

Influence of Musical Expertise on the processing of Musical Features in a Naturalistic Setting

Dipankar Niranjana¹ (dipankar.niranjana@research.iiit.ac.in)

Iballa Burunat² (iballa.burunat@jyu.fi)

Petri Toiviainen² (petri.toiviainen@jyu.fi)

Elvira Brattico³ (elvira.brattico@clin.au.dk)

Vinoo Alluri¹ (vinoo.alluri@iiit.ac.in)

¹ Cognitive Science Lab, Kohli Center on Intelligent Systems, IIIT Hyderabad, Hyderabad, India

² Department of Music, University of Jyväskylä, Jyväskylä, Finland

³ Department of Clinical Medicine - Center for Music In the Brain, Aarhus University, Aarhus, Denmark

Abstract

Musical training causes structural and functional changes in the brain due to its sensory-motor demands, but the modulatory effect of musical training on music feature processing in the brain in a continuous music listening paradigm, has not been investigated thus far. In this work, we investigate the differences between musicians and non-musicians in the encoding of musical features encompassing musical timbre, rhythm and tone. 18 musicians and 18 non-musicians were scanned using fMRI while listening to 3 varied stimuli. Acoustic features corresponding to timbre, rhythm and tone were computationally extracted from the stimuli and correlated with brain responses, followed by t-tests on group level maps to uncover encoding differences between the two groups. The musicians demonstrated greater involvement of limbic and reward regions, and regions possessing adaptations to music processing due to training, indicating greater analytic processing. However, as a group, they did not exhibit large regions of consistent correlation patterns, especially in processing high-level features, due to differences in processing strategies arising out of their varied training. The non-musicians exhibited broader regions of correlations, implying greater similarities in bottom-up sensory processing.

Keywords: fMRI; Musicians vs Non-Musicians; Music Processing; Acoustic Feature Extraction; Naturalistic Stimulus

Introduction

Musical training is known to cause structural and functional changes in the brain (Angulo-Perkins et al., 2014; Gaser & Schlaug, 2003). Moreover, since listening to music recruits several brain regions other than the sensory auditory cortices, musical training is also associated with transfer effects such as enhanced cognitive function in language processing, motor abilities, attention and memory, visuo-spatial abilities, and overall cognitive development in general (Miendlarzewska & Trost, 2014; Tierney et al. 2013). This makes music a great

tool to study brain adaptation, and musicians an ideal group to study brain changes driven by experience.

The repetitive sensory-motor workload that musical training entails has been known to engender enhanced sensitivity in the processing, representation, and discrimination of sounds and music (Musacchia et al., 2008), especially in subcortical encodings of acoustic features such as pitch, timbre and timing (Kraus et al., 2009). To further investigate brain plasticity and transfer effects, it is of great importance to understand specifically, which brain regions and circuits are recruited during music processing and the modulatory effect of musical expertise thereof. In this study, we adopt an interdisciplinary approach combining neuroimaging with computational acoustic feature extraction to investigate the differences in how musicians and non-musicians process acoustic features during continuous music listening in a naturalistic setting. We look at a set of six acoustic features which were computationally extracted from the musical stimuli (Alluri et al., 2012) - Activity, Brightness, Fullness, Timbral Complexity, Pulse Clarity and Key Clarity. Activity, Brightness, Fullness and Timbral Complexity are lower level (requiring low-level sensory processing), short term Timbral features while Pulse Clarity (Rhythm) and Key Clarity (Tone) are higher level (requiring processing mechanisms which also depend on prior knowledge and exposure to music), long term features. We thus also investigate how musical expertise modulates the processing of features of varying complexity and temporal nature. Prior work consists of either studies of group differences (musicians vs. non-musicians) in a controlled auditory setting (acoustic features presented in isolation), or, studies in a naturalistic setting where musical training has not been taken into account. Our work bridges this gap. As music is multidimensional in nature, investigations in a naturalistic setting present a more holistic picture of the neural underpinnings of music processing.

Methods

Participants and Stimuli The participant pool consisted of 18 musically trained (9 female, age 28.2 ± 7.8 years) and 18 un-



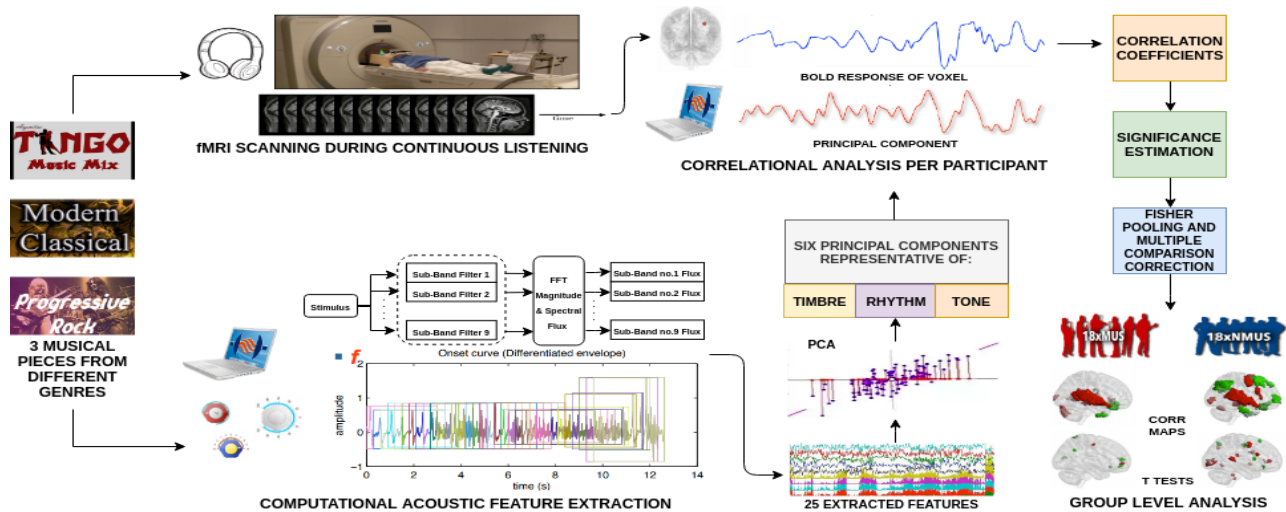


Figure 1: An overview of our study.

trained ($10, 29.2 \pm 10.7$) participants. Both groups were comparable with respect to cognitive measures (WAIS-WMS III scores) and socioeconomic status (Hollingshead's FFI). The total number of years of training for musicians was 16 ± 5.7 years. The number of hours spent practicing music on average per week was 16.6 ± 11 . The data was collected as part of a broader project ("Tunteet") involving other tests (neuroimaging and neurophysiological measures). The study protocol was approved by the ethics committee of the Coordinating Board of the Helsinki and Uusimaa Hospital District. Written consent was obtained from all the participants. They were asked to listen to three instrumental pieces - Adios Nonino by Astor Piazzolla (tango nuevo), the first three dances of the Rite of Spring by Igor Stravinsky (modern classical), and Stream of Consciousness by Dream Theater (progressive rock). Each piece was roughly around 8 min long and each belonged to a different genre (to allow for generalization of the findings). All three pieces contained a high amount of variation in acoustic features such as timbre, tonality, rhythm etc., while exhibiting a comparable musical structure.

fMRI data acquisition Participants brain responses were acquired while they listened to the musical stimuli presented in a counterbalanced order. Their only task was to attentively listen to the music delivered via MR-compatible insert earphones while keeping their eyes open. MRI data was collected at the Advanced Magnetic Imaging Centre, Aalto University, Finland, on a 3T Siemens MAGNETOM Skyra, TR = 2s, TE = 32ms, whole brain, voxel size: $2 \times 2 \times 2 \text{ mm}^3$, 33 slices, FoV: 192 mm (64×64 matrix), interslice skip = 0 mm. fMRI scans were preprocessed on Matlab using SPM8, VBM5 and custom scripts. Normalization to MNI segmented tissue template was carried out. Head movement related components were regressed out, followed by spline interpolation and filtering.

Acoustic feature extraction and processing We chose acoustic features that broadly capture the timbral, tonal and rhythmic aspects of the stimuli. The feature set of 25 features can be classified into short and long-term features based on the duration of the analysis window (25 ms and 3s). The

short-term features encapsulate timbral properties, and are, the zero crossing rate, spectral centroid, high energy-low energy ratio, spectral spread, spectral roll-off, spectral entropy, spectral flatness (Wiener entropy), roughness, RMS energy, spectral flux, and Sub-Band Flux (10 coefficients). The long-term features encapsulate context-dependent aspects of music, such as tonality and rhythm, and are, pulse clarity, fluctuation centroid, fluctuation entropy, mode and key clarity. All the features were extracted from each stimulus using the MIR-Toolbox (Lartillot & Toivainen, 2007). The time series' of these features were processed to account for haemodynamic response lag, filtered for noise, and downsampled to match the fMRI data. To reduce the number of features, PCA was performed. The first 9 PCs explained 95% of the variance. 6 PCs which were perceptually validated in prior studies (Alluri et al., 2010, 2012) were chosen and labelled as: Activity, Fullness, Brightness, Timbral Complexity (these four constitute timbral features encapsulating sound character distinct from its pitch and intensity), Pulse Clarity (regularity of pulse/rhythm), and Key Clarity (strength of tonality). These will henceforth be referred to as acoustic components.

Statistical Analysis To boost the statistical power of the results, we present further analysis on the concatenated stimuli (behavioral ratings revealed no significant differences between the groups for the mean valence and arousal ratings for each of the stimuli). First, the inter-subject consistency of the fMRI data was checked using mean inter-subject correlation as a measure. Then, per participant, the Pearson correlation coefficients (r) were obtained per voxel, per acoustic component. The r maps were converted to Fisher Z-score maps normalized by the estimated effective degrees of freedom (estimated using the approach described by Pyper and Peterman (1998)). For each acoustic component, to obtain group maps (musicians and non-musicians), we utilized the combining tests procedure described by Lazar (2008). Individual Z-score maps are converted to p-maps, and are then pooled using Fisher's p-value technique to create group maps. The obtained group maps for each acoustic component were thresh-

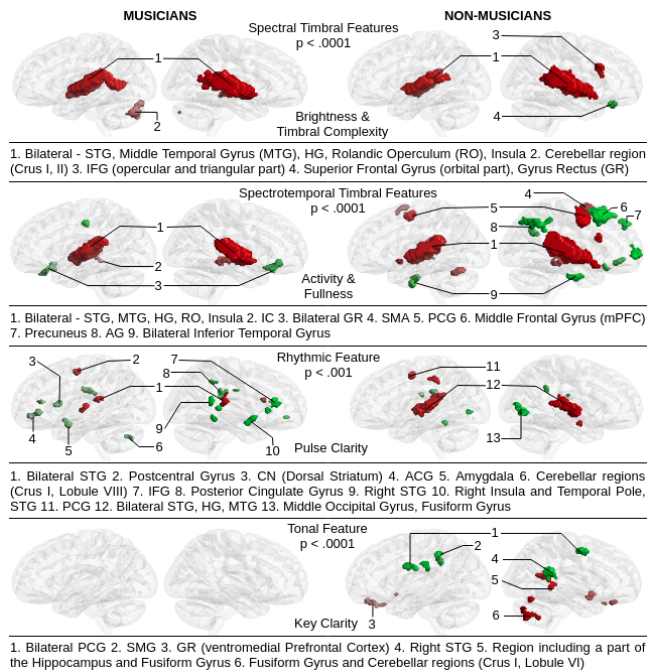


Figure 2: Group level correlations with acoustic components. Lateral view of the left and right hemispheres indicating positive (red) and negative (green) correlations.

olded at a significance level of $p < .0001$. To minimize Type I errors, we performed multiple comparisons correction using cluster size thresholding. To determine the thresholds for this, we performed a Monte Carlo simulation of the approach described by Ledberg et al. (1998). Next, in order to examine the differences in correlation between the two groups, voxel-wise t-tests of the group maps were performed per acoustic component. The results were thresholded at a significance level of $p < .05$ followed by cluster-size thresholding.

Results and Discussion

In this study, we investigate the differences in neural responses arising between musicians and non-musicians in processing acoustic components while listening to music in a naturalistic setting. First, voxel-wise mean inter-subject consistency analysis revealed maximal values in bilateral auditory cortices, with a larger proportion in the right hemisphere for both groups. Subsequent correlational analysis revealed results that corroborate previous findings in addition to evidencing first-time differences between the groups in a continuous listening paradigm.

Per acoustic component, we first present and discuss the results for the group level correlational analysis. Then we discuss the t-test results (to uncover significant group differences). For conciseness, we group and plot the maps of the timbral components as follows - Brightness and Timbral Complexity as Spectral components, and, Activity and Fullness as Spectrotemporal components.

Timbral feature processing High levels of timbral components were associated with activations in the auditory regions

of the bilateral Superior Temporal Gyrus (STG) and the Heschl's Gyrus (HG) for both musicians and non-musicians, in line with previous literature, demonstrating that timbral processing is primarily associated with the sensory auditory processing regions (primary auditory cortex) for both groups. For musicians, positive correlations were observed in the cognitive areas of the cerebellum (Crus I, II) for the spectral components of Brightness and Timbral Complexity. Also, we demonstrate for the first time in a naturalistic setting, the involvement (positive correlations) of the Inferior Colliculus (IC) in processing spectrotemporal acoustic properties represented by Activity (roughness & flux from higher frequency bands) and Fullness (fluctuations in lower frequency bands). The IC is known to encode coarse spectral decompositions of complex sounds (Rodrigues, Read & Escabi, 2010). On the other hand, non-musicians displayed additional positive correlations in the Primary Motor Cortex (Precentral Gyrus (PCG)) and Supplementary Motor Area (SMA), which are part of the dorsal auditory pathway responsible for processing complex sounds focusing on the "how" and "where" of the sound (Bizley & Cohen, 2013). For non-musicians, we find that key regions belonging to the DMN including the Angular Gyrus (AG), Precuneus, medial Prefrontal Cortex (mPFC) show significant negative correlations with the spectrotemporal timbral features, indicating low auditory cognitive load at those times (Pallesen et al., 2009).

T-tests revealed that the musicians displayed significantly higher positive correlations than non-musicians focused around the left HG, for spectral components. Musicians are known to possess larger amounts of gray matter particularly in the left HG (Gaser and Schlaug, 2003) and hence this could further manifest as functional changes occurring due to musical expertise. On the other hand, non-musicians exhibited significantly higher positive correlations focused around the right-hemispheric Superior Temporal Sulcus (STS), PCG and SMA. The right secondary auditory areas (especially the STS) have been found to be key in encoding timbre (Alluri & Kadire, 2019). Significantly higher positive correlations observed in the PCG and SMA, part of the dorsal timbre processing pathway, in the non-musicians, suggests similarities in how they process the potential origins of the sound and that musical expertise modulates this pathway. The absence of significant correlations in these regions for musicians might indicate changes in coupling and functioning of these regions as a result of musical training. Alluri et al., (2015) observed that greater coupling exists between the limbic regions and the SMA in musicians, thereby potentially altering its role from that of a bottom-up one of processing low-level sensory timbral features, to a top-down one processing higher level attributes related to the emotional content of the music. The significant higher negative correlations observed in the non-musicians in the key nodes of the DMN (i.e. right AG & mPFC) is in line with Alluri et al.'s (2017) observation of higher coupling of these areas in non-musicians as compared to musicians (for whom similar correlations are observable at more lenient thresholds of $p < .005$), implying possible changes in intra-DMN connec-

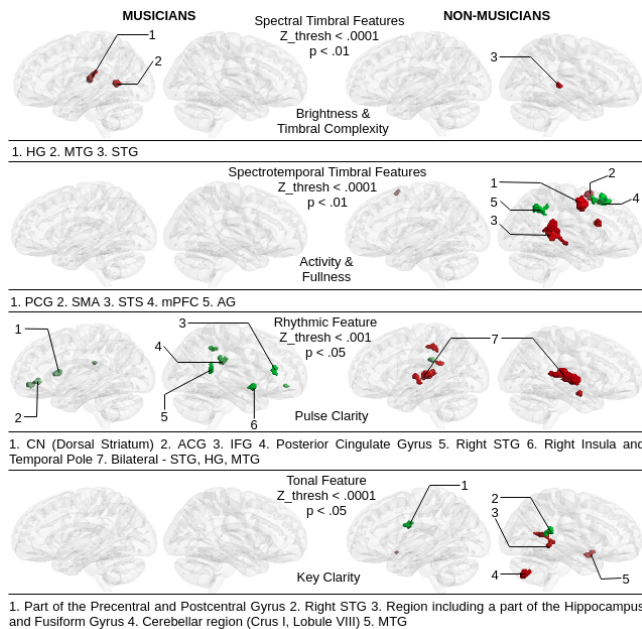


Figure 3: T-Tests to uncover significant group differences.

tivity due to musical training.

Rhythmic and Tonal feature processing For Pulse Clarity, musicians displayed negative correlations in the left hemispheric Caudate Nucleus (CN)(Dorsal Striatum), Anterior Cingulate and Paracingulate Gyrus (ACG), Amygdala, and Cerebellar regions (Crus I, Lobule VIII), and in the right hemispheric Inferior Frontal Gyrus (IFG), STG, Posterior Cingulate Cortex, and Insula. The CN has been associated with encoding information related to musical meter, and also in the anticipation and prediction of reward (Salimpoor et al., 2015; Trost et al., 2014). The ACG is a prominent node in the salience network and is implicated in numerous complex functions by means of integrating sensory, emotional and cognitive information and plays also plays a key role in attentional control (Menon & Uddin, 2010). The right IFG and STG have been associated with increased activations in musically trained individuals, more so when violations arise in well defined musical regularities (Koelsch & Siebel, 2005). On the other hand, the non-musicians primarily demonstrated large areas of positive correlations in the primary auditory cortices (bilateral STG, HG) demonstrating primarily sensory processing. Negative correlations were found in the bilateral Middle Occipital Gyrus and the right Fusiform Gyrus. Also, both groups showed positive correlations in a small area in the motor region of PCG, possibly indicating monitoring/maintenance of pre-existing rhythm in the music (Tanaka et al., 2005).

The t-test results indicated significantly greater positive correlations in the auditory regions for non-musicians, while the musicians showed significantly greater negative correlations in the limbic and reward regions, and in regions possessing adaptations due to musical training (CN, ACG, rIFG/rSTG). Unlike the non-musicians group who primarily seek to monitor pre-existing rhythm, the musicians (in addition to monitoring pre-existing rhythm) also exhibit activations during moments

of low Pulse Clarity - possibly suggesting that they try to internally anticipate/seek pulse during those moments (Grahn & Rowe, 2009). This sense of anticipation and tension during moments of low Pulse Clarity could also be the reason for pleasurable sensations leading to negative correlations in the limbic regions (Salimpoor et al., 2011).

For Key Clarity, the non-musicians displayed positive correlations in the ventromedial prefrontal cortex (Gyrus Rectus) and in the Fusiform/Middle Temporal Gyrus region. These regions have shown activations in tone related processing (Janata et al., 2002). Positive correlations were also seen in the reward regions - CN and the Hippocampus, and also in the cerebellum (limbic region) - indicating a possible link between non-musicians and their preference for tonally clear parts of the music. Negative correlations were found in the PCG, the Supramarginal Gyrus (SMG) and the right STG. The PCG, as explained earlier is a key component of the dorsal auditory pathway responsible for processing complex sounds, thus possibly leading to activations in parts of the music with unclear tone. The SMG is known to be involved in auditory memory (Schaal, Pollock & Banissy, 2017). Further investigation with respect to its role in tonal processing is called for. The musicians did not show any significant correlations.

T-tests indicated significantly greater positive correlations for non-musicians in the Cerebellar (limbic) and Hippocampal (reward) regions, and also in the MTG. The non-musicians also displayed significantly higher negative correlations in left Pre and Postcentral gyrus and in the right STG. These results indicate broader regions of activations in non-musicians, as compared to musicians, indicating more commonality in listening strategies. Key Clarity is a high level, top-down cognitive feature and hence its perception and processing would differ in musicians based on differences in training (genre, primary instrument, prior experience, etc.) (Burunat et al., 2015).

To summarize, we investigate the modulatory effect of musical training on music feature processing in the human brain during continuous music listening. The musicians as a group, show increased integration and involvement of limbic and reward regions, along with regions possessing adaptations to music feature processing due to training. This training and prior experience enables top-down analytic processing of high-level rhythmic and tonal features leading to differences in listening and processing strategies. Hence, as a group, they do not exhibit large regions of consistent correlation patterns. On the other hand, the non-musicians exhibit broader regions of correlations implying greater similarity in sensory listening strategies. They also exhibit consistent patterns in DMN regions in addition to auditory cortical regions.

Acknowledgments

This work was supported by the Academy of Finland (project numbers 272250 and 274037), and the Danish National Research Foundation (DNRF117).

References

References have been cited and linked inline with the text.