 2014), and it has been assumed that the communication of emotion in both domains may be based on the same neural systems associated with social cognition, including the medial superior frontal gyrus (SFG) and the anterior cingulate cortex (ACC; Escoffier et al., 2013). Similar to music, processing of emotional speech prosody has traditionally been associated with right hemispheric activation (Schirmer and Kotz, 2006; Wildgruber et al., 2006) but this view has recently been challenged by multi-phase models that assume several stages to be involved in emotional prosody processing recruiting both the left and the right hemisphere (e.g., Brück et al., 2011; Kotz and Paulmann, 2011; Witteman et al., 2012; Grandjean and Frühholz, 2013; Kotz et al., 2013). The network of brain areas involved in processing emotional prosody is assumed to mainly consist of the primary auditory cortices, the superior temporal gyrus (STG) and the inferior frontal gyrus, as well as subcortical regions including the amygdala and the hippocampus (Ethofer et al., 2012; Frühholz et al., 2012, 2014; Frühholz and Grandjean, 2013a; Kotz et al., 2013; Belyk and Brown, 2014).

Music training has been shown to alter the neural processing of music presumably based on functional and structural changes in the musician's brain (Hyde et al., 2009; Kraus and Chandrasekaran, 2010). Are the transfer effects of musical training on speech prosody processing also observable on the neural level? Research has been supportive of this view and it has been suggested (Besson et al., 2011; Strait and Kraus, 2011b; Patel, 2014) that intense and continuing musical training leads to structural and functional changes of the brain that advance cognitive processes and increases sensitivity to acoustic features in music processing (Besson et al., 2011; Strait and Kraus, 2011b) which may subsequently also improve speech and specifically prosody processing. A number of studies have described differences between musicians and non-musicians in speech and prosody processing on the neural level (see Wong et al., 2007; Strait et al., 2009a,b; Patel, 2014). However, these studies have investigated the advantages in musicians compared to non-musicians on the level of subcortical auditory processing (Kraus and Chandrasekaran, 2010). To our knowledge, no brain imaging study has to date explicitly investigated the effects of musical training on cortical activation patterns in response to emotions conveyed in speech prosody. In line with previous studies showing that individual differences, such as stable personality traits, and also acquired musical expertise (Park et al., 2013, 2014), alter the neural responses to musically conveyed emotions such as sadness and fear, we aimed at identifying a potential cross-modal effect of musical training on the neural processing of speech prosody conveying different emotions. We expected musical training to be associated with an enhanced competence of emotional recognition, and distinctive differences in neural responses to emotional speech prosody.

METHODS
PARTICIPANTS
Twenty four healthy volunteers participated in the study. Twelve were non-musicians (7 female, mean age = 19.00, SD = 0.60)

who had no previous musical training and did not play any instruments, and 12 were musicians (7 female, mean age = 20.25, SD = 1.76 years) who had received formal music training (mean years of training = 13.83, SD = 2.58 years) in a variety of musical instruments (stringed instruments: 29%, accordion: 24%, piano: 35%, flute 12%). All participants were right-handed. All of them were German native speakers. None of them had a record of neurological or psychiatric illness, head trauma or psychoactive substance abuse, or had contraindications for MRI (e.g., pacemaker implant, pregnancy). Musicians and non-musicians did not differ in general health (GHQ-12, German Version by Linden et al., 1996), (independent t-test: $t_{(21)} = 1.88$, $p > 0.05$) or general intelligence ($t_{(22)} = -0.65$, $p > 0.05$). There was no difference between the groups in mood, measured by the "Delighted-Terrible Scale" (Andrews and Withey, 1976), before (Mann-Whitney U-test: $z = 1.17$, $p > 0.05$), or after the experiment ($z = -0.06$, $p > 0.05$), also there was no differences within neither the non-musician (Mann-Whitney U-test: $z = 0.46$, $p > 0.05$) nor the musician ($z = 1.38$, $p > 0.05$) group before and after the experiment. The study was performed in accordance to the Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by the ethics committee of the Medical Faculty of the University of Munich. All participants signed an informed consent.

MATERIAL
Items from the Berlin Database of Emotional Speech (Burkhardt et al., 2005) were used for the study. The database includes pre-evaluated semantically neutral sentences spoken in German in six different emotional tones (happiness, sadness, fear, disgust, boredom, neutral) by five different male and female actors. For the present study, sentences spoken by both male and female voices with three different emotional intonations conveying happiness, sadness and fear were selected. Neutral sentences spoken with a neutral intonation served as the control condition. The stimuli set has been evaluated for correct identification rates and naturalness of expression (Burkhardt et al., 2005) and for the present study, only stimuli with high values for correct detection (>65%) and naturalness (>65%) were chosen. To provide comparable and relatively long duration times, several original recordings of a given emotional quality by the same speaker were combined to last about 21 s each.

EXPERIMENTAL PROCEDURE
During scanning, participants listened to the stimuli binaurally via pneumatic, noise attenuating and non-magnetic headphones. Sound level was individually adjusted to be comfortable, and light was dimmed to suppress further visual stimulation. The participants listened passively to the sentences and were asked to keep their eyes closed during the experiment.

During three measurement sessions (runs) three emotional qualities (happiness, sadness, fear) and a control condition (neutral) were presented twice (same sentences and same emotional intonation but spoken by a female and a male speaker respectively). In total, six iterations (trials) of each emotion were presented. The different conditions were presented under computer control in a pseudo-randomized order. To control for order

effects, two versions of stimuli sequences were created and participants were randomly assigned to either one of them. Each stimulation-interval was followed by a pause. After scanning, participants listened to the set of stimuli again and were asked to identify the expressed emotion after each sentence by selecting an emotion from a provided list (happiness, fear, anger, disgust, sadness, surprise, neutral) or by choosing an individual label.

IMAGE ACQUISITION AND fMRI DATA ANALYSES

The experimental set-up was similar to a previous study (Park et al., 2014). MRI was performed using a 3 T whole body system (Magnetom VERIO, Siemens, Erlangen, Germany) at the University Hospital of the LMU Munich. The scanner was equipped with a standard TIM head coil (12 elements) and the participant's head was securely but comfortably fastened by a foam cushions in order to minimize head movements. For acquiring the blood oxygen level dependent (BOLD) functional images, an T2*-weighted Echo-Planar Imaging (EPI) sequence was used with the following parameters: repetition time (TR) = 3000 ms, echo time (TE) = 30 ms, flip angle (FA) = 80°, number of slices = 28, slice thickness = 4 mm, inter-slice gap = 0.4 mm, interleaved acquisition, field of view (FOV) = 192 × 192 mm, matrix = 64 × 64, in-plane resolution = 3 × 3 mm. Functional images were obtained in axial orientation, covering the whole cerebrum and dorsal cerebellum. A total of 183 scans were conducted for each participant over all three runs. The functional measurement session lasted approximately 10 min in total. To provide an anatomical reference and to rule out structural abnormalities, a sagittal high-resolution 3D T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence was performed: TR = 2400 ms, TE = 3.06 ms, FA = 9°, number of slices = 160, FOV = 240 × 256 mm, spatial resolution = 1 mm.

Data were analyzed with SPM8 (Statistical Parametric Mapping[1]). The first five volumes were discarded due to T1 saturation effects. All functional images were realigned ("estimate and reslice"), co-registered ("estimate"; EPI template; Montreal Neurologic Institute, MNI), spatially normalized ("estimate and write") into standard stereotaxic space using standard SPM8 parameters, re-sliced to 2 × 2 × 2 mm voxels, and smoothed with an [8 8 8] mm full-width at half maximum (FWHM) Gaussian kernel. Each condition was modeled by a boxcar function convolved with the canonical hemodynamic response function. At the first level, t-tests were computed for each subject and for each condition vs. the baseline. The baseline of statistical parametric maps in our study is comprised of time periods not defined as conditions in the first-level model (i.e., happy, sad, fearful, and neutral prosody). The individual contrast images for each subject were used for the random-effects second level analysis (Full factorial design with one between-subjects (musicians, non-musicians) and one within-subjects (happy, sad, fearful, neutral prosody) factors). The statistical parametric maps were cluster-level thresholded (cluster-level thresholded at p(FDR) < 0.05, starting from p uncorrected < 0.01; cluster-size threshold = 300

----
[1]http://www.fil.ion.ucl.ac.uk/spm

voxels). Anatomical description was done referring to the AAL atlas (Automated Anatomical Labeling of Activations; Tzourio-Mazoyer et al., 2002).

RESULTS

IDENTIFICATION TASK

A main effect of emotion was revealed by a two-way analysis of variance (ANOVA) with *emotion* as within-subject variable and *group* as between-subject variable, $F_{(3,66)} = 9.454$, $P < 0.001$. Further paired t-tests showed that sadness conveyed by speech prosody was as easily identified as neutral voice (0.69 vs. 0.70 in correct identification rate, $P > 0.05$), while happy and fearful voices were equally difficult to be identified (0.48 vs. 0.58 in correct identification rate, $P > 0.05$), as significant differences were only observed between the two categories (i.e., sadness and neutral vs. happy and fear, $P < 0.05$). Importantly, no significant main effect of *group* was observed, $F_{(1,22)} = 1.546$, $p > 0.05$, and no significant two-way interaction was observed either, $F_{(3,66)} = 1.728$, $p > 0.05$. These results seemed to indicate that both musicians and non-musicians are equally capable to identify emotions conveyed in speech prosody, although both groups are better at recognizing sadness as compared to fearful and happy emotions.

SIMILARITIES BETWEEN GROUPS—CONJUNCTION ANALYSIS

Conjunction analysis (conj. null) for the three basic emotions (happiness, sadness, fear) vs. baseline revealed bilateral activation in the temporal cortex, specifically in middle temporal (BA 21) and STG (BA 22) (**Table 1**, **Figure 1**). Possibly due to scanner noises, no distinct increases of activation were found in primary auditory cortices in response to the three emotions.

DIFFERENCES BETWEEN GROUPS

We observed significant differences in neural activation between the groups in response to sentences with sad prosody. In response to sad prosody musicians showed a significant increase of activation in the frontal cortex (BA 10, BA 9, 46), ACC (BA 32), posterior cingulate (BA 23, 31) and retrosplenial cortex (BA 29, 30) (**Table 1**, **Figure 2**). We did not observe any differences in neural activation between musicians and non-musicians in response to happy or fearful prosody. No increases of activation for non-musicians relative to musicians in response to any of the emotions were found.

DISCUSSION

The present study revealed similarities and differences between musicians and non-musicians in processing of emotional speech prosody expressing happiness, sadness and fear.
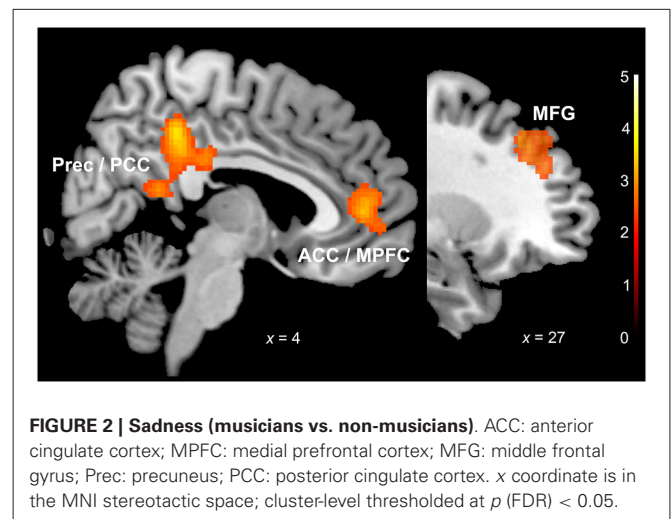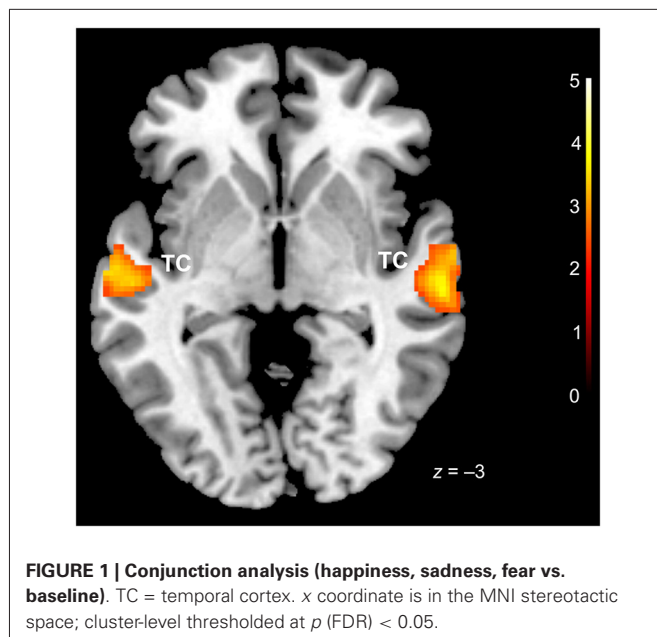
Conjunction analysis for fear, happiness and sadness revealed bilateral activations in temporal cortex, in the middle temporal gyrus (MTG) and the STG in both musicians and non-musicians. These areas are part of an auditory processing stream for categorizing auditory information (Hickok and Poeppel, 2007), including the identification and processing of linguistic and paralinguistic features of speech (e.g., Wildgruber et al., 2005; Schirmer and Kotz, 2006; Ethofer et al., 2012). The STG and the MTG crucially involved in processing emotional prosody

Table 1 | Neurofunctional correlates.

Brain region	Cluster	kE	Coordinates			Z-value
			x	y	z	
A. Conjunction (all emotions vs. baseline)						
R. superior temporal g., R. middle temporal g. (BA 21, 22)	1	548	66	−10	−6	3.72
L. middle temporal g., L. superior temporal g. (BA 21, 22)	2	359	−64	−16	−2	3.37
B. Sadness (musicians vs. non-musicians)						
R./L. cingulate g., middle part, R./L. precuneus, R./L. cingulate g., posterior part (BA 23, 31, 7, 29, 30)	1	1591	2	−40	40	3.72
R./L. cingulate g., anterior part, R. middle frontal g., R. superior frontal g., R. superior frontal g., medial part (BA 9, 10, 46, 32)	2	962	12	44	8	3.64

Note. kE = size in voxels (2 × 2 × 2 mm). R. = right, L. = left, g. = gyrus. The x, y and z coordinates are in the MNI stereotactic space.

FIGURE 1 | Conjunction analysis (happiness, sadness, fear vs. baseline). TC = temporal cortex. x coordinate is in the MNI stereotactic space; cluster-level thresholded at *p* (FDR) < 0.05.

FIGURE 2 | Sadness (musicians vs. non-musicians). ACC: anterior cingulate cortex; MPFC: medial prefrontal cortex; MFG: middle frontal gyrus; Prec: precuneus; PCC: posterior cingulate cortex. x coordinate is in the MNI stereotactic space; cluster-level thresholded at *p* (FDR) < 0.05.

(e.g., Mitchell et al., 2003; Leitman et al., 2010; Frühholz and Grandjean, 2013b; Grandjean and Frühholz, 2013) and reliably, the (right) STG is found in studies on emotional prosody processing (Brück et al., 2011; Ethofer et al., 2012; Frühholz et al., 2012; Frühholz and Grandjean, 2013a; Kotz et al., 2013; Belyk and Brown, 2014). It is assumed to play a major role in the early stages of prosody processing and has recently been referred to as the crucial part of the "emotional voice area" (Ethofer et al., 2012).

These common activations suggest that in musicians and non-musicians similar neural mechanisms are recruited for early stage processing of emotional vocal stimuli.

Apart from these similarities we also observed differences in neural responses to emotional speech prosody between the groups. Specifically, musicians showed enhanced activations in several brain areas when responding to sentences spoken with sad prosody, suggesting higher sensitivity in emotion processing.

Our observations will be discussed in the context of local neural activations and their assumed associations with subjective representations, being well aware of the conceptual problems when attributing high level cognitive processes to local neural modules or distributed neural networks (Bao and Pöppel, 2012).

We observed activation increases in the musician group in response to sad prosody in right frontal areas, in the middle and SFG (BA 10, BA 9, BA 46). Structural plasticity in right frontal regions has previously been associated with musical training (Hyde et al., 2009). Consistently, models on prosody processing agree in assuming the frontal cortex to play a crucial role in higher levels of prosody processing (see Witteman et al., 2012), specifically in the detection and judgment of emotional speech prosody (see Schirmer and Kotz, 2006). Specifically, the middle frontal gyrus has previously been found to be associated with processing of incongruity of in emotional prosody (Mitchell, 2013) and the detection of sad emotional tone (Buchanan et al., 2000). The stronger activations in right prefrontal areas may thus reflect processes related to the evaluation and categorization

of emotional prosody and it might also point to an enhanced sensitivity in the musician group specifically for the sad emotional content of the stimuli.

The increases in frontal activation for the group of musicians in response to sad speech prosody also included the an area comprising the medial part of the SFG and the ACC (BA 10, 32); areas that are both particularly associated with emotional processing, the appraisal and the regulation of emotions (Etkin et al., 2011), and also the induction of emotions (Amodio and Frith, 2006). The ACC is assumed to be part of a network specifically sensitive to monitoring of uncertainty and emotional saliency (Nomura et al., 2003; Cieslik et al., 2013) and the ACC and the medial prefrontal cortex have been specifically associated with the induction of sadness (Beauregard et al., 1998; Mayberg et al., 1999; Bush et al., 2000). Furthermore, the medial prefrontal cortex has been observed to be involved in emotional voice processing (Johnstone et al., 2006; Ethofer et al., 2012), and activation in the ACC has been found to play a special role in processing of emotional prosody (Bach et al., 2008; Belyk and Brown, 2014). We previously found increased activation in prefrontal regions in musicians in response to sadness in a study on musically conveyed emotions (Park et al., 2014) and Escoffier et al. (2013) found activations in the superior frontal cortex and the ACC during the processing of emotions that were expressed in music and through vocalization. The authors assumed that specific social processes might underlie emotion perception in both domains as both the superior frontal cortex and the ACC play a crucial role in mentalizing and other theory of mind (TOM) mechanisms (Escoffier et al., 2013). In fact, the medial prefrontal cortex and the ACC have consistently been associated with empathic processes and perspective taking (Amodio and Frith, 2006; Decety and Jackson, 2006; Etkin et al., 2011) and in particular the medial prefrontal cortex has been termed a "hub of a system mediating inferences about one's own and other individual's mental states" (Ochsner et al., 2004). The increased activations in the medial prefrontal cortex and the ACC in the group of musicians in response to sad sentences might thus suggest stronger emotional responses specifically related to the sad prosody of the stimuli. The increases of activation might furthermore point towards specific empathic processes related to the perceived sadness expressed in the stimuli (Harrison et al., 2007).

We also observed stronger activation in musicians in response to sad speech prosody in the posterior cingulate (PCC, BA 23, 31) and the retrosplenial cortex (BA 29, 30). The PCC and retrosplenial region have been associated with internally directed thought and episodic memory functions (Vann et al., 2009; Leech et al., 2012), and they are also involved in the "neural network correlates of consciousness", playing an important role in cognitive awareness, self-reflection (Vogt and Laureys, 2005) and control of arousal (Leech and Sharp, 2014). The PCC and retrosplenial region are also assumed to be involved in processing of the salience of emotional stimuli (Maddock, 1999) and the emotional content of external information (Cato et al., 2004), specifically of emotional words (Maddock et al., 2003). The increased activation we observed in the PCC and retrosplenial region in response to the sad prosody might, thus, reflect enhanced memory processes as well as increased assessment of

emotional saliency of the sad prosodic stimuli and monitoring of arousal.

Some of the areas in which we found activation increases for musicians in response to sad speech prosody can be considered parts of the default mode network (DMN, Raichle et al., 2001; Buckner et al., 2008), specifically the cortical midline structures ACC and PCC and the anterior medial regions of the prefrontal cortex. The DMN shows strong activity at rest and deactivation during tasks that call for external attention. The DMN as a functional system has been associated with processing of self (Zaytseva et al., 2014) and reflects introspective activities and stimulus-independent thought. Such "mentalizing" detaches from the present moment in which stimulus processing takes place (Pöppel and Bao, 2014). Furthermore, the DMN has been associated with induction of emotions, processing of affective saliency (Andrews-Hanna et al., 2010) and with social-emotional processing (Schilbach et al., 2008, 2012), such as attributing mental states to self and others (e.g., Mars et al., 2012).

It may be a puzzling result that the only significant differences between the groups were observed in the neural response to prosody expressing sadness but not in response to the other emotions. However, sadness is consistently found to be one of the emotions that are easiest to recognize (see Thompson et al., 2004). It is characterized by a particularly relevance to social loss (Panksepp, 2005) and may therefore be considered a highly salient and socially relevant signal. Furthermore, the expression of sadness in both music and speech prosody relies on similar acoustic features (Curtis and Bharucha, 2010), which musicians, due to their enhanced acoustic skills, may be able to extract more readily. In a previous study on musical emotions (see Park et al., 2014), we also found that musicians showed stronger neural activations to musical excerpts conveying negative emotions including sadness, and indicated stronger arousal in response to sad music. It was hypothesized that musicians may possibly be at an advantage to respond to the high social saliency of this emotion due to certain gains in social-emotional sensibility. In fact, the social functions and effects of music making have recently received increased attention (Koelsch, 2013) and listening to music has been shown to automatically engage TOM processes such as mental state attributions (Steinbeis and Koelsch, 2009), possibly implying that musicians because of their ongoing training may be particularly experienced in those specific aspects of social-emotional cognition. In fact, there is some empirical indication that musical training does indeed positively influence social emotional and communication development (Gerry et al., 2012) and that musical interventions effectively improve social skills (Gooding, 2011). Thus, a specific increase of social competence and social-emotional sensibility may be one cross-functional benefit of long-term musical training. Assuming these potentially enhanced social cognitive and empathic competences, musicians might thus be more responsive to the high social saliency of sadness in speech prosody. However, while several studies have reported advantages in recognition of emotional speech prosody due to musical training (Thompson et al., 2004; Lima and Castro, 2011), we only observed the difference between musicians and non-musicians in identifying sadness on the neural level, but we did not find any significant differences on the behavioral level.

This dissociation between neural responses and verbal reports to sadness supports the general concept to distinguish between the levels of explicit and implicit experience (Pöppel and Bao, 2011). The fact that the difference between the groups was only observed on the neural level suggests that for musicians sadness may be characterized by a unique implicit representation. The neural activations we observed in response to the sad prosody, in particular the activations in the MPFC and other parts of the DMN (Ochsner et al., 2004; Mitchell et al., 2005; Amodio and Frith, 2006), may possibly reflect these social-emotional mechanisms that crucially involve implicit introspective, i.e., self-referential, processes to infer the mental state of the speaker.

Finally, while a transfer effect of musical training to speech processing may mainly depend on acoustic and rhythmic similarities between music and speech (see Jäncke, 2012) temporal mechanisms might constitute another driving force for this cross-functional learning effect. Temporal mechanisms are of utmost importance in coordinating cognitive processes and can be considered to be an anthropological universal (Bao and Pöppel, 2012). Positive learning effects related to temporal training have been observed previously on the level of temporal order thresholds (Bao et al., 2013) of native speaker of the tonal language Chinese who show different thresholds compared to subjects from a non-tonal language environment. Furthermore, temporal mechanisms are crucial for conveying poetry (Turner and Pöppel, 1988) and they can be regarded basic to the expression and experience of music (Pöppel, 1989). Since neuro-imaging studies have shown music and language to rely on similar neural structures (Abrams et al., 2011) and considering the temporal similarities between music and speech it might be suspected that musical training also positively impacts temporal processing, and the observed effects thus may reflect enhanced temporal sensitivity as an effect of inter-modal transfer (Pöppel, 1989, 2009) which might also involve a higher competence to detect sadness in speech.

In conclusion, consistent with a previous study showing differences in emotion processing presumably due to musical training (Park et al., 2014), our study supports the notion that such training also alters the neural processing of distinct emotions conveyed in speech prosody. In particular, while musicians and non-musicians do not differ in their performance in recognizing sadness in speech, they process this particular emotion significantly differently on the neural level. Musicians show distinct increases of neural activations only in response to the sad prosody, possibly due to a higher affective saliency that the sentences spoken with sad intonation might possess. Our results imply that the cross-modal transfer effects of musical training go beyond auditory processing and explicit emotional recognition skills; we suggest that such training may also impact the empathic aspects in human communication.

ACKNOWLEDGMENTS

REFERENCES

Abrams, D., Bhatara, A., and Ryali, S. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* 21, 1507–1518. doi: 10.1093/cercor/bhq198

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Andrews, F., and Withey, S. (1976). *Social Indicators of Well-Being: Americans' Perceptions of Life Quality.* New York: Plenum Press.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron* 65, 550–562. doi: 10.1016/j.neuron.2010.02.005

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., and Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage* 42, 919–927. doi: 10.1016/j.neuroimage.2008.05.034

Bao, Y., and Pöppel, E. (2012). Anthropological universals and cultural specifics: conceptual and methodological challenges in cultural neuroscience. *Neurosci. Biobehav. Rev.* 36, 2143–2146. doi: 10.1016/j.neubiorev.2012.06.008

Bao, Y., Szymaszek, A., Wang, X., Oron, A., Pöppel, E., and Szelag, E. (2013). Temporal order perception of auditory stimuli is selectively modified by tonal and non-tonal language environments. *Cognition* 129, 579–585. doi: 10.1016/j.cognition.2013.08.019

Barrett, K. C., Ashley, R., Strait, D. L., and Kraus, N. (2013). Art and science: how musical training shapes the brain. *Front. Psychol.* 4:713. doi: 10.3389/fpsyg.2013.00713

Beauregard, M., Leroux, J. M., Bergman, S., Arzoumanian, Y., Beaudoin, G., Bourgouin, P., et al. (1998). The functional neuroanatomy of major depression: an fMRI study using an emotional activation paradigm. *Neuroreport* 9, 3253–3258. doi: 10.1097/00001756-199810050-00022

Belyk, M., and Brown, S. (2014). Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Soc. Cogn. Affect. Neurosci.* 9, 1395–1403. doi: 10.1093/scan/nst124

Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., and Levitin, D. J. (2011). Perception of emotional expression in musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 921–934. doi: 10.1037/a0021922

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U S A* 98, 11818–11823. doi: 10.1073/pnas.191355898

Brown, S., Martinez, M. J., and Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi: 10.1097/00001756-200409150-00008

Brück, C., Kreifelts, B., and Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.* 8, 383–403. doi: 10.1016/j.plrev.2011.10.002

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res. Cogn. Brain Res.* 9, 227–238. doi: 10.1016/s0926-6410(99)00060-9

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N Y Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Burkhardt, F., Paeschke, A., and Rolfes, M. (2005). A database of German emotional speech. *Interspeech* 5, 1517–1520.

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222. doi: 10.1016/s1364-6613(00)01483-2

Cato, M. A., Crosson, B., Gökçay, D., Soltysik, D., Wierenga, C., Gopinath, K., et al. (2004). Processing words with emotional connotation: an fMRI study of time course and laterality in rostral frontal and retrosplenial cortices. *J. Cogn. Neurosci.* 16, 167–177. doi: 10.1162/089892904322984481

Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246. doi: 10.1111/j.1469-8986.2009.00928.x

Chartrand, J.-P., and Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci. Lett.* 405, 164–167. doi: 10.1016/j.neulet.2006.06.053

Chartrand, J.-P., Peretz, I., and Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Res.* 1220, 191–198. doi: 10.1016/j.brainres.2008.01.014

Cieslik, E. C., Zilles, K., Caspers, S., Roski, C., Kellermann, T. S., Jakobs, O., et al. (2013). Is there "one" DLPFC in cognitive action control? Evidence for heterogeneity from co-activation-based parcellation. *Cereb. Cortex* 23, 2677–2689. doi: 10.1093/cercor/bhs256

Curtis, M. E., and Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion* 10, 335–348. doi: 10.1037/a0017928

Decety, J., and Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Curr. Dir. Psychol. Sci.* 15, 54–58. doi: 10.1111/j.0963-7214.2006.00406.x

Escoffier, N., Zhong, J., Schirmer, A., and Qiu, A. (2013). Emotional expressions in voice and music: same code, same effect? *Hum. Brain Mapp.* 34, 1796–1810. doi: 10.1002/hbm.22029

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., and Vuilleumier, P. (2012). Emotional voice areas: anatomic location, functional properties and structural connections revealed by combined fMRI/DTI. *Cereb. Cortex* 22, 191–200. doi: 10.1093/cercor/bhr113

Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93. doi: 10.1016/j.tics.2010.11.004

Frühholz, S., Ceravolo, L., and Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* 22, 1107–1117. doi: 10.1093/cercor/bhr184

Frühholz, S., and Grandjean, D. (2013a). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 37, 24–35. doi: 10.1016/j.neubiorev.2012.11.002

Frühholz, S., and Grandjean, D. (2013b). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* 37, 2847–2855. doi: 10.1016/j.neubiorev.2013.10.007

Frühholz, S., Trost, W., and Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Prog. Neurobiol.* 123, 1–17. doi: 10.1016/j.pneurobio.2014.09.003

Gerry, D., Unrau, A., and Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Dev. Sci.* 15, 398–407. doi: 10.1111/j.1467-7687.2012.01142.x

Gooding, L. (2011). The effect of a music therapy social skills training program on improving social competence in children and adolescents with social skills deficits. *J. Music Ther.* 48, 440–462. doi: 10.1093/jmt/48.4.440

Grandjean, D., and Frühholz, S. (2013). "An integrative model of brain processes for the decoding of emotional prosody," in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt and E. Zimmermann (Oxford, NY: OUP Oxford), 211–228.

Harrison, N. A., Wilson, C. E., and Critchley, H. D. (2007). Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion* 7, 724–729. doi: 10.1037/1528-3542.7.4.724

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025. doi: 10.1523/jneurosci.5118-08.2009

Jäncke, L. (2012). The relationship between music and language. *Front. Psychol.* 3:123. doi: 10.3389/fpsyg.2012.00123

Johnstone, T., van Reekum, C. M., Oakes, T. R., and Davidson, R. J. (2006). The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Soc. Cogn. Affect. Neurosci.* 1, 242–249. doi: 10.1093/scan/nsl027

Juslin, P., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Koelsch, S. (2013). From social contact to social cohesion–the 7 Cs. *Music Med.* 5, 204–209. doi: 10.1177/1943862113508588

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., and Haynes, J.-D. (2013). Predicting vocal emotion expressions from the human brain. *Hum. Brain Mapp.* 34, 1971–1981. doi: 10.1002/hbm.22041

Kotz, S. A., and Paulmann, S. (2011). Emotion, language and the brain. *Lang. Linguist. Compass* 5, 108–125. doi: 10.1111/j.1749-818X.2010.00267.x

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Leech, R., Braga, R., and Sharp, D. J. (2012). Echoes of the brain within the posterior cingulate cortex. *J. Neurosci.* 32, 215–222. doi: 10.1523/jneurosci.3689-11.2012

Leech, R., and Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain* 137, 12–32. doi: 10.1093/brain/awt162

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., et al. (2010). "It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Front. Hum. Neurosci.* 4:19. doi: 10.3389/fnhum.2010.00019

Levitin, D. J., and 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. doi: 10.1016/j.neuroimage.2003.08.016

Lima, C. F., and Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion* 11, 1021–1031. doi: 10.1037/a0024521

Linden, M., Maier, W., Achberger, M., Herr, R., Helmchen, H., and Benkert, O. (1996). Psychiatric diseases and their treatment in general practice in Germany. Results of a World Health Organization (WHO) study. *Nervenarzt* 67, 205–215.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends Neurosci.* 22, 310–316. doi: 10.1016/s0166-2236(98)01374-5

Maddock, R. J., Garrett, A. S., and Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Hum. Brain Mapp.* 18, 30–41. doi: 10.1002/hbm.10075

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502

Magne, C., Schön, D., and 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. doi: 10.1162/jocn.2006.18.2.199

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., and Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain". *Front. Hum. Neurosci.* 6:189. doi: 10.3389/fnhum.2012.00189

Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., et al. (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am. J. Psychiatry* 156, 675–682.

Mitchell, R. L. C. (2013). Further characterisation of the functional neuroanatomy associated with prosodic emotion decoding. *Cortex* 49, 1722–1732. doi: 10.1016/j.cortex.2012.07.010

Mitchell, J. P., Banaji, M. R., and Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.* 17, 1306–1315. doi: 10.1162/0898929055002418

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., and Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41, 1410–1421. doi: 10.1016/s0028-3932(03)00017-4

Moreno, S., and Besson, M. (2005). Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann. N Y Acad. Sci.* 1060, 93–97. doi: 10.1196/annals.1360.054

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U S A* 104, 15894–15898. doi: 10.1073/pnas.0701498104

Nilsonne, Å., and Sundberg, J. (1985). Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. *Music Percept.* 2, 507–516. doi: 10.2307/40285316

Nomura, M., Iidaka, T., Kakehi, K., Tsukiura, T., Hasegawa, T., Maeda, Y., et al. (2003). Frontal lobe networks for effective processing of ambiguously expressed

emotions in humans. *Neurosci. Lett.* 348, 113–116. doi: 10.1016/s0304-3940(03)00768-7

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., et al. (2004). Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.* 16, 1746–1772. doi: 10.1162/0898929042947829

Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* 14, 30–80. doi: 10.1016/j.concog.2004.10.004

Park, M., Gutyrchik, E., Bao, Y., Zaytseva, Y., Carl, P., Welker, L., et al. (2014). Differences between musicians and non-musicians in neuro-affective processing of sadness and fear expressed in music. *Neurosci. Lett.* 566, 120–124. doi: 10.1016/j.neulet.2014.02.041

Park, M., Hennig-Fast, K., Bao, Y., Carl, P., Pöppel, E., Welker, L., et al. (2013). Personality traits modulate neural responses to emotions expressed in music. *Brain Res.* 1523, 68–76. doi: 10.1016/j.brainres.2013.05.042

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011

Pöppel, E. (1989). The measurement of music and the cerebral clock: a new theory. *Leonardo* 22, 83–89. doi: 10.2307/1575145

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1887–1896. doi: 10.1098/rstb.2009.0015

Pöppel, E., and Bao, Y. (2011). "Three modes of knowledge as basis for intercultural cognition and communication: a theoretical perspective," in *Culture and Neural Frames of Cognition and Communication*, eds S. Han and E. Pöppel (Heidelberg: Springer), 215–231.

Pöppel, E., and Bao, Y. (2014). "Temporal windows as bridge from objective time to subjective time," in *Subjective Time: The Philosophy, Psychology and Neuroscience of Temporality*, eds V. Arstila and D. Lloyd (Cambridge: MIT Press), 241–261.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U S A* 98, 676–682. doi: 10.1073/pnas.98.2.676

Schilbach, L., Bzdok, D., Timmermans, B., Fox, P. T., Laird, A. R., Vogeley, K., et al. (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social and unconstrained cognition. *PLoS One* 7:e30920. doi: 10.1371/journal.pone.0030920

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., and Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Conscious. Cogn.* 17, 457–467. doi: 10.1016/j.concog.2008.03.013

Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x

Steinbeis, N., and Koelsch, S. (2009). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb. Cortex* 19, 619–623. doi: 10.1093/cercor/bhn110

Strait, D. L., and Kraus, N. (2011a). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fpsyg.2011.00113

Strait, D. L., and Kraus, N. (2011b). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Percept.* 29, 133–146. doi: 10.1525/mp.2011.29.2.133

Strait, D. L., Kraus, N., Skoe, E., and Ashley, R. (2009a). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668. doi: 10.1111/j.1460-9568.2009.06617.x

Strait, D. L., Kraus, N., Skoe, E., and Ashley, R. (2009b). Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Ann. N Y Acad. Sci.* 1169, 209–213. doi: 10.1111/j.1749-6632.2009.04864.x

Su, Y., and Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychol. Res.* 76, 373–382. doi: 10.1007/s00426-011-0346-3

Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U S A* 109, 19027–19032. doi: 10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46

Turner, F., and Pöppel, E. (1988). "Metered poetry, the brain and time," in *Beauty and the Brain—Biological Aspects of Aesthetics*, eds I. Rentschler, B. Herzberger and D. Epstein (Basel: Birkhäuser), 71–90.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802. doi: 10.1038/nrn2733

Vogt, B. A., and Laureys, S. (2005). Posterior cingulate, precuneal and retrosplenial cortices: cytology and components of the neural network correlates of consciousness. *Prog. Brain Res.* 150, 205–217. doi: 10.1016/s0079-6123(05)50015-3

Wildgruber, D., Ackermann, H., Kreifelts, B., and Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268. doi: 10.1016/S0079-6123(06)56013-3

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage* 24, 1233–1241. doi: 10.1016/j.neuroimage.2004.10.034

Witteman, J., Van Heuven, V. J., and Schiller, N. O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia* 50, 2752–2763. doi: 10.1016/j.neuropsychologia.2012.07.026

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Zatorre, R. J., and Schönwiesner, M. (2011). "Cortical speech and music processes revealed by functional neuroimaging," in *The Auditory Cortex*, eds J. A. Winer and C. E. Schreiner (Boston, MA: Springer), 657–677. doi: 10.1007/978-1-4419-0074-6

Zaytseva, Y., Gutyrchik, E., Bao, Y., Pöppel, E., Han, S., Northoff, G., et al. (2014). Self processing in the brain: a paradigmatic fMRI case study with a professional singer. *Brain Cogn.* 87, 104–108. doi: 10.1016/j.bandc.2014.03.012