

# Musicianship and Tone Language Experience Are Associated with Differential Changes in Brain Signal Variability

Stefanie Hutka<sup>1</sup>, Sarah M. Carpentier<sup>1</sup>, Gavin M. Bidelman<sup>2</sup>, Sylvain Moreno<sup>1</sup>, and Anthony R. McIntosh<sup>1</sup>

## Abstract

■ Musicianship has been associated with auditory processing benefits. It is unclear, however, whether pitch processing experience in nonmusical contexts, namely, speaking a tone language, has comparable associations with auditory processing. Studies comparing the auditory processing of musicians and tone language speakers have shown varying degrees of between-group similarity with regard to perceptual processing benefits and, particularly, nonlinguistic pitch processing. To test whether the auditory abilities honed by musicianship or speaking a tone language differentially impact the neural networks supporting nonlinguistic pitch processing (relative to timbral processing), we employed a novel application of brain signal variability (BSV) analysis. BSV is a metric of information processing capacity and holds great potential for understanding the neural underpinnings of experience-dependent plasticity. Here, we measured BSV in electroencephalograms of musicians,

tone language-speaking nonmusicians, and English-speaking nonmusicians (controls) during passive listening of music and speech sound contrasts. Although musicians showed greater BSV across the board, each group showed a unique spatio-temporal distribution in neural network engagement: Controls had greater BSV for speech than music; tone language-speaking nonmusicians showed the opposite effect; musicians showed similar BSV for both domains. Collectively, results suggest that musical and tone language pitch experience differentially affect auditory processing capacity within the cerebral cortex. However, information processing capacity is graded: More experience with pitch is associated with greater BSV when processing this cue. Higher BSV in musicians may suggest increased information integration within the brain networks subserving speech and music, which may be related to their well-documented advantages on a wide variety of speech-related tasks. ■

## INTRODUCTION

Psychophysiological evidence supports an association between music and speech such that experience in one domain is related to processing in the other (e.g., Bidelman, Gandour, & Krishnan, 2011; Koelsch, Maess, Gunter, & Friederici, 2001). Musicianship has been associated with benefits to auditory processing, such as enhanced spectral acuity for the perception of degraded speech (Zendel & Alain, 2012; Bidelman & Krishnan, 2010; Parbery-Clark, Skoe, Lam, & Kraus, 2009), lexical pitch judgments (e.g., Chandrasekaran, Krishnan, & Gandour, 2009; Schon, Magne, & Besson, 2004), and binaural sound processing (Parbery-Clark, Strait, Hittner, & Kraus, 2013). It is unclear, however, whether pitch processing experience in nonmusical contexts, namely, speaking a tone language, has comparable associations with auditory processing.

Tone languages, unlike other types of languages, use pitch phonemically (i.e., at the word level; e.g., Yip, 2002) to distinguish lexical meaning. Of all tone lan-

guages, Cantonese has one of the largest tonal inventories, comprising six tones—three of which are level, and three of which are contour (Rattanasone, Attina, Kasisopa, & Burnham 2013; Wong et al., 2012). These level pitch patterns are differentiable based on pitch height (Khouw & Ciocca, 2007; Gandour, 1981). The proximity of tones is approximately one semitone (i.e., a 6% difference in frequency, calculated from Peng, 2006), which is also the smallest distance found between pitches in music (Bidelman, Hutka, & Moreno, 2013). Note that this does not mean that Cantonese language experience is on par with musicians' auditory experience. Cantonese speakers have less pitch processing experience than musicians who have extensive experience with 12 level tones (i.e., the number of semitones in a scale) at several octaves, the processing of pitch contours as a result of the demands of musicianship, and the perception and production of complex melodies and harmonies. Furthermore, musicians' auditory demands include processing simultaneous tones (e.g., chords) and attending to the tone quality (i.e., timbre) of their instrument and other instruments around them. In comparison, tone language speakers have lesser auditory demands, typically

<sup>1</sup>Baycrest Centre for Geriatric Care, Toronto, Canada, <sup>2</sup>University of Memphis

processing a single, sequential stream of speech, without the same emphasis as musicians on tracking timbral cues. Because of the higher auditory demands faced by musicians (relative to tone language speakers), one might predict that benefits to auditory processing in musicians might be greater than to tone language speakers. Furthermore, one might predict that benefits to auditory processing in tone language speakers might be greater than in controls without musical training or tone language experience. However, studies comparing the auditory processing of musicians and tone language speakers have shown varying degrees of between-group similarity with regard to perceptual processing benefits and, particularly, nonlinguistic pitch processing.

Tone language experience (e.g., Mandarin: Bidelman et al., 2011; Cantonese: Bidelman, Hutka, et al., 2013) has been shown to similarly affect the neural encoding (Bidelman et al., 2011) and perception (Bidelman, Hutka, et al., 2013) of pitch. These studies imply that tone language experience may confer some benefits to spectral acuity that are comparable with those conferred by musicianship. However, behavioral studies have also revealed contradictory findings on tone language speakers' nonlinguistic pitch perception abilities, ranging from weak (Wong et al., 2012; Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011) to no enhancements (Schellenberg & Trehub, 2008; Bent, Bradlow, & Wright, 2006; Stargay & Downs, 1993). Neuroimaging studies have also been unclear in this regard (e.g., Bidelman et al., 2011). Our group has found that enhanced preattentive processing in brainstem and cortical auditory evoked potentials in musicians (Hutka, Bidelman, & Moreno, 2015; Bidelman et al., 2011), as well as tone language speakers (Bidelman et al., 2011) for nonlinguistic pitch. Yet, neural advantages do not necessarily coincide with behavioral benefits in nonlinguistic pitch discrimination and vice versa (Hutka et al., 2015).

In Bidelman et al. (2011), musicians and tone language speakers had similar, stronger brainstem representation of the defining pitches of musical sequences, as compared with controls (i.e., nonmusicians, non-tone language speakers). However, only musicians showed enhanced behavioral, musical pitch discrimination, relative to tone language speakers and controls. These findings suggest that enhanced processing at the brainstem level (i.e., preattentive stages of auditory processing) do not necessarily equate to perceptual, behavioral benefits. More puzzling is the absence of neural effects, given the presence of perceptual benefits, which was observed in Hutka et al. (2015). Hutka et al. (2015) measured the behavioral (i.e., difference limens) and automatic change detection response (i.e., MMN) to variations in pitch and timbre in musicians, Cantonese speakers, and non-musicians controls. Musicians and Cantonese speakers outperformed controls on the behavioral pitch discrimination task. Only musicians showed enhanced behavioral timbral processing, relative to tone language speakers

and controls. Parallel enhancements of behavioral spectral acuity in early auditory processing were observed in musicians only. That is, tone language users' advantages in pitch discrimination that were observed behaviorally were not reflected in early cortical MMN responses to pitch changes.

If both musicianship and speaking a tone language hone a common cue (pitch), why is this not reflected in their automatic cortical responses to pitch changes (Hutka et al., 2015)? It is possible that the mean activation over a cortical patch (i.e., the ERP, MMN measures) used in Hutka et al. (2015) may not adequately represent neural processes underlying the processing pitch. Musicians arguably have a greater range of experience with pitch (e.g., manipulating and producing complex melodies and harmonies) than do tone language speakers. By this logic, tone language speakers should not show neural responses to—nor behavioral benefits in—pitch discrimination that is comparable to that of musicians. However, because such behavioral benefits were observed in Cantonese speakers in Hutka et al. (2015), it is possible that there are unique neural circuitries associated with pitch processing in these individuals that were not adequately captured in ERP measures. More generally, this discrepancy between brain and behavior raises the question of the extent to which musicianship and tone language experience similarly shape the processing of nonlinguistic pitch processing, relative to controls.

To test whether automatic, nonlinguistic pitch processing is supported by common neural network activations in musicians and Cantonese speakers, two requirements emerge. First, one would need to apply a methodology that could detect nuanced effects in the brain signal that might underlie the differences between auditory processing for Cantonese speakers and musicians. Second, one would need to apply this methodology to the existing EEG data set from Hutka et al. (2015), which could then be directly compared with the ERP data from this study. Both of these requirements were met by the measurement of brain signal variability (BSV) in the continuous EEG signal as a metric of information integration (Heisz, Shedden, & McIntosh, 2012; Misić, Mills, Taylor, & McIntosh, 2010; Ghosh, Rho, McIntosh, Kötter, & Jirsa, 2008; McIntosh, Kovacevic, & Itier, 2008).

### **Brain Signal Variability**

BSV is the brain signal's transient, temporal fluctuations (McIntosh et al., 2013). There is strong evidence that BSV conveys important information about network dynamics (Deco, Jirsa, & McIntosh, 2011). The modeling of neural networks involves mapping an integration of information across widespread brain regions, via variations in correlated activity between areas across multiple timescales (Honey, Kötter, Breakspear, & Sporns, 2007; Jirsa & Kelso 2000; see Garrett et al., 2013, for a discussion). These transient changes result in fluctuations in the dynamics of the

corresponding brain signal (Garrett et al., 2013). The networks with more potential configurations produce a more variable response (Garrett et al., 2013). Therefore, BSV appears to represent the system's information processing capacity, in which higher variability reflects greater information integration (Garrett et al., 2013; McIntosh et al., 2008, 2013; Heisz et al., 2012; Misić et al., 2010; Ghosh et al., 2008). However, like any non-linear system, there is a theoretical "sweet spot" around which too little or too much variability may compromise information processing (Deco, Jirsa, & McIntosh, 2013).

As a metric of neural network dynamics, BSV provides valuable information about these dynamics that could not be obtained through the sole measurement of mean neural activity (e.g., using ERPs; Heisz et al., 2012; Vakorin, Misić, Krakovska, & McIntosh, 2011; McIntosh et al., 2008; see also Garrett, Kovacevic, McIntosh, & Grady, 2011; Ghosh et al., 2008). Averaging across trials (i.e., as in a traditional ERP analysis) removes the variability in each trial (see Hutka, Bidelman, & Moreno, 2013, for a discussion). This variability is not noise, instead providing information about network dynamics (Deco et al., 2011) and considers the entire neural network's activation and interactions (Hutka et al., 2013).

We posit that BSV might have great potential for studying the implicit neuroplasticity afforded by experience and learning. Studies have shown that the more information available to a listener about a given stimulus, the greater the BSV in response to that stimulus (Heisz et al., 2012; Misić et al., 2010). Variability should therefore increase as a function of learning, such that the more information one acquires for a stimulus, the greater information carried in the brain signal (Heisz et al., 2012). A study of Heisz et al. (2012) confirmed this expectation, showing greater BSV was associated with greater knowledge representation for certain faces—a result that was not reflected in the mean ERP amplitude in the same data set. Variability increased with face familiarity, suggesting that the perception of well-known stimuli engages a broader network of brain regions, manifesting in greater spatiotemporal changes in BSV (Heisz et al., 2012). These findings suggest that BSV is a useful metric of knowledge representation, capable of conveying information above and beyond what could be learned for mean neural activity. BSV can therefore be applied to the study experience-dependent plasticity and, in particular, pitch processing in musicians and tone language speakers (Hutka et al., 2013).

At present, we applied BSV analysis to the continuous EEG data set (i.e., not the ERP data) of Hutka et al. (2015), with the objective of studying the implicit impact of neuroplasticity afforded by experience and learning<sup>1</sup> in the auditory processing of nonlinguistic pitch in musicians, tone language speakers, and controls. Specifically, this design tested whether pitch processing, relative to another auditory cue (timbral processing), is supported by common neural network activations in musicians

and Cantonese speakers, relative to controls. Note that, throughout this manuscript, we are not seeking to make claims regarding fine-grained anatomical differences between groups or conditions. Instead, we sought to examine activation patterns of information integration during automatic processing of music (i.e., nonlinguistic pitch) and speech (linguistic timbre) in the three aforementioned groups.

## Hypotheses

If auditory experience via musicianship and tone language experience are associated with comparable information processing capacities supporting automatic music processing (i.e., pitch), then both groups would show greater BSV supporting auditory processing, as compared with controls (i.e., musicians = Cantonese speakers > controls). If the auditory expertise honed by musicianship and tone language are associated with different information processing capacities supporting automatic pitch processing, then we would predict different BSV between musicians and tone language speakers. This latter prediction would also manifest in unique spatio-temporal distributions for each group, as each group would be using a different brain network to support processing of pitch versus timbre.

## METHODS

Participants, stimuli, and EEG recording and preprocessing are the same as in Hutka et al. (2015).

## Participants

Sixty right-handed, young adult participants were recruited from the University of Toronto and Greater Toronto Area. All participants provided written, informed consent in compliance with an experimental protocol approved by the Baycrest Centre research ethics committee and were provided financial compensation for their time. English-speaking musicians (M;  $n = 21$ , 14 women) had at least 8 years of continuous training in Western classical music on their primary instrument ( $\mu \pm \sigma$ : 15.43  $\pm$  6.46 years), beginning formal music training at a mean age of 7.05 ( $\pm$ 3.32 years). English-speaking nonmusicians (NM;  $n = 21$ , 14 women) had  $\leq 3$  years of formal music training on any combination of instruments throughout their lifetime ( $\mu \pm \sigma$ : 0.81  $\pm$  1.40 years). Neither Ms nor NMs had experience with a tonal language of any kind. Native Cantonese-speaking participants (C;  $n = 18$ ; 11 women) also had minimal musical training throughout their lifetime (0.78  $\pm$  0.94 years). Importantly, NM and C did not differ in their minimal extent of music training,  $F(1, 37) = 0.007$ ,  $p = .935$ . C were born and raised in mainland China or Hong Kong, started formal instruction in

English at mean age of 10.27 ( $\pm 5.13$  years), and used Cantonese on a regular basis ( $>40\%$  of daily language use).

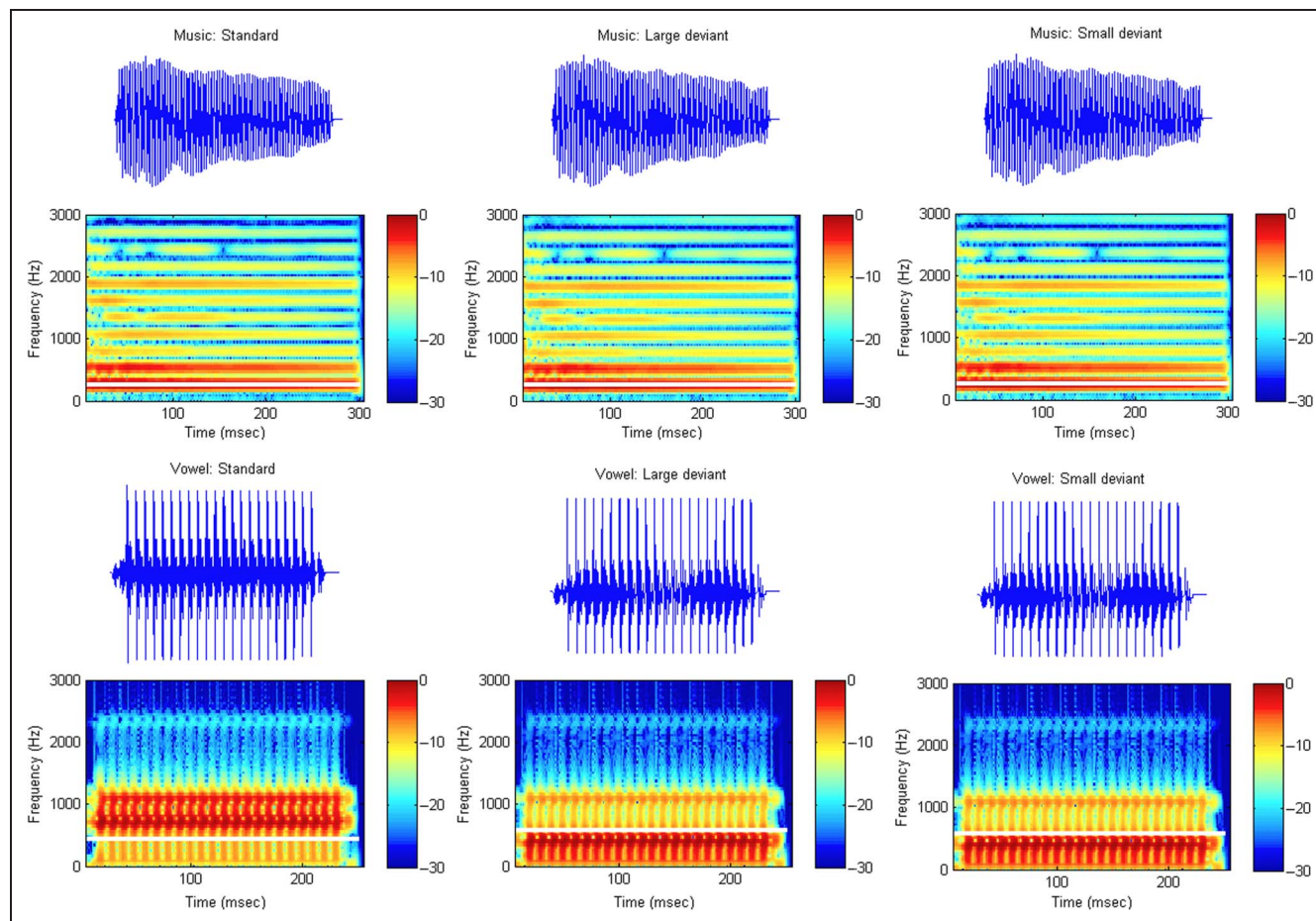
The three groups were closely matched in age (M:  $25.24 \pm 4.17$  years, C:  $24.17 \pm 4.12$  years, NM:  $23.38 \pm 4.07$  years;  $F(2, 57) = 1.075, p = .348$ ) and years of formal education (M:  $18.19 \pm 3.25$  years, C:  $16.94 \pm 2.46$  years; NM:  $16.67 \pm 2.76$  years;  $F(2, 57) = 1.670, p = .198$ ). All groups performed comparably on a measure of general fluid intelligence (Raven's Advanced Progressive Matrices; Raven, Raven, & Court, 1998) and non-verbal, short-term visuospatial memory (Corsi blocks; Corsi, 1972),  $p > .05$ .

### EEG Task Stimuli

EEGs were recorded using a passive, auditory oddball paradigm, consisting of two conditions—namely, music and speech sound contrasts presented in separate blocks (Figure 1). There were a total of 780 trials in each condition including 90 large deviants (12% of the trials) and 90 small deviants (12% of the trials). The notes (piano timbre) consisted of middle C (C4,  $F_0 = 261.6$  Hz), middle C mistuned by an increase of 0.5 semitones (large deviant;

269.3 Hz; 2.9% increase in frequency from standard), and middle C mistuned by an increase of 0.25 semitones (small deviant; 265.4 Hz; 1.4% increase in frequency from standard). Tone durations were 300 msec, including 5 msec of rise/fall time to reduce spectral splatter in the stimuli. Note that these changes were selected because previous behavioral research has demonstrated that both Cantonese speakers and musicians can distinguish between half-semitone changes in a given melody better than controls, whereas musicians outperform Cantonese speakers and controls when detecting a quarter-semitone change (Bidelman, Hutka, et al., 2013).

Speech stimuli consistent of three steady-state vowel sounds (Bidelman, Moreno, & Alain, 2013), namely, “oo” as in “book” [U], “aw” as in “pot” [a], and “uh” as in “but” [ʌ], as the standard, large deviant, and small deviant (on the border of categorical perception between the standard and large deviant; Bidelman, Moreno, et al., 2013), respectively. The duration of each vowel was 250 msec, including 10 msec of rise/fall. Note that the speech and note stimuli durations are different, as we were interested in maintaining natural acoustic features and presenting the sound as naturally as possible (Hutka et al., 2015). The



**Figure 1.** Spectrograms from Hutka et al. (2015), demonstrating the standard, large deviant, and small deviant stimuli for the music (top row) and speech (bottom row) conditions. White lines mark the frequencies of each tone’s fundamental frequency and each vowel’s first formant, respectively. Reprinted from Hutka et al. (2015), p. 55, copyright 2016, with permission from Elsevier.

sound onset asynchrony was 1000 msec in both conditions so that the stimulus repetition rates (and thus, neural adaptation effects) were comparable for both speech and music EEG recordings.

The standard vowel had a first formant (F1) of 430 Hz, the large deviant 730 Hz (41.1% increase in frequency from standard), and the small deviant 585 Hz (26.5% increase in frequency from standard). Speech tokens contained identical voice fundamental (F0), second (F2), and third (F3) formant frequencies (F0: 100, F2: 1090, and F3: 2350 Hz, respectively), chosen to match prototypical productions from a male speaker (Peterson & Barney 1952). The magnitude of F1 change between the standard and each speech deviant was chosen to parallel the magnitude of change in the music standard and deviants. However, it is notable that a greater magnitude of change was required to detect the standard large deviant and standard small deviant changes for F1 than F0. This difference was informed by past findings showing that participants require a larger percent change between two vowel sounds (i.e., F1) to detect a difference, as compared with between two pitches (i.e., F0; Bidelman & Krishnan, 2010). Pilot testing was used at present to determine the specific F0 and F1 standard deviant changes that musicians and nonmusicians could reliably detect.

### EEG Recording and Preprocessing

EEG was recorded using a 76-channel ActiveTwo amplifier system (Biosemi, Amsterdam, The Netherlands) with electrodes placed around the scalp according to standard 10–20 locations (Oostenveld & Praamstra, 2001). Continuous EEG recordings were sampled at 512 Hz and band-pass filtered online between 0.01 and 50 Hz. Source estimation was performed on the EEG data at 72 ROIs<sup>2</sup> defined in Talairach space (Diaconescu, Alain, & McIntosh, 2011) using sLORETA (Pascual-Marqui, 2002), as implemented in Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). Source reconstruction was constrained to the cortical mantle of the standardized brain template MNI/Colin27 defined by the Montreal Neurological Institute in Brainstorm. Current density for one source orientation ( $X$  component) was mapped at 72 brain ROIs, adapting the regional map coarse parcellation scheme of the cerebral cortex developed in Kötter and Wanke (2005). Multiscale entropy (MSE) was calculated on the source waveform at each ROI for each participant.

### Multiscale Entropy

To characterize BSV, MSE (Costa, Goldberger, & Peng, 2002, 2005) was measured, as it indexes sample entropy across multiple timescales. MSE quantifies sample entropy (Richman & Moorman, 2000) at multiple timescales (Costa et al., 2002, 2005). We calculated MSE in two steps using the algorithm available at [www.physionet.org/physiotools/mse](http://www.physionet.org/physiotools/mse) (Goldberger et al., 2000). First, the EEG signal was

progressively down-sampled into multiple coarse-grained timescales where, for scale  $t$ , the time series is constructed by averaging the data points with nonoverlapping windows of length  $t$ . Each element of a coarse-grained time series,  $y_j^{(\tau)}$ , is calculated according to Equation 1:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{N}{\tau} \quad (1)$$

The number of scales is determined by a function of the number of data points in the signal and the data in this study supported 12 timescales [sampling rate (512 Hz)  $\times$  epoch (1200 msec)/50 time points per epoch = maximum of 12 scales]. To convert timescale into milliseconds, the timescale was divided by the EEG sampling rate (512 Hz).

Second, the algorithm calculates the sample entropy ( $S_E$ ) for each coarse-grained time series ( $y_j^{(\tau)}$ ; Equation 2):

$$S_E(m, r, N) = \ln \frac{\sum_{i=1}^{N-m} n'_i m}{\sum_{i=1}^{N-m} n'_i m + 1} \quad (2)$$

Sample entropy quantifies the predictability of a time series by calculating the conditional probability that any two sequences of  $m$  consecutive data points that are similar to each other within a certain criterion ( $r$ ) will remain similar at the next point ( $m + 1$ ) in the data set ( $N$ ), where  $N$  is the length of the time series (Richman & Moorman, 2000). In this study, MSE was calculated with pattern length<sup>3</sup> set to  $m = 5$ , and the similarity criterion<sup>4</sup> was set to  $r = 1$ .

MSE estimates were obtained for each participant as the mean across single-trial entropy measures for each timescale.

### Spectral Analysis

Power spectral density (PSD) was also measured for all trials. This spectral analysis was conducted because previous studies suggested that changes in MSE tend to closely follow changes in spectral power, while providing unique information about the data (Misic et al., 2010; Lippe, Kovacevic, & McIntosh, 2009; McIntosh et al., 2008; Gudmundsson, Runarsson, Sigurdsson, Eiriksdottir, & Johnsen, 2007). Therefore, changes in sample entropy across sources and temporal scales were examined, as well as at changes in PSD across sources and frequency bands.

Single-trial power spectra were computed using the fast Fourier transform. To capture the relative contribution from each frequency band, all time series were first normalized to mean 0 and  $SD$  1. Given a sampling rate of 512 Hz and 614 data points per trial, the effective frequency resolution was 0.834 Hz. Hence, all spectral analyses were constrained to a bandwidth of 0.834–50 Hz.

## Statistical Analysis

### *Task Partial Least Squares Analysis*

Task partial least squares analysis (PLS) is a multivariate statistical technique that was used to assess between- and within-subject changes in MSE during listening (McIntosh & Lobaugh, 2004; McIntosh, Bookstein, Haxby, & Grady, 1996). Similar to multivariate techniques, such as canonical correlation analysis, PLS operates on the entire data structure at once, extracting the patterns of maximal covariance between two mean-centered data matrices, in the present case either group membership or condition (i.e., task design), and MSE measures (McIntosh et al., 2013). The analysis was done to emphasize two aspects of the experiment: (1) Between-group, which emphasizes main effects by centering group means to the overall grand mean, and (2) Between-condition, which identifies potential interactions by mean-centering each group to its own grand mean, which eliminates the between-group effects.

The PLS model is constructed with a singular value decomposition applied to the mean-centered MSE or PSD matrices. Singular value decomposition identified the strongest group and/or condition differences and the corresponding scalp topography, producing a set of orthogonal latent variables (LVs), with descending order of magnitude of accounted-for covariance. Each LV consists of (1) a pattern of design scores, (2) a singular image showing the distribution across brain regions and sampling scales, and (3) a singular value representing the covariance between the design scores and the singular image (McIntosh & Lobaugh, 2004; McIntosh et al., 1996). Statistical assessment in PLS consists of two steps. First, the overall significance of each LV that related the two data matrices was assessed with permutation testing (Good, 2000), which generates an estimated null distribution of the data. An LV was considered significant if the observed pattern (i.e., its singular value) was present less than 5% of the time in random permutations (i.e.,  $p < .05$ ). The dot product of an individual subject's raw MSE data and the singular image from the LV produces a brain score. The brain score is similar to a factor score in factor analysis that indicates how strongly an individual participant expresses the patterns on the LV. Analysis of brain scores allowed us to estimate 95% confidence intervals for the mean effects in each group and task condition.

Second, the reliability of the scalp topographies was determined using bootstrap resampling. This bootstrap resampling estimated standard error confidence intervals around the individual singular vector weights in each LV, assessing the relative contribution of particular locations and timescales and the stability of the relation with either group or condition (Efron & Tibshirani, 1986). For scalp topographies, the singular vector weights for each channel were divided by the bootstrap estimated standard error, giving bootstrap ratios. A bootstrap ratio is similar to a  $z$  score if the distribution of singular vector weights is

Gaussian, but are best interpreted as approximating a confidence interval (McIntosh et al., 2013). Brain regions with a singular vector weight over standard error ratio  $>3.0$  correspond to a 99% confidence interval and were considered to be reliable (Sampson, Streissguth, Barr, & Bookstein, 1989).

The large and small deviant conditions were combined into a single condition for all analyses, as preliminary analysis showed there were no differences in MSE or PSD between these conditions.

## RESULTS

All groups and conditions were entered into the task PLS analysis. Figures 2, 4, and 5 show both MSE and spectral data.

### **Between-group Comparisons**

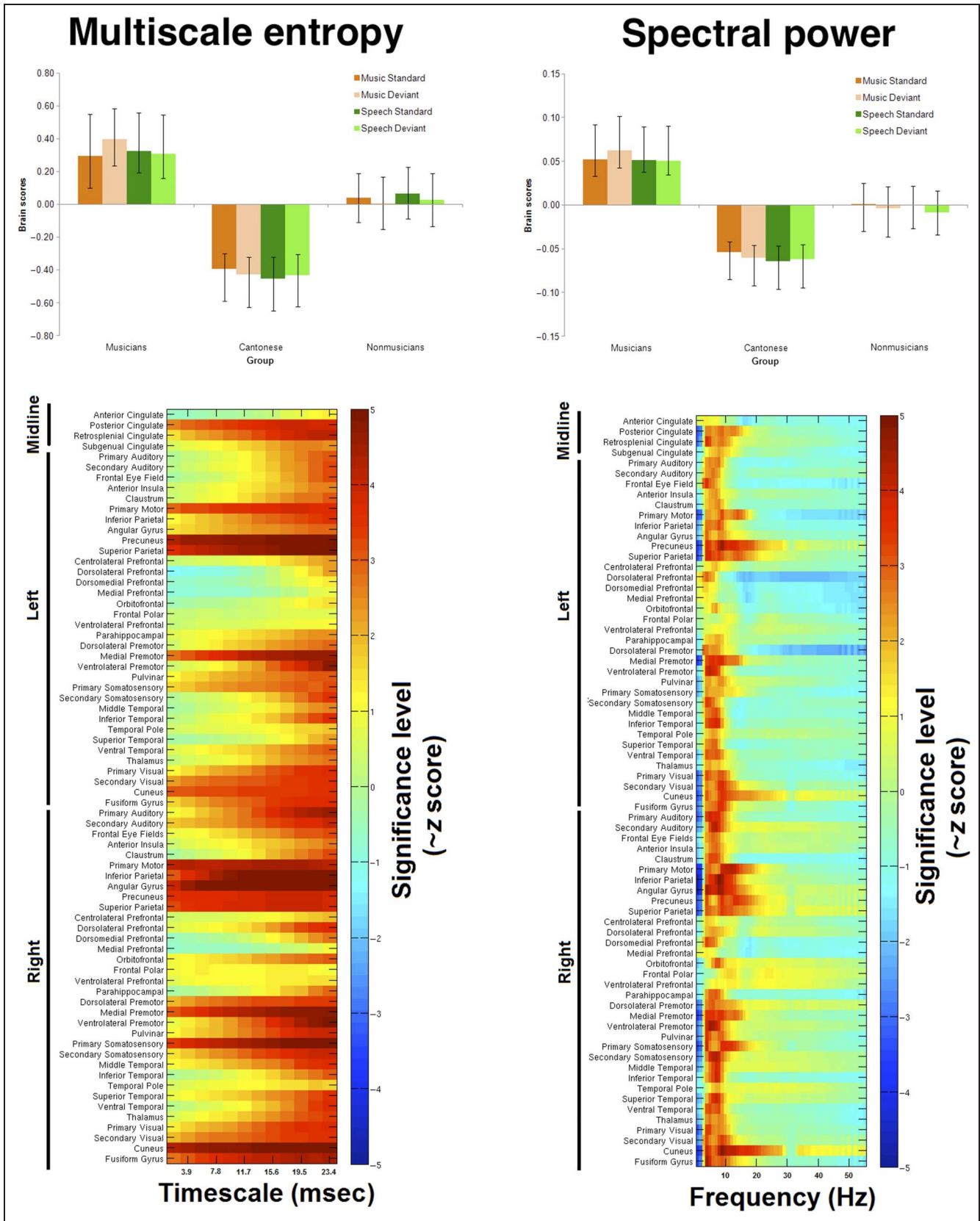
When comparing groups across all conditions (Figure 2; Figure 3, showing sample entropy curves for each timescale, across all conditions), the first LV (LV1) of the MSE analysis captured greater sample entropy in the musician group as compared with the Cantonese group (LV1,  $p = .004$ , singular value = 1.0856 corresponding to 43.82% of the covariance). This difference was robustly expressed across both fine and coarse timescales across all neural ROIs, particularly in the right hemisphere. The largest effects were seen across all timescales (particularly, in coarse scales) in the right inferior parietal, angular gyrus, and primary somatosensory area; medial posterior cingulate; and bilateral primary motor, medial premotor, precuneus, cuneus, and superior parietal area.

LV1 of the spectral analysis captured differences in the musician group as compared with the control and Cantonese groups (LV1,  $p = .012$ , singular value = 0.1626, corresponding to 37.01% of the covariance). This difference was robustly expressed across frequencies that were lower than 20 Hz (primarily theta/alpha band: 4–12 Hz) in a number of brain regions similar or identical to those observed in the MSE results.

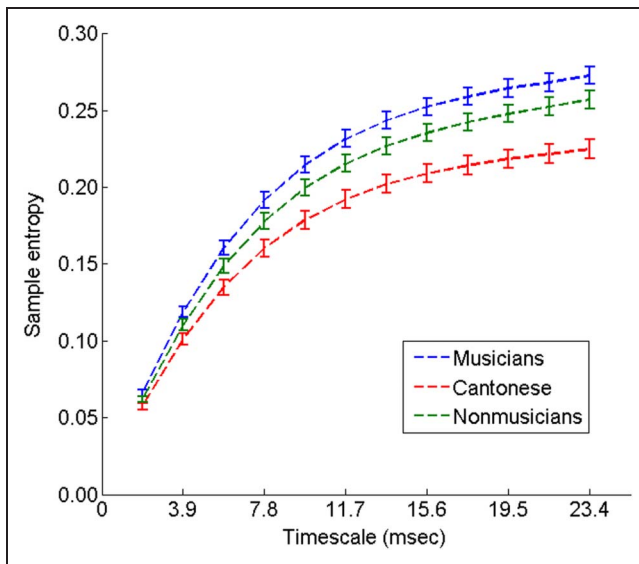
Collectively, PLS analyses revealed that each group could be distinguished based on the variability (MSE) and spectral details of their EEG (particularly in the right hemisphere), when listening to speech and music stimuli. Furthermore, in the areas in which these contrasts were robustly expressed (e.g., right angular gyrus; Figure 3), musicians had the greatest sample entropy across all conditions; Cantonese speakers had the lowest sample entropy; nonmusicians were in-between these two groups.

### **Between-condition Comparisons**

LV1 for the MSE analysis (Figure 4) captured differences in sample entropy between the music and speech conditions for nonmusicians ( $p = .002$ , singular value = 0.2518,



**Figure 2.** First LV (LV1), between-group comparison: Contrasting the EEG response to the music and speech conditions across measures of MSE (left) and spectral power (right). The bar graphs (with standard error bars) depict brain scores that were significantly expressed across the entire data set as determined by permutation at 95% confidence intervals. These results are consistent with a main effect of group. The image plot highlights the brain regions and timescale or frequency at which a given contrast was most stable; values represent  $\sim z$  scores, and negative values denote significance for the inverse condition effect.



**Figure 3.** MSE curves for all groups, averaged across all conditions, at the right angular gyrus. Musicians have the highest sample entropy of all groups, followed by nonmusicians and Cantonese speakers.

corresponding to 22.85% of the covariance). These differences were robustly expressed at fine timescales in several left hemisphere areas, namely, the anterior insula, centrolateral and dorsomedial pFC, frontal polar area, and secondary visual areas. Specifically, greater information integration supporting speech processing, as compared with music processing, was observed in these left hemisphere regions. Differences were also robustly expressed in the right primary and secondary visual areas and the cuneus. Namely, greater information processing capacity supporting music processing, rather than speech processing, was observed in these right hemisphere regions.

Similarly, LV1 of the spectral analysis captured spectral differences between the music and speech conditions for nonmusicians ( $p < .001$ , singular value = 0.0533, corresponding to 25.28% of the covariance). Processing of music, as compared with speech, was robustly expressed at frequencies below 10 Hz (e.g., theta, 4–7 Hz for the music condition) in multiple left hemisphere regions, namely, the left anterior insula, claustrum, centrolateral and dorsomedial pFC, frontal polar, parahippocampal cortex, thalamus, and dorsolateral and ventrolateral premotor cortex. These differences were also expressed in the midline posterior cingulate cortex, and the right cuneus, thalamus, and ventrolateral pFC. Processing of speech, as compared with music, was robustly expressed in frequencies above 12 Hz (e.g., beta, 12–18 Hz; gamma, 25–70 Hz for the speech stimuli) in multiple left hemisphere areas (left anterior insula, centrolateral and dorsomedial pFC, OFC, frontal polar, and dorsolateral premotor cortex), and the right primary motor area, precuneus, and dorsolateral pFC.

LV2 for the MSE analysis (Figure 5) captured differences in sample entropy between the music and speech conditions for Cantonese speakers ( $p = .052$ , singular

value = 0.2029, corresponding to 18.41% of the covariance). Specifically, greater information processing capacity for music processing, rather than speech processing, was robustly expressed in the midline posterior cingulate and retrosplenial cingulate cortex at fine timescales and the primary visual area at coarse timescales. Greater information processing capacity for speech processing, rather than music processing, was expressed in the left medial premotor cortex and right medial premotor cortex at coarse timescales.

Similarly, LV2 of the spectral analysis captured differences between the music and speech conditions for Cantonese speakers ( $p = .036$ , singular value = 0.0382, corresponding to 18.12% of the covariance). The processing of speech, as compared with music, was robustly expressed at frequencies below 10 Hz (e.g., theta, 4–7 Hz) in the bilateral medial premotor cortex. The processing of the music condition, as compared with speech, was robustly expressed in low-frequency activity (e.g., theta, 4–7 Hz) in the left parahippocampal cortex, and right anterior insula, ventral temporal cortex, and fusiform gyrus. Processing of music was also robustly expressed at frequencies above 12 Hz (e.g., beta, 12–18 Hz; gamma, 25–70 Hz), in the midline posterior and retrosplenial cingulate cortex, left superior parietal cortex, and bilateral primary and secondary visual areas.

Interestingly, a third LV (LV3), contrasting music and speech conditions for the musician group, was not significant (MSE:  $p = .256$ ; spectral analysis:  $p = .210$ ). Although it is possible that this effect would become significant at a larger sample size, the bootstrap-estimated standard errors were small, suggesting that this lack of an effect was robust (i.e., a stable-zero estimate; see McIntosh & Lobaugh 2004). The fact that we failed to detect a difference between musicians' processing of music or speech stimuli suggests that this group used a similar neural architecture to process acoustic information, regardless of the stimulus domain (i.e., music  $\approx$  speech).

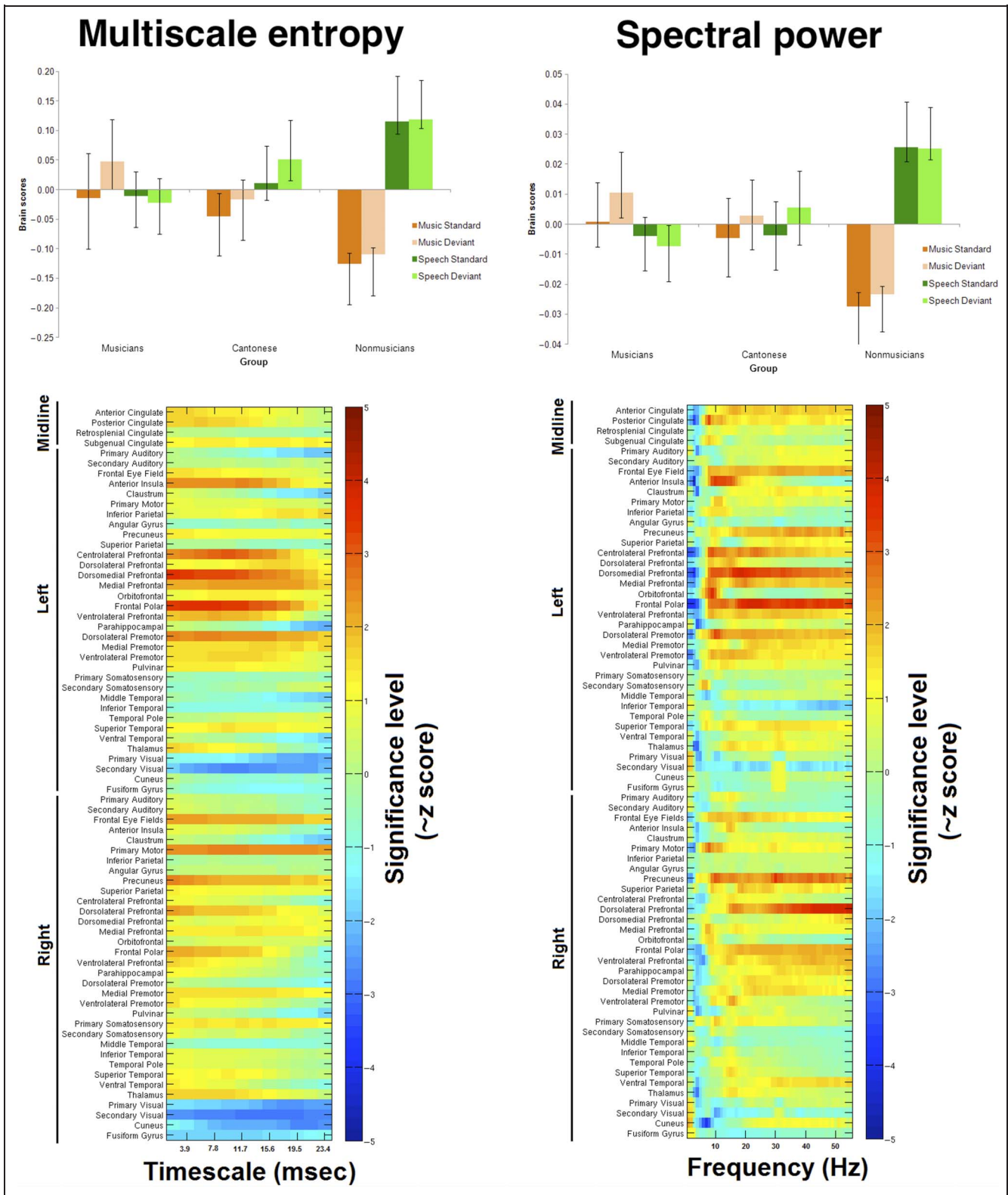
Collectively, the between-condition analyses revealed that each group processed the distinction between music and speech using a unique spatiotemporal network. LV1 showed that nonmusicians had greater sample entropy and higher-frequency activity for speech than music at several left hemisphere areas. LV2 showed that Cantonese speakers had greater sample entropy, for music than speech, particularly in midline regions. The spectral analyses revealed that this contrast was also expressed across multiple frequency bands. LV3, which was not significant, suggested that musicians used similar neural networks to support the processing of both music and speech stimuli.

## DISCUSSION

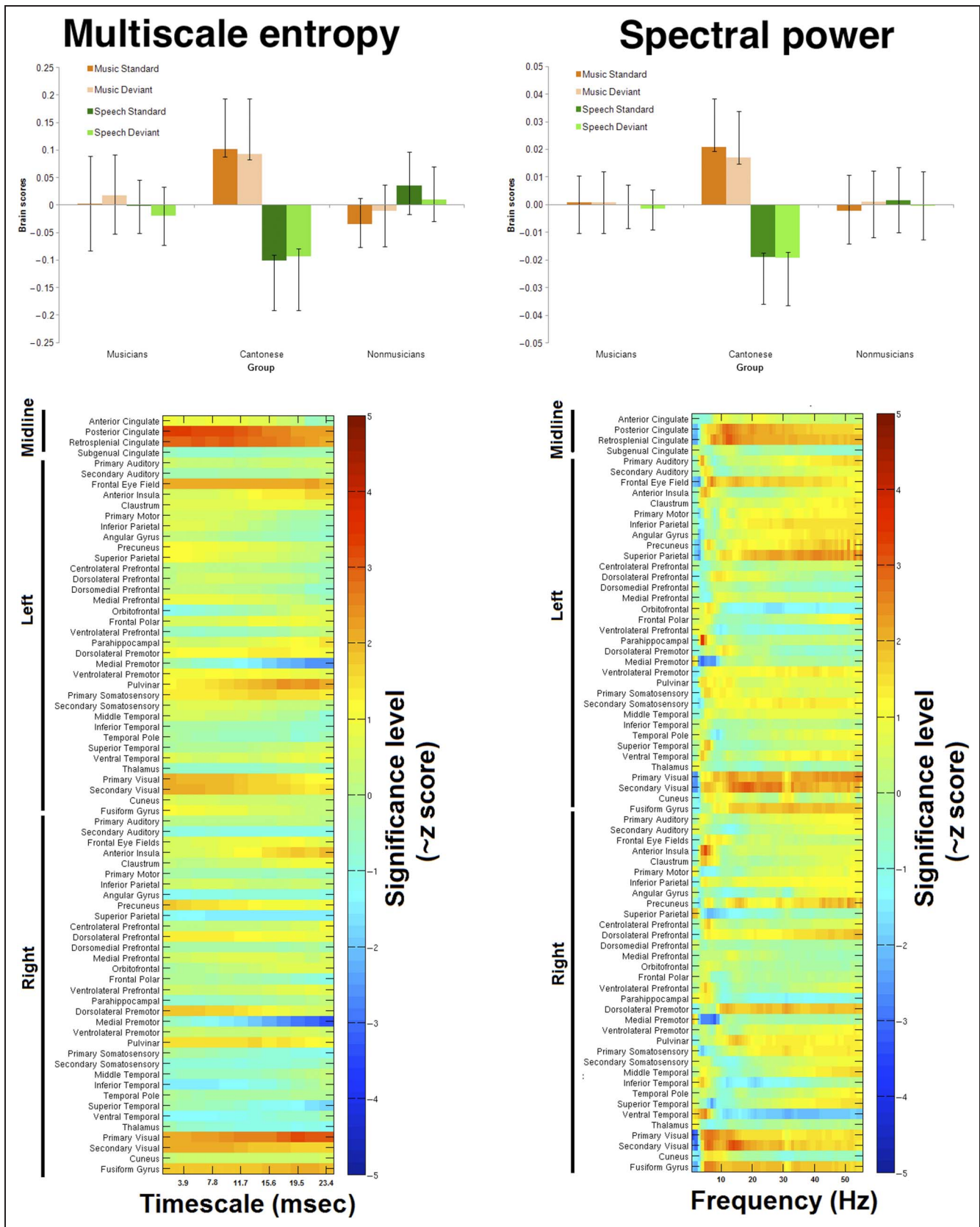
### MSE Data

By examining sample entropy between groups, we sought to test if musicianship and tone language (Cantonese)





**Figure 4.** First LV (LV1), between-condition comparison: Contrasting the EEG response to the music and speech conditions across measures of MSE (left) and spectral power (right) for nonmusicians. The bar graphs (with standard error bars) depict brain scores that were significantly expressed across the entire data set as determined by permutation tests at 95% confidence intervals. These results are consistent with an interaction effect. The image plot highlights the brain regions and timescale or frequency at which a given contrast was most stable; values represent  $\sim z$  scores, and negative values denote significance for the inverse condition effect.



**Figure 5.** Second LV (LV2), between-condition comparison: Contrasting the EEG response to the music and speech conditions across measures of MSE (left) and spectral power (right) for Cantonese speakers. The bar graphs (with standard error bars) depict brain scores that were significantly expressed across the entire data set as determined by permutation tests at 95% confidence intervals. These results are consistent with an interaction effect. The image plot highlights the brain regions and timescale or frequency at which a given contrast was most stable; values represent  $\sim z$  scores, and negative values denote significance for the inverse condition effect.

experience are associated with comparable patterns of information integration during automatic processing of music (i.e., nonlinguistic pitch) and speech (linguistic timbre). Between groups, we found that musicians had greater BSV than nonmusicians when listening to both music and speech stimuli. Cantonese speakers had the lowest entropy of all three groups for both stimulus conditions. Although this pattern of results was evident across multiple neural regions and timescales, it was particularly prominent in right hemisphere regions at coarse timescales. These data support the hypothesis that musicianship and tone language differentially impact information integration supporting both music (pitch) and speech (timbre) processing. It is notable that, although pitch cues are used extensively in both musicians' and Cantonese speakers' auditory experience, their information processing networks for pitch appears to be differentially shaped by their unique, domain-specific use and knowledge of this cue (Cantonese: linguistic pitch context; musicians: nonlinguistic pitch context).

The finding that musicians' increased BSV was most prominent in the right hemisphere corroborates the finding that this hemisphere is engaged in fine spectral features of auditory input, as compared with the left hemisphere, which is more specialized for temporal processing (see Zatorre, Belin, & Penhune, 2002, for a review). Similarly, expression in coarse timescales suggests that the dynamics supporting pitch and timbre processing are distributed, rather than locally based (Vakorin et al., 2011). Collectively, our findings indicate that musicians' processing of fine spectral features—both for pitch and timbre—is likely supported by a more expansive network than in Cantonese speakers and English-speaking nonmusicians. These data align with evidence that musicianship benefits a wide range of spectral processing (e.g., Parbery-Clark et al., 2013; Zendel & Alain, 2012; Bidelman & Krishnan, 2010; Chandrasekaran et al., 2009; Parbery-Clark et al., 2009; Schon et al., 2004) and, particularly, timbre (Hutka et al., 2015; Bidelman & Krishnan, 2010).

Overall, the results suggest that the extent of information integration during pitch processing is associated with whether one gained pitch experience via musicianship or speaking Cantonese. These results support our earlier prediction that, because of the higher auditory demands faced by musicians (relative to tone language speakers), the benefits to auditory processing in musicians are greater than to tone language speakers. It is interesting to contemplate if the present differences are rooted in the relative contributions of nature to each type of pitch experience. Specifically, there is evidence that suggests that musicianship is self-selected, with factors such as genetics (e.g., Tan, McPherson, Peretz, Berkovic, & Wilson, 2014), intelligence (Schellenberg, 2011), socioeconomic status (e.g., Sergeant & Thatcher, 1974), and personality traits (e.g., Corrigan, Schellenberg, & Misura, 2013), causing certain individuals to begin and continue music training, as compared with others (e.g., Schellenberg, 2015). Conversely, the networks

supporting Cantonese speakers' pitch processing are only subject to nurture, being born into a language that happens to use pitch to convey lexical meaning. Future studies could examine the link between preexisting factors and BSV related to pitch processing in musicians versus Cantonese speakers to determine the extent to which information processing capacity is shaped by nature rather than nurture.

In the between-conditions results, we found that each group engaged unique spatiotemporal distributions to process the differences between music and speech. Nonmusicians had greater BSV supporting speech processing, as compared with music processing (Figure 4). This difference was primarily expressed in several left hemisphere areas at fine timescales. The lateralization of this result is consistent with reports that in musically naive listeners, speech processing is more left-lateralized than music, given the left hemispheres' specialization for temporal processing (see Zatorre et al., 2002, for a review). These findings also suggest that nonmusicians may have greater, locally based information integration supporting speech processing, as compared with music processing (see Vakorin et al., 2011). Unlike for speech, this group's processing of music was right-lateralized, aligning with evidence for right hemisphere specialization for spectral processing (Zatorre et al., 2002).

Cantonese speakers had greater sample entropy for music as compared with speech (Figure 5). This distinction was primarily expressed in the midline posterior cingulate and retrosplenial cingulate cortex at fine timescales. This finding suggests that Cantonese speakers' use of lexical pitch may manifest for greater sample entropy for this cue, as compared with timbre. This finding aligns with the idea that the more familiar one is with a stimulus, the greater sample entropy associated with processing that stimulus (i.e., familiar vs. unfamiliar faces; Heisz et al., 2012). Finally, we did not detect a difference in musicians' BSV when processing music and speech sounds. This null result is consistent from what we would expect in musicians, as the auditory acuity honed by musicianship may enhance the information integration supporting both pitch and timbral cues in nonspeech and speech signals (i.e., music training benefitting speech processing; see Patel, 2011). Collectively, our data demonstrate that each group processes the distinction between music and speech using a different spatiotemporal network. Furthermore, the activation patterns for each group suggest a gradient of pitch processing capacity, such that the more experience one has with pitch (i.e., musicians > Cantonese > nonmusicians), the greater sample entropy associated with processing this cue. Namely, nonmusicians had greater sample entropy for speech as compared with music; Cantonese speakers had a greater sample entropy capacity for music than speech; musicians had similar levels of sample entropy for both conditions. An analogous gradient was observed in behavioral data for a pitch memory task in Bidelman, Hutka, et al. (2013). This gradient effect

suggests that musicianship hones more than just spectral acuity (unlike in Cantonese speakers and nonmusicians) and is thus associated with greater information integration supporting both pitch and timbre processing. Cantonese speakers only use pitch in a lexical context and thus have less information integration than musicians, but still more than that of nonmusicians.

### **Comparing MSE Results to the Spectral Analysis Results**

The MSE analyses yielded some unique information that was not obtained in the spectral analyses, as well as data that were complementary to the spectral analysis. Between-group comparisons of sample entropy revealed that musicians had greater brain signal complexity than tone language speakers, across all conditions. In contrast, spectral analyses revealed that musicians' processing of all conditions drew more heavily upon low theta/alpha (4–12 Hz) frequencies than in the other groups. Low frequencies of the EEG have traditionally been interpreted as reflecting long-range neural integration (von Stein & Sarnthein, 2000). Both the MSE and spectral results were also observed in similar neural regions. Collectively, both types of analyses suggest long range and more “global” processing of auditory stimuli in musicians compared with tone language speakers or nonmusicians. Indeed, the observation that whenever there is a preponderance of low frequencies, the entropy at longer timescales is higher suggests that there is a close relationship between the MSE and PSD. We have noted elsewhere, however, that MSE is more dependent on higher-order relations of the signal that are not present in measures of spectral density (McIntosh et al., 2008).

This global processing aligns with multiple neuroimaging findings in which musicians have regional anatomical differences that could facilitate interhemispheric communication, as compared with nonmusicians. For example, musicians—relative to nonmusicians—have a larger anterior corpus callosum, which is responsible for such interhemispheric communications and connecting premotor, supplementary motor, and motor cortices (Schlaug, Jancke, Huang, & Steinmetz, 1995). Numerous studies have since found differences in the corpus callosum between musicians and nonmusicians (e.g., Steele, Bailey, Zatorre, & Penhune, 2013; Hyde et al., 2009; Schlaug et al., 2009; Schlaug, Norton, Overy, & Winner, 2005), particularly in regions connecting motor areas (Schlaug et al., 2005, 2009). These differences may be honed by the bimanual coordination related to playing an instrument (Moore, Schaefer, Bastin, Roberts, & Overy, 2014).

Between-condition comparison of sample entropy revealed that each group showed unique spatiotemporal distributions in their response to processing music and speech. Nonmusicians had greater BSV for speech processing than music processing at fine timescales in several left hemisphere areas (e.g., anterior insula, centrolateral

and dorsomedial pFC, frontal polar area). The spectral data revealed beta and gamma frequency activity when processing speech (as compared with music) in similar neural regions as found in the MSE analysis. High-frequency activity has been associated with local perceptual processing (von Stein & Sarnthein, 2000) and is in accordance with the fine timescale (i.e., local) activation observed in our MSE analysis (Vakorin et al., 2011).

The spectral data characterizing the nonmusician effect differed from the MSE analyses via the results for the music condition. Specifically, low-frequency (theta) activation was associated with music processing in many of the same regions that expressed higher frequencies when processing speech. This suggests that nonmusicians may utilize longer-range neural integration to process music (von Stein & Sarnthein, 2000). However, this difference was not reflected in the MSE analysis (i.e., no increase in sample entropy at coarse timescales for the music condition), suggesting that nonmusicians do not have less information integration for music, relative to speech. This is plausible, as nonmusicians may have experience casually listening to/processing music,<sup>5</sup> but not the precise pitch experience present in musicians or Cantonese speakers.

In the MSE results for the Cantonese speakers, there was greater sample entropy for music as compared with speech—a difference that was primarily expressed at fine timescales in midline regions. Similarly, the spectral data showed that processing of music, as compared with speech, was associated with beta and gamma frequencies in similar neural regions as in the MSE results. Both the fine timescale and high-frequency activity suggest that the processing of music versus speech in Cantonese speakers relies on locally—rather than globally—distributed networks (Vakorin et al., 2011; von Stein & Sarnthein, 2000). There was also low-frequency (i.e., theta) activation associated with processing music, particularly in several right hemisphere areas (e.g., anterior insula, ventral temporal cortex, and fusiform gyrus), and with processing speech in the bilateral medial cortex. This low-frequency activity may suggest that Cantonese speakers utilize long-range neural integration to process music and speech (von Stein & Sarnthein, 2000). This does not align with either the local complexity supporting pitch processing, as suggested in the MSE data, or the low-frequency communication supporting such processing, as suggested in the high-frequency spectral data. Clarifying the global versus local nature of neural networks supporting music and speech processing in Cantonese speakers could be further investigated in future studies.

### **Comparisons to ERPs**

The EEG time series analyzed at present was previously examined in our previous ERP study (Hutka et al., 2015). MMNs (e.g., Näätänen, Paavilainen, Rinne, & Alho, 2007) were measured in these same groups to index early,

automatic cortical discrimination of music and speech sounds. In that study, only musicians showed an enhanced MMN response to both music and speech, aligning with the current between-group effects. Our collective findings suggest that musicians show greater automatic processing (Hutka et al., 2015) and information integration (present study) supporting the automatic processing of both music and speech, as compared with Cantonese speakers and controls.

However, we previously failed to find a difference in MMN amplitude between music and speech stimuli for any group (Hutka et al., 2015). In the current study, both sample entropy and spectral characteristics between conditions were observed in controls and Cantonese speakers. Furthermore, each group had a unique spatio-temporal distribution in response to music and speech. Despite having lower sample entropy than musicians or nonmusicians across all conditions, Cantonese speakers showed greater sample entropy for music as compared with speech. These data suggest that Cantonese speakers have greater information integration supporting pitch processing, as compared with timbral processing. In contrast, MMNs did not reveal a difference in automatic processing of music versus speech in the Cantonese group (Hutka et al., 2015). The differences between the MMN findings and the present results suggest that the nonlinear analyses currently applied afforded additional, more fine-grained information about these between-condition effects (see Hutka et al., 2013, for a discussion). That is, the averaging conducted to increase the signal-to-noise ratio in ERP analyses may eliminate important signal variability that carries information about brain functioning (Hutka et al., 2013).

## Conclusions

The present data suggest that the use of pitch for musicians versus for tone language speakers is associated with different information processing capacities supporting the automatic processing of pitch. Furthermore, each group's pitch processing was associated with a unique spatiotemporal distribution, suggesting that musicianship and tone language do not share processing resources for pitch, but instead, use different networks. This recruitment of different networks may help explain how similar behavioral pitch discrimination benefits in musicians and Cantonese speakers are not reflected in mean activations in response to pitch (i.e., Hutka et al., 2015). Collectively, these results further elaborate the discussion of music and speech processing in the context of experience-dependent plasticity. These data also serve as a proof of concept of the theoretical premise outlined in Hutka et al. (2013), namely, how applying a nonlinear approach to the study of the music–language association can advance our knowledge of each domain, as well as experience-dependent plasticity in general.

## Acknowledgments

This work was supported by the Ontario Graduate Scholarship (to S. H.), a GRAMMY Foundation grant (to G. M. B.), the Ministry of Economic Development and Innovation of Ontario (to S. M.), and the Natural Sciences and Engineering Research Council (RGPIN-06196-2014 to S. M.).

Reprint requests should be sent to Stefanie Hutka, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada, or via e-mail: stefanie.hutka@mail.utoronto.ca.

## Notes

1. That is, BSV does not in itself measure changes in connectivity; it simply measures the changes in dynamics associated with different connectivity patterns that might be the product of experienced-dependent plasticity.
2. The 72-region parcellation scheme was meant to reduce the dimensionality of the source map to something more meaningful than the 15,000 vertices generated by Brainstorm (Tadel et al., 2011). This specific parcellation scheme was used to maximize the definitional overlap of the regions with other reported regions in human and macaque data. These regions were mapped based on maximally agreed upon boundaries in the literature (see Kötter & Wanke, 2005).
3. Estimating sample entropy is based on nonlinear dynamics and employs a procedure called time delay embedding, for which we need to specify the embedding dimension. Time delay embedding is a critical step in the nonlinear analysis, but the choice of parameters is essentially based on heuristics. At the same time, Takens' (1981) embedding theorem states that, for reconstructing macrocharacteristics of a dynamical system underlying observed time series (such as entropy), embedding dimension should be relatively large. We used a priori  $m = 5$  for embedding dimension as a compromise between the requirements imposed by Takens' theorem and the fact that our time series are not only finite, but also relatively short.
4. The selection of this similarity criterion was guided by the simulations performed by Richman and Moorman (2000). Using a series of tests, they showed that the reconstructed values of sample entropy were close to the theoretical ones, when the tolerance parameter  $r$  was approaching 1, especially for relatively short time series.
5. Note that Bigand and Poulin-Charronnat (2006) discuss the large overlap in neural activity in musically trained and untrained listeners, in response to Western musical features (e.g., structure). On the basis of this evidence, one might predict that examining BSV in musicians versus untrained controls while listening to more complex musical excerpts might show a smaller BSV difference than one might initially anticipate.

## REFERENCES

- Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 97–103.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*, 425–434.
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One*, *8*, e60676.

- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125.
- Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, *79*, 201–212.
- Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*, 2662–2673.
- Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*, 1–9.
- Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*, 222.
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain [PhD thesis]*. Montreal: McGill University.
- Costa, M., Goldberger, A. L., & Peng, C. K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, *89*, 068102.
- Costa, M., Goldberger, A. L., & Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *71*, 021906.
- Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, *12*, 43–56.
- Deco, G., Jirsa, V. K., & McIntosh, A. R. (2013). Resting brains never rest: Computational insights into potential cognitive architectures. *Trends in Neurosciences*, *36*, 268–274.
- Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, *106*, 2896–2909.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*, 54–75.
- Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*, 20–36.
- Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2011). The importance of being variable. *Journal of Neuroscience*, *31*, 4496–4503.
- Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W., Lindenberger, U., McIntosh, A. R., & Grady, C. L. (2013). Moment-to-moment brain signal variability: A next frontier in human brain mapping?. *Neuroscience & Biobehavioral Reviews*, *37*, 610–624.
- Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R., & Jirsa, V. K. (2008). Cortical network dynamics with time delays reveals functional connectivity in the resting brain. *Cognitive Neurodynamics*, *2*, 115–120.
- Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, *2*, 146.
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, *101*, E215–E220.
- Good, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.
- Gudmundsson, S., Runarsson, T. P., Sigurdsson, S., Eiriksdottir, G., & Johnsen, K. (2007). Reliability of quantitative EEG features. *Clinical Neurophysiology*, *118*, 2162–2171.
- Heisz, J. J., Shedden, J. M., & McIntosh, A. R. (2012). Relating brain signal variability to knowledge representation. *Neuroimage*, *63*, 1384–1392.
- Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 10240–10245.
- Hutka, S., Bidelman, G. M., & Moreno, S. (2013). Brain signal variability as a window into the bidirectionality between music and language processing: Moving from a linear to a nonlinear model. *Frontiers in Psychology*, *4*, 984.
- Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, *71*, 52–63.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*, 3019–3025.
- Jirsa, V. K., & Kelso, J. S. (2000). Spatiotemporal pattern formation in neural systems with heterogeneous connection topologies. *Physical Review E*, *62*, 8462.
- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*, 104–117.
- Koelsch, S., Maess, B., Gunter, T. C., & Friederici, A. D. (2001). Neapolitan chords activate the area of Broca. A magnetoencephalographic study. *Annals of the New York Academy of Sciences*, *930*, 420–421.
- Kötter, R., & Wanke, E. (2005). Mapping brains without coordinates. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *360*, 751–766.
- Lippe, S., Kovacevic, N., & McIntosh, A. R. (2009). Differential maturation of brain signal complexity in the human auditory and visual system. *Frontiers in Human Neuroscience*, *3*, 48.
- McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, *3*, 143–157.
- McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLOS Computational Biology*, *4*, e1000106.
- McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: Applications and advances. *Neuroimage*, *23(Suppl. 1)*, S250–S263.
- McIntosh, A. R., Vakorin, V., Kovacevic, N., Wang, H., Diaconescu, A., & Protzner, A. B. (2013). Spatiotemporal dependency of age-related changes in brain signal variability. *Cerebral Cortex*, bht030.
- Misic, B., Mills, T., Taylor, M. J., & McIntosh, A. R. (2010). Brain noise is task dependent and region specific. *Journal of Neurophysiology*, *104*, 2667–2676.
- Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., & Overy, K. (2014). Can musical training influence brain connectivity? Evidence from diffusion tensor MRI. *Brain Science*, *4*, 405–427.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590.
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*, 713–719.
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*, 653–661.

- Parbery-Clark, A., Strait, D. L., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *Journal of Neuroscience*, *33*, 16741–16747.
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, *24*(Suppl. D), 5–12.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, Article 142.
- Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*, 134.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.
- Rattanasone, N. X., Attina, V., Kasisopa, B., & Burnham, D. (2013). How to compare tones. In *South and Southeast Asian psycholinguistics*. Cambridge: Cambridge University Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998). Advanced progressive matrices. In *Harcourt Assessment*. San Antonio, TX: Harcourt Assessment.
- Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology: Heart and Circulatory Physiology*, *278*, H2039–H2049.
- Sampson, P. D., Streissguth, A. P., Barr, H. M., & Bookstein, F. L. (1989). Neurobehavioral effects of prenatal alcohol: Part II. Partial least squares analysis. *Neurotoxicology and Teratology*, *11*, 477–491.
- Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, *102*, 283–302.
- Schellenberg, E. G. (2015). Music training and speech perception: A gene-environment interaction. *Annals of the New York Academy of Sciences*, *1337*, 170–177.
- Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, *25*, 241–252.
- Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A., & Winner, E. (2009). Training-induced neuroplasticity in young children. *Annals of the New York Academy of Sciences*, *1169*, 205–208.
- Schlaug, G., Jancke, L., Huang, Y., & 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. *Annals of the New York Academy of Sciences*, *1060*, 219–230.
- 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.
- Sergeant, D., & Thatcher, G. (1974). Intelligence, social status and musical abilities. *Psychology of Music*, *2*, 32–57.
- Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, *21*, 143–163.
- Steele, C. J., Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2013). Early musical training and white-matter plasticity in the corpus callosum: Evidence for a sensitive period. *Journal of Neuroscience*, *33*, 1282–1290.
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Braintstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 879716.
- Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand & L. S. Young (Eds.), *Dynamical systems and turbulence, lecture notes in mathematics* (Vol. 898, pp. 366–381). Berlin: Springer-Verlag.
- Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology*, *5*, 658.
- Vakorin, V. A., Misić, B., Krakovska, O., & McIntosh, A. R. (2011). Empirical and theoretical aspects of generation and transfer of information in a neuromagnetic source network. *Frontiers in Systems Neuroscience*, *5*, 96.
- von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology*, *38*, 301–313.
- Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L. H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS One*, *7*, e33424.
- Yip, M. (2002). *Tone*. New York: Cambridge University Press.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*, 410–417.