Musicianship can be decoded from magnetic resonance images

Tuomas Puoliväli[a,*], Tuomo Sipola[a,1], Anja Thiede[b,e], Marina Kliuchko[e], Brigitte Bogert[e], Petri Toiviainen[c], Asoke K. Nandi[d,i], Lauri Parkkonen[b,j], Elvira Brattico[h,e,j], Tapani Ristaniemi[a], Tiina Parviainen[f,g]

[a]*Faculty of Information Technology, University of Jyväskylä, Finland*
[b]*Department of Neuroscience and Biomedical Engineering, Aalto University, Finland*
[c]*Finnish Centre for Interdisciplinary Music Research, Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*
[d]*Department of Electronic and Computer Engineering, Brunel University London, United Kingdom*
[e]*Cognitive Brain Research Unit, Department of Psychology and Logopedics, University of Helsinki, Finland*
[f]*Department of Psychology, University of Jyväskylä, Finland*
[g]*Centre for Interdisciplinary Brain Research, University of Jyväskylä, Finland*
[h]*Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University & Royal Academy of Music Aarhus / Aalborg (RAMA), Denmark*
[i]*College of Electronic and Information Engineering, Tongji University, Shanghai, China*
[j]*Advanced Magnetic Imaging Centre, Aalto Neuroimaging, School of Science, Aalto University, Espoo, Finland*

Abstract

Learning induces structural changes in the brain. Especially repeated, long-term behaviors, such as extensive training of playing a musical instrument, are likely to produce characteristic features to brain structure. However, it is not clear to what extent such structural features can be extracted from magnetic resonance images of the brain. Here we show that it is possible to predict whether a person is a musician or a non-musician based on the thickness of the cerebral cortex measured at 148 brain regions encompassing the whole cortex. Using a supervised machine-learning technique, we achieved a significant ($\kappa = 0.321, p < 0.001$) agreement between the actual and predicted participant groups of 30 musicians and 85 non-musicians. The areas contributing to the prediction were mostly in the frontal, parietal, and occipital lobes of the left hemisphere. Our results suggest that decoding musicianship from magnetic resonance images of brain structure is feasible. Further, the distribution of the areas that were informative in the classification, which mostly, but not entirely, overlapped with earlier findings on areas relevant for musical skills, implies that decoding-based analyses of structural properties of the brain can reveal novel aspects of musical aptitude. In particular, our results highlight differences in visual areas in addition to the already more established differences located in motor networks and networks of higher-order cognitive function.

Keywords: Decoding, Learning, Magnetic resonance imaging, Machine learning, Musician, Support vector machines

1. Introduction

The gross structural properties of the brain are shaped during peri- and postnatal development, and genetic guidance continues to influence brain formation throughout infancy and childhood (see e.g. Pletikos et

*Corresponding author at Faculty of Information Technology, University of Jyväskylä, PL 35, Jyväskylä, Finland. Email: tuomas.a.b.puolivali@jyu.fi, tuomas.puolivali@ieee.org.

[1]T. Sipola left the University of Jyväskylä during the preparation of this manuscript, and currently works at JAMK University of Applied Sciences.

al., 2014). Especially during postnatal development, structural changes reflect interactions between genetic and environmental factors. However, also learning affects brain structure, and the role of the environment and individual's own actions in influencing gray- and white-matter properties are likely to be particularly strong during development. Training-related structural changes have been observed to occur throughout the human lifespan in both the gray and white matter (Maguire et al, 2006; Scholz et al. 2009; Zatorre et al. 2012; Sampaio-Baptista et al. 2014). Clear evidence of the influence of sensory environment can be seen for example in the auditory cortices of bilingual children: more variable language and auditory environment increases the volume of the Heschl's gyri (Ressel et al. 2010). Similarly, learning a second language increases the gray matter density of the left inferior parietal cortex in both adolescents and adults (Mechelli et al, 2004). Taken together, a magnetic resonance image (MRI) of a person's brain should contain information about individual abilities, especially for well-trained skills that are acquired early in life.

Hence, does possessing a skill manifest as a detectable trace in the structural MRI of a person's brain? The answer is likely to depend on at least the magnitude and type of the structural changes associated with the acquisition of a specific skill. Since playing music requires mastering of a complex set of motor and multi-sensory skills, the learning-related structural changes could be expected to be highly distributed across the cortex. Indeed, there is now a large body of evidence showing functional and structural differences between musicians' and non-musicians' brains (Elbert et al., 1995; Münte et al., 2002; Gaser & Schlaug, 2003a,b; Wong et al., 2007; Bermudez et al. 2009; Barrett et al. 2013; Miendlarzewska et al. 2013; Moore et al. 2014; Reybrouck & Brattico, 2015). Musicians are therefore an ideal population for testing whether the structural information obtained with MRI would allow decoding of acquired expertise, since adult professional musicians have typically started the intensive training at an early age, and it is easy to find control participants with little or no musical training.

In previous neuroscience research, machine learning has been more widely applied to functional but not so much to structural neuroimaging data. For example, studies conducted with functional magnetic resonance imaging (fMRI) have succeeded in reliably decoding reward-based decision-making processes (Hampton & O'Doherty, 2007), visual experiences and visual imagery during sleep (Nishimoto et al. 2011; Horikawa et al. 2013), moral judgments (Koster-Hale et al. 2013), and unconscious or covert mental states (Haynes & Rees, 2006). Experiments using other brain imaging modalities, namely electroencephalography (EEG) and magnetoencephalography (MEG), have demonstrated that decoding hand movement directions (Waldert et al. 2008), semantic categories of words (Cichy et al. 2014) and aware vs. unaware visual percepts (Salti et al. 2015) is possible, to name a few examples. In the specific domain of music-related expertise, it has been previously possible to decode individuals' musicianship classes from fMRI data that was recorded while the participants were listening to different pieces of music (Saari et al., 2018). While these and other functional imaging techniques allow inferring and contrasting brain states, traits are perhaps better quantified from structural data. Indeed, functional imaging techniques only capture processes elicited by the performed task.

There are many clinical applications based on data-driven classification of structural MRIs , such as diagnosis of Alzheimer's disease (Moradi et al. 2015), or identification of individuals at risk for bipolar disorder (Hajek et al. 2015). However, relatively few investigations on the feasibility of using machine learning techniques for classification of non-clinical MRI features have been reported to date. Given that brain structure reflects a multitude of influences and processes, and which must be relatively highly preserved to allow normal function, such tasks are inherently difficult. In this study, we show for the first time that it is possible to decode musicianship from the brain structure: to predict whether a person is a musician or a non-musician by analyzing cerebral cortical thickness data with support vector machines (SVM), which is a supervised machine-learning technique for binary classification problems (Cortes & Vapnik, 1995).

A difficult problem in structural MRI studies is identifying which findings are at the same time statistically and neurobiologically significant for a given cognitive function or behavioral skill. In principle, a small difference in one region can be much more relevant for a given cognitive function than a larger difference in

another region: behaviorally (psychologically, clinically) meaningful variation is embedded within variation due to other causes across individuals. It is possible to address causal questions by disrupting brain activity for example by using transcranial magnetic stimulation, but the relative stability of brain structure renders any interventional approaches impossible in humans. Hence, how should one decide whether a statistically significant difference in cortical thickness, reported in millimeters, is meaningful in behavioral terms?

In this regard, machine-learning or decoding models may be useful: if a set of measurements systematically predicts the desired target value, one can conclude that the measurements contain information about the property of interest, which often has more value than merely reporting group differences. In this study, we attempt to decode musicianship by classifying the participants as either professional performing musicians or non-musician control participants based on their MRIs. While structural differences between musicians' and non-musicians' brains have been investigated for decades, it has remained unknown which cortical structures best allow decoding musicianship. By using reliable and extensively validated machine-learning and MRI processing methods, we address this issue in this study.

2. Materials and methods

2.1. Participants

A total of 121 professional, amateur, and non-musician participants were recruited to this study. Six participants were excluded from the data analysis due to neurological or psychiatric disorders. The recruited group of professionals consisted of rock, pop, jazz, and classical musicians who had started playing their first instrument before the age of thirteen. The non-musician and amateur musician groups were merged for the binary classification task. The final participant demographics are summarized to Table I. The study was part of the "Tunteet" protocol approved by the Coordinating ethics committee of the Helsinki and Uusimaa Hospital District. Written informed consent was obtained from each participant prior to the measurements. The research was performed in compliance with the Declaration of Helsinki by World Medical Association.

2.2. MRI data acquisition

The MR images were acquired at the Advanced Magnetic Resonance Imaging (AMI) center of Aalto Neuroimaging, Aalto University (Espoo, Finland). A Siemens Magnetom Skyra 3 T whole-body scanner (Siemens Healthcare, Erlangen, Germany) with a standard 20-channel head-neck coil was used. A gradient-echo (MP-RAGE) T1-weighted sequence with repetition time, echo time, inversion time, and flip angle of 2530 ms, 3.3 ms, 1100 ms, and 7 degrees, respectively, was used. Voxel size was set as 1 mm^3.

2.3. Extraction of cerebral cortical thickness data

Automated cerebral cortical thickness measurements were performed using the FreeSurfer MRI analysis software suite (Dale et al. 1999; Fischl et al. 1999), which is available online at `https://surfer.nmr.mgh.harvard.edu/`. FreeSurfer has been shown to make accurate and precise volumetric estimates in several validation studies (Fischl et al. 2002; Han et al. 2006; Wonderlick et al. 2009; Cardinale et al. 2014). The left and right hemispheres of the cerebral cortex were both parcellated into 74 regions according to the Destrieux atlas (Destrieux et al. 2010). The average thickness of gray matter at each of the 148 regions was estimated. The intermediate steps of the automated procedures were manually checked for obvious errors, and corrected, if needed. The smallest measured cortical thickness was 1.47 mm and the largest 4.14 mm, suggesting that the data did not contain any major outliers, since the human cerebral cortical thickness is known to vary approximately between 1.0 and 4.5 mm (Fischl et al. 2000).

2.4. *Control for confounding effects*

Cortical thickness decreases with age (Salat et al. 2004; Thambisetty et al. 2010; Toga et al. 2011). Since the participants in our study spun a wide range of ages, the effect may have consequences for our study. If, for example, we trained a classifier with young musicians, and tested it with old musicians, the performance would be probably worse than with age-balanced groups. To test the existence of this effect in our data, we fitted a regression model using age as the independent variable and mean cortical thickness as the dependent variable, and examined the coefficient of determination. A successful replication of the previous studies would imply a need for age-balanced training and test sets. Because previous studies have also identified structural differences between the brains of males and females (Im et al. 2006; Sowell et al. 2007), we also examined the mean cortical thickness with respect to the participants' gender.

2.5. *Classification and cross-validation*

Supervised machine learning is about finding a mapping $X \rightarrow Y$ where $x_i \in X$ is a sample and $y_i \in Y$ is a group label, when having a limited number of labeled samples $(y_1, x_1), \ldots, (y_n, x_n)$. Using SVMs (Cortes and Vapnik, 1995), we studied a two-class problem with $y_i \in \{$musician, non-musician$\}$ and each sample x_i containing the cerebral cortical thickness measurements of the participant i. The SVM classifier was selected because it is known to perform well in many kinds of classification tasks (Fernández-Delgado et al. 2014) and its implementation is available in many machine-learning packages.

To use the data as efficiently as possible, we used nested stratified k-fold cross-validation for model selection and validation (Kohavi, 1995) (Figure 1). We selected to test linear and radial basis function kernels for the SVM and k = 10 for both the inner and outer cross-validation loops. Since age but not gender was identified to have an effect on the mean cortical thickness (see Sections 3.1 and 3.2), we balanced the folds with respect to participant and age groups. The coarse logarithmic grid $\{1, 10, 100, 1000\}$ was searched to optimize the SVM C parameter, which controls the generalization ability. Since the participant groups were imbalanced, the parameter C was adjusted for group j as Cw_j where w_j is a weight inversely proportional to the group frequency.

The machine learning pipeline was implemented in Python version 3.2.0 from the Python Software Foundation. For cross-validation and classification we employed the Scikit-learn module (Pedregosa et al. 2011; Abraham et al. 2014), which internally uses the LIBSVM library (Chang and Lin, 2011) for SVMs. The cortical surface visualizations were produced with PySurfer, which is available online at `https://pysurfer.github.io/`.

2.6. *Variable selection and scaling*

To identify an informative subset of the most relevant variables for constructing the classifier, we used a filter-type variable selection procedure as a preprocessing step (Weston et al., 2000; Gyuon & Elisseef, 2003). The Student's t-test for two independent samples with the critical level $\alpha = 0.05$ was used to discard variables for which the null hypothesis of equal means could not be rejected. The variable selection was repeated on each cross-validation iteration. Therefore, we used the number of times a variable was selected as an indicator of its relative importance. The selected variables were transformed to standard scores by removing the mean and scaling to unit variance to avoid variables with large numerical ranges to dominate the classifier fitting process.

2.7. *Evaluation of classifier performance*

We used Cohen's kappa (Landis and Koch, 1977) as the primary metric for evaluating the classifier's performance. This metric is zero for chance-level agreement between the actual and predicted participant groups, $+1$ for complete agreement, and -1 for complete disagreement. It is calculated as $\kappa = (\pi_o - \pi_e)/(1-\pi_e)$ where π_o is the observed agreement (percentage of correct predictions) and π_e is the probability of an agreement occurring by chance. The metric can be rewritten as $\kappa = 1 - (1 - \pi_o)/(1 - \pi_e)$, where $1 - \pi_o$ is the error rate and $1 - \pi_e$ is the probability of a disagreement occurring by chance.

2.8. *Statistical assessment of over-fitting*

To investigate the reliability of the classifier, we used the following permutation test. First, the group labels of the entire data were randomly permuted. The classifier training and testing were then performed with the expectation of achieving chance-level prediction metrics on average. Next, this process was repeated 1000 times and the prediction metrics collected as the null distribution. Finally, the classifier's performance was compared to this distribution. If the probability of obtaining equal or better results with the randomly-permuted group labels was less than the critical level $\alpha = 0.05$, we considered the classifier reliable.

3. Results

3.1. *Mean cortical thickness as a function of age*

The relationship between the participants' age and mean cerebral cortical thickness appeared to be linear; see Figure 3 for an illustration. A regression analysis showed that the mean cortical thickness decreased significantly with age ($F_{1,114} = 39.72, p < 0.001$); mean cortical thickness $= 2.652 - 0.005 \times$ age; $r^2 = 0.260$. The 95 % confidence interval for the slope was $[-0.007, -0.003]$. The result was in agreement with previous studies (Salat et al. 2004; Thambisetty et al. 2010; Toga et al. 2011) and implied a need for age-balanced cross-validation folds.

3.2. *Mean cortical thickness as a function of gender*

The mean cerebral cortical thickness was 2.51 mm (standard deviation was 0.90 mm) for the male participants and 2.51 mm (0.70 mm) for the female participants. The 95 % confidence intervals of the means were $[2.26, 2.75]$ and $[2.34, 2.68]$ for males and females respectively. The measurements were approximately normally distributed in both groups. There was no significant difference between the means according to an independent two-sample t-test: $t(113) = 0.036, p = 0.971$. The data are illustrated with box plots in Figure 2. The result implies that the cross-validation folds do not need to be balanced with respect to gender.

3.3. *Classifier performance*

The linear SVM with nested cross-validation and group-frequency-weighted samples achieved a significant inter-rater agreement ($\kappa = 0.321, p < 0.001$) between the actual and predicted participant groups which was an imbalanced sample of musicians and control participants. It correctly classified 19 of the 30 (63.3 %) musician and 62 of the 85 (72.9 %) non-musician participants. In the set of 73 participants classified as non-musicians, 62 (84.9 %) were actually non-musicians. In the set of 42 participants classified as musicians, 19 (45.2 %) were actually musicians. The confusion matrix of the classification results is shown in Table II. A radial basis kernel function for the SVM was also tested, but the results were not analyzed further due to over-fitting, which was assessed using the permutation test described in section 5.8 ($p > 0.05$).

Post-hoc tests were performed to assess possible causes of participant misclassification. For the demographic data, there was a significant difference in age between the correctly and incorrectly classified non-musicians (Student's t-test; $p < 0.005$) such that incorrectly classified non-musicians were older than correctly classified non-musicians. For the imaging data, there was a significant difference in mean cortical thickness between the incorrectly and correctly classified non-musicians (Student's t-test; $p < 0.001$). The same effect was observed for musicians (Student's t-test; $p < 0.005$). Given that cortical thickness was smaller in musicians than in non-musicians in the frequently selected regions of the machine learning pipeline (see section 3.5), and that cortical thickness decreased with age (see section 3.1), this suggests that sometimes the classifier confused the effects of musical training and aging because cortical thinning is associated with both of these two processes in these data.

3.4. Assessment of over-fitting with permutation tests

The classifier training and testing was repeated 1000 times with randomly permuted group labels. The mean inter-rater agreement was $\kappa_\mu = -0.002$ and the standard deviation was 0.124. The 95 % confidence interval for the mean inter-rater agreement was $[-0.010, 0.006]$. The permutation distribution was approximately normally distributed. The probability of observing a Cohen's kappa equal to or larger than $\kappa = 0.321$ by chance was $p = 0.006$. The permutation test results are illustrated as a histogram in Figure 4. The results suggest that the actual classifier was reliably fit to the data and not prone to over-fitting, since it did not perform well with the randomly permuted group labels.

3.5. Variable selection

The variable selection results of the linear SVM are summarized to Table III and illustrated in Figure 5. The results were fairly stable, meaning that similar decisions were made on each of the ten cross-validation iterations. There was a total of thirteen regions, nine in the left hemisphere and four in the right hemisphere, which were selected on every iteration. Five more regions, all located in the left hemisphere, were selected on more than half of the iterations. In total, 39 different regions were selected during the ten cross-validation iterations. Because it is difficult to assess the importance or meaning of those regions that were selected infrequently, we chose to analyze further only those 18 regions that were selected on more than half of the iterations.

The frequently selected regions were located in the frontal, parietal, and occipital lobes. Since these feature selection results themselves did not contain information about the direction or magnitude of effects, we compared the cortical thickness in the frequently selected regions between musicians and non-musicians as a post-hoc comparison in order to understand the effects of any underlying musical ability better. The results are presented in Table III. The cortical thickness of every frequently selected region was smaller in musicians than in non-musicians. To have a reference for the effect sizes in these regions, one can compare the rate of cortical thinning with aging to the differences observed between musicians and non-musicians. The rate of cortical thinning was 0.005 mm per year on average in our data ($p < 0.001$, see results in Section 3.1). Therefore, the differences between the musicians' and non-musicians' brains were approximately 10–20 times larger than yearly cortical thinning.

4. Discussion

Our results demonstrate that decoding musicianship from brain MRIs is feasible. The linear SVM achieved a significant inter-rater agreement between the actual and predicted participant groups. In other words, it was able to learn a function from labeled MRI data that could be used to classify new previously unseen unlabeled MRI data. It correctly labeled 63.3 % of the musician and 72.9 % of the non-musician participants.

Since playing music is a combined multi-sensory and motor skill, we expected that the relevant regions for classifying the participants would be widely distributed across the cortex. Indeed, the classifier's ability to distinguish musicians from non-musicians was based on differences in cortical thickness in several regions located in the frontal, parietal, and occipital lobes with a greater emphasis on the left compared to the right hemisphere. This is in general agreement with the previous literature, which demonstrates that several areas in frontal, parietal and occipital lobes are linked to musical ability.

In line with our results, James and colleagues (2012) found a decreased gray matter density in amateur and professional musicians compared to non-musician participants in the right postcentral gyrus, bilateral paracentral lobule, bilateral precuneus, left inferior occipital gyrus, and bilateral striatal areas. Our results are in line with these findings especially for the medial surface regions, however suggesting more emphasized role for the left than right hemisphere. In our study, the thickness of the left precuneus and the left paracentral lobule were observed to be relevant predictors of musical ability. James and colleagues (2012)

interpreted their findings to reflect automatized sensorimotor function and argued that fewer neurons are eventually needed for producing the playing-related movements, after a certain level of proficiency has been acquired, which could explain also our results.

Indeed, increases in proficiency during development are often associated with decreases in gray matter, and as musical skills are often acquired early in life, it would make sense that the areas that are engaged in increased amounts of sensorimotor training show stronger expertise-related pruning. The volume of left precuneus, which is a part of the superior parietal lobule, and left paracentral lobule were observed as relevant predictors of musical ability both by James and co-authors and the current study, highlighting the role of these regions in acquired musical aptitude. James and colleagues (2012) found also regions with increased gray-matter densities but they did not overlap with the regions we observed to be important. Achieving detailed understanding on the differences between expertise-related decrease vs. increase in gray matter thickness and volume would require analysis at the level of underlying neuronal and vascular changes. It may well be, that the skill acquisition during development builds on very different neuroplastic processes which are integrated with maturational changes, than the skill acquisition later in life (Steele et al., 2012; Vaquero et al., 2015; de Manzano and Ullén, 2017; Weisberg et al., 2019).

Further, while many studies that investigated skill- and experience-related changes in the human brain structure report increases in volume, area, or thickness, the large amounts of knowledge, experience, and skills obtained during a lifetime are ultimately not likely to be represented in the brain by cumulative, continual increases in structural measures (Wenger et al., 2017; Calmels et al., 2019). Indeed, selection, deletion and pruning are effective mechanisms that occur throughout the nature, including brain development. The expansion-renormalization model posits that learning is a two-step process, in which the first step involves an expansion of gray matter structure, followed by selection and pruning in the second step, thus keeping the overall brain volume approximately constant (Wenger et al., 2017). Since musical expertise is acquired over very long time scales, these data could be explained in the context of the expansion–renormalization model by a gradually accumulated negative net effect, which is either due to over-pruning over time, declined expansion volumes over time, or both. These processes might be modulated by the already acquired experience or aging.

Gray-matter differences in the frontal lobe between musicians and non-musicians have also been reported in previous experiments. Gaser & Schlaug (2003a) found an increased gray-matter volume in the left inferior frontal gyrus in musicians. Another study (Bermudez et al. 2009) observed greater cortical thickness bilaterally in the middle frontal gyrus, and the left superior frontal gyrus. Our results are partially in conflict with these observations, as we detected greater cortical thickness in the left middle frontal gyrus and sulcus in non-musicians than musicians. Unfortunately, it is difficult to point out what could be the reason for the disagreement due to differences in participant recruitment and data analysis methods. While Bermudez and colleagues performed voxel-level analyses based on the general linear model, we used supervised machine learning and analyzed the data using a predefined cortical parcellation.

Our findings on the significance of occipital areas for classifying musicians vs. non-musicians may seem at first contradictory, as playing an instrument is strongly linked with auditory-motor -integration. There are however also earlier studies showing this association. Luders and colleagues (2004) identified structural changes in the occipital lobe related to musical ability. They found gray matter asymmetries in musicians in the cuneus and the medial occipital gyrus. Along the same lines, Bengtsson and co-workers (2005) showed that practicing playing music during adolescence correlated with fractional anisotropy in the splenium and body of the corpus callosum extending into the white matter of the occipital lobe. Foster and Zatorre (2010) found a positive correlation between a relative pitch task and cortical thickness in the left precuneus, left inferior parietal lobule, bilateral parieto-occipital sulcus, left middle frontal gyrus, left precentral sulcus, right precentral gyrus, and left ventrolateral frontal cortex. A recent fMRI study also indicated large increases in brain-scale networks connected to music-selective visual areas (Mongelli et al., 2017).

The observed cortical differences between musicians and non-musicians in the occipital lobe could be related to an audiovisual or visuo-motor integration process, that presumably forms an essential component in learning to fluently translate visual notes to auditory and motor representations: Interestingly, modulating attentional demands for auditory stimuli has been shown to activate the peripheral visual cortex not only in blind subjects but also in normally sighted adults (Cate et al. 2009). The cross-modal sensory interactions might thus be more readily available for relevant behavioral needs than is typically assumed (Shimojo & Shams, 2001). Music training might well be one of the skills that readily maintains integration between auditory, visual, and motor cortices, at the early level of cortical hierarchy. Indeed, previous studies have found musicians to score better than non-musicians in visual attention tests (Rodrigues et al. 2013), as well as in audiovisual timing tests where musicians were better in judging whether auditory and visual stimuli were presented synchronously (Lu et al. 2014). Moreover, an EEG study found that musicians are able to expect auditory endings based on visually-presented musical sequences (Schön and Besson, 2003).

As with any comparison between groups differing in the level of expertise, it is difficult, if not impossible, to tease apart the influence of expertise and innate differences in brain structural properties. Most of the previous studies investigating the neural basis of musical expertise have relied on cross-sectional experimental designs. A potential problem shared by such studies, including the present one, is that they are able to rule out only certain confounding effects. Therefore, it is interesting to compare the results on relevant cortical areas for musical expertise between cross-sectional and longitudinal or intervention studies. For example, Hyde and colleagues (2009) studied the effect of fifteen months of musical training in musically untrained children, compared to control children receiving no training. The children who received the musical training showed a decrease in the relative volume of the left middle occipital gyrus, which was a relevant predictor of musical ability in our study, too. These common findings give rather strong evidence of the importance of this region in acquiring musical aptitude. Interestingly, learning to play a musical instrument is a task that first relies much on visual feedback for achieving the desired motor function (Groussard et al. 2014). Early sensory areas show also the fastest developmental trajectory. Therefore, it is not surprising that the region that was found relevant for musical aptitude in both studies was located in the occipital lobe.

Hyde and colleagues (2009) also found increased relative volumes in frontal and temporal lobes but we did not observe such effects. The disagreement could be related to the participants' age, since Hyde and colleagues studied young children whereas we investigated adults. Indeed, the cortical gray matter shows considerable changes, both increase and decrease, in thickness and volume during protracted period of development (Toga et al, 2006). These developmental processes continue well into adolescence and even adulthood for some brain regions, and reflect the combined influence of genetic and environmental factors. The discrepancy might also stem from the very different duration of received musical training; our musician participants had practiced their instrument thirteen years on average.

To make the classifier generalize optimally, we investigated the effects of age and gender on the participants' mean cortical thickness. Since an analysis of the entire sample of musicians and non-musicians before training the classifier would lead to over-fitting, the effects of age and gender were studied without knowledge of participant groups. The mean cortical thickness decreased with age, which replicated previously reported results (Salat et al. 2004; Thambisetty et al. 2010; Toga et al. 2011). While some studies have also reported cortical differences between males' and females' brains (Im et al. 2006; Sowell et al. 2007), we did not observe such an effect on the level of the whole cortex. Therefore, the participants' group and age – but not gender – were used in balancing the classifier's cross-validation folds. It should be noted that since age and the number of years of experience in playing music are naturally correlated in professional musicians, a limitation in this study is that it is not possible to disentangle all interactions between development, learning, and aging.

Our feature-selection results indicate that the left-hemispheric regions were more informative than the right-hemispheric regions for classifying the participants as either musicians or non-musicians; only four reliable regions got selected from the right hemisphere, and as many as 14 from the left hemisphere (see

Table III). Thus, future studies could benefit from expanding the feature set by calculating suitable inter-hemispheric lateralization scores, since differences in hemispheric asymmetry between musicians and non-musicians have been reported in previous studies (Schneider et al. 2005; Ellis et al. 2013; Burunat et al. 2015). However, this imbalance could be alternatively partially related to sampling; most of our participants were right-handed.

Remarkably, the temporal regions of the Destrieux atlas were not considered informative by our feature-selection procedure, which was based on cross-validation and the Student's t-test. However, this may be simply due to the relatively large sizes of the Destrieux atlases temporal regions resulting in the loss of information regarding musical ability.Given the availability of many different cortical parcellations (Fischl et al., 2004; Fan et al., 2016; Glasser et al., 2016; Gordon et al., 2016; Schaefer et al., 2018) which are based on different delineation methods and region sizes, alternative choices can be tested in future studies. It is also possible that cortical area or volume would reflect training-related changes better than cortical thickness, which can be also tested in subsequent studies. Indeed, cortical thickness and area are known to be influenced by different sets of genetic factors and neuroplastic processes (Panizzon et al., 2009; Raznahan et al., 2011; Walhovd et al., 2016). In addition, thickness and area also have different developmental trajectories and are generally differentially associated with cognitive abilities (Vuoksimaa et al., 2016; Winkler et al., 2017). Taking the opposite view, the present results allow speculation on regions that are critical for identifying professional musicians. It may be that fine-tuned auditory sensory function is not sufficient for separating musicians from non-musicians. In fact, sensory expertise might be generally much more balanced over different sub-populations than motor expertise. Professional musicians must master fine motor movements and repertoires and be efficient in integrating visuo-motor functions (and memory) with auditory perception, which might cause more unique structural patterns that become represented in musicians' brains. Thus, the brain networks related to actions performed by musicians might be much more discriminative than the brain networks related to perception.

On a technical and more general level, it is of importance also to consider the advantages and disadvantages of using a brain-atlas-based analysis approach in the first place. On one hand, a voxel-level analysis would allow detecting very focal effects. However, as the number of voxels would be orders of magnitude larger than the number of brain regions in a typical atlas, much more emphasis would be required on handling over-fitting and multiple testing in particular, which in turn would usually imply a need for obtaining a larger sample of participants. On the other hand, brain atlases allow making neuroanatomical definitions, communicating findings easily, and facilitate comparisons across studies. Ideally, the regions of a brain atlas also delineate structures with specific functions, allowing efficient data interpretation. Disadvantages of using brain atlases have been recently reviewed by Dickie and colleagues (2017) and include, for example, experimental effects being localized between regions and risk of systematic biases. Finally, it should be noted that voxel-based analyses typically require the data to be smoothed substantially, which can make the actual effective resolution close to using a very fine-grained brain atlas. From this perspective, these two approaches can be fairly similar with particular choices of parameters.

In this study, we chose to decode musicianship from MRIs. Since this was shown to be feasible, as a next step it would be interesting to decode more specific musical skills. For example, it might be possible to decode musicians' main instruments using a multi-class classification method. In such approach, several one-vs-one or one-vs-many SVMs could be used (Hsu & Lin, 2002). Here, while our sample did include musicians with different main instruments, the sample size of thirty professional musicians would have been too small to allow analysis of sub-groups. Another way to make the classification problem more fine-grained could be to predict a continuous value such as years of practice instead of musicianship. Including both structural and functional features in such decoding analyses (see e.g. Albouy et al., 2019) would be particularly interesting, since it would become possible to assess whether particular structural differences are associated with corresponding functional differences and how such relationships change as a function of (or type of) accumulated experience.

Overall, our data are consistent with the previous studies that suggest professional musicians' brains to be structurally different compared to non-musicians' brains. Here we demonstrated that using SVMs, such differences can be detected from MRIs on the level of individual participants. While the most plausible explanation is that the observed differences are caused by the decades-lasting intensive musical training, it is impossible to make a definite statement since the musicians' brains could be different to begin with; Merrett and colleagues (2013) have recently reviewed possible confounding variables such as genetics, personality, and early auditory environment. Similarly, we are not able to rule out transfer effects from other multi-sensory and motor skills associated with musical ability. In future studies, it should be considered if behavioral data could be included as covariates (e.g. data from an auditory or motor task in which musicians could be assumed to perform better). Studies investigating training- or learning-induced structural changes in the brain have also been criticized for small effect sizes compared to imaging accuracy (Thomas & Baker, 2013). However, it should be acknowledged that not all structural changes are necessarily created equal: for a given behavioral function small structural differences could sometimes be more relevant than larger ones. In our study, the relatively large sizes of the analyzed regions emphasize differences in cortical thickness that occur consistently over the defined regions; very focal changes could become insignificant in the average that is computed from each region. Further, the novelty in our approach is that in addition to finding which brain regions are different between musicians and non-musicians, we also obtain an estimate of how important those regions are for decoding musicianship.

Future studies should explore ways of improving the classification accuracy. The performance of SVMs could be compared to other available classification methods such as random forests or neural networks. Different feature selection methods could be also tested. In particular, it would be interesting to contrast filter and wrapper type methods (Weston et al., 2000). Increasing the sample size or MRI resolution (e.g. performing imaging at 7 T) could also improve the results.

5. Conclusions

In this study, we have demonstrated that it is feasible to decode musicianship from structural MRIs of the brain. Using well-known and previously extensively validated machine learning and MRI processing techniques, we analyzed a fairly large sample of 30 professional musician and 85 non-musician control participants. We were able to correctly categorize 63.3 % of the musician and 72.9 % of the non-musician participants based on cerebral cortical thickness. The ability to distinguish the musicians from the non-musicians was based on the thickness of several regions located mostly in the frontal, parietal, and occipital lobes of the left hemisphere. While several previous studies have compared structural MRIs of musicians' and non-musicians' brains, it has remained unclear how relevant the revealed differences are for musical aptitude. In addition to simply identifying differences in the brains of musicians and non-musicians, we could estimate how important they are in predicting a person's musical ability. Overall, our results suggest that predicting musicianship from brain structure is a challenging task but solvable to a certain degree.

Conflicts of interest

The authors declare no competing financial or other interests.

Acknowledgements

This research was financially supported by TEKES of Finland under the grant 40334/10 "Machine learning for future music and learning technologies". We thank IT Center for Science Ltd, administered by the Finnish Ministry of Education and Culture, for providing data storage and computing resources. We also

thank Ilkka Pölönen and Jarmo Hämäläinen for helpful discussions, Jussi Numminen for performing a medical screening of the collected MRIs, and Jyrki Mäkelä for contributing to initial stages of the project and for supervision of participant recruitment. We thank Marita Kattelus and Toni Auranen for helping in the data acquisition. A. K. Nandi thanks TEKES for the award of Finland Distinguished Professorship.

References

[1] Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G (2014): Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics* 8:14. DOI: 10.3389/fninf.2014.00014

[2] Albouy P, Caclin A, Norman-Haignere SV, Lévêque Y, Peretz I, Tillmann B, Zatorre RJ (2019): Decoding task-related functional brain imaging data to identify developmental disorders: the case of congenital amusia. *Frontiers in Neuroscience* 13:1165. DOI: 10.3389/fnins.2019.01165

[3] Barrett KC, Ashley R, Strait DL, Kraus N (2013): Art and science: How musical training shapes the brain. *Frontiers in Psychology* 4:713. DOI: 10.3389/fpsyg.2013.00713

[4] Bengtsson SL, Nagy Z, Skare S, Forsman L, Forssberg H, Ullén F (2005): Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience* 8(9):1148–1150. DOI: 10.1038/nn1516

[5] Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009): Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex* 19(7):1583–1596. DOI: 10.1093/cercor/bhn196

[6] Burunat I, Brattico E, Puoliväli T, Ristaniemi M, Sams M, Toivianen P (2015): Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE* 10(9):e0138238. DOI: 10.1371/journal.pone.0138238

[7] Calmels C (2019): Neural correlates of motor expertise: Extensive motor training and cortical changes. *Brain Research* 146323. DOI: 10.1016/j.brainres.2019.146323

[8] Cardinale F, Chinnici G, Bramerio M, Mai R, Sartori I, Cossu M, Lo Russo G, Castana L, Colombo N, Caborni C, De Momi E, Ferrigno G (2014): Validation of FreeSurfer-estimated brain cortical thickness: Comparison with histologic measurements. *Neuroinformatics* 12(4): 535–542. DOI: 10.1007/s12021-014-9229-2

[9] Cate AD, Herron TJ, Yund EW, Stecker GC, Rinne T, Kang X, Petkov CI, Disbrow EA, Woods DL (2009): Auditory attention activates peripheral visual cortex. *PLoS ONE* 4(2):e4645. DOI: 10.1371/journal.pone.0004645

[10] Chang CC, Lin CJ (2011): LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3):27:1–27:27. DOI: 10.1145/1961189.1961199

[11] Cichy RM, Pantazis D, Oliva A (2014): Resolving human object recognition in space and time. *Nature Neuroscience* 17:455–462. DOI: 10.1038/nn.3635

[12] Cortes C, Vapnik V (1995): Support-vector networks. *Machine Learning* 20:273-297. DOI: 10.1007/BF00994018

[13] Dale AM, Fischl B, Serono MI (1999): Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage* 9(2):179–194. DOI: 10.1006/nimg.1998.0395

[14] De Manzano Ö, Ullén F (2017): Same genes, different brains: Neuroanatomical differences between monozygotic twins discordant for musical training. *Cerebral Cortex* 28:387–394. DOI: 10.1093/cercor/bhx299

[15] Destrieux C, Fischl B, Dale A, Halgren E (2010): Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53(1):1–15. DOI: 10.1016/j.neuroimage.2010.06.010

[16] Dickie DA, Shenkin SD, Anblagan D, Lee J, Cabez MB, Rodriguez D, Boardman JP, Waldman A, Job DE, Wardlaw JM (2017): Whole brain magnetic resonance image atlases: a systematic review of existing atlases and caveats for use in population imaging. *Frontiers in Neuroinfomatics* 11:1. DOI: 10.3389/fninf.2017.00001

[17] Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1995): Increased cortical representation of the fingers of the left hand in string players. *Science* 270:305–307

[18] Ellis RJ, Bruijn B, Norton AC, Winner E, Schlaug G (2013): Training-mediated leftward asymmetries during music processing: a cross-sectional and longitudinal fMRI analysis. *NeuroImage* 75:97–107. DOI: 10.1016/j.neuroimage.2013.02.045

[19] Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird A, Fox P, Eickhoff S, Yu C, Jiang T (2016): The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral Cortex* 26:3508–3526

[20] Fernández-Delgado M, Cernadas E, Barro S (2014): Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research* 15:3133–3181.

[21] Fischl B, Dale AM (2000): Measuring the thickness of human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* 97(20):11050–11055. DOI: 10.1073/pnas.200033797

[22] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Kouwe A van der, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002): Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3):341–355. DOI: 10.1016/S0896-6273(02)00569-X

[23] Fischl B, Serono IM, Dale AM (1999): Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9(2):195–207. DOI: 10.1006/nimg.1998.0396

[24] Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004): Automatically parcellating the human cerebral cortex. *Cerebral Cortex* 14:11–22. DOI: 10.1093/cercor/bhg087

[25] Foster NVE, Zatorre RJ (2009): A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex* 20:1350–1359. DOI: 10.1093/cercor/bhp199

[26] Foster NEV, Zatorre RJ (2010): Cortical structure predicts success in performing musical transformation judgments. *NeuroImage* 53:23–36. DOI: 10.1016/j.neuroimage.2010.06.042

[27] Gaser C, Schlaug G (2003a): Gray matter differences between musicians and non-musicians. *Annals of the New York Academy of Sciences* 999:514–517. DOI: 10.1196/annals.1284.062

[28] Gaser C, Schlaug G (2003b): Brain structures differ between musicians and non-musicians. *Journal of Neuroscience* 23(27):9240–9245. DOI: 10.1523/jneurosci.23-27-09240.2003

[29] Glasser M, Coalson T, Robinson E, Hacker C, Yacoub JHE, Ugurbil K, Andersson J, Beckmann C, Jenkinson M, Smith S, van Essen D (2016): A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615):171–178

[30] Gordon E, Laumann T, Adeyemo B, Huckins J, Kelley W, Petersen S (2016): Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex* 26:288–303

[31] Groussard M, Viader F, Landeau B, Desgranges B, Eustache F, Platel H (2014): The effects of musical practice on structural plasticity: The dynamics of grey matter changes. *Brain and Cognition* 90:174–180. DOI: 10.1016/j.bandc.2014.06.013

[32] Gyuon I, Elisseef A (2003): An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157–1182.

[33] Hajek T, Cooke C, Kopecek M, Novak T, Hoschl C, Alda M (2015): Using structural MRI to identify individuals at genetic risk for bipolar disorders: A 2-cohort, machine learning study. *Journal of Psychiatry and Neuroscience* 40(5):316–324. DOI: 10.1503/jpn.140142

[34] Hampton AN, O'Doherty JP (2007): Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America* 104(4):1377–1382. DOI: 10.1073/pnas.0606297104

[35] Han X, Jovicich J, D Salat, Kouwe A van der, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B (2006): Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage* 32:180–194. DOI: 10.1016/j.neuroimage.2006.02.051

[36] Haynes JD, Rees G (2006): Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7(7):523–534. DOI: 10.1038/nrn1931

[37] Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013): Neural decoding of visual imagery during sleep. *Science* 340(6132):639–642. DOI: 10.1126/science.1234330

[38] Hsu C-W, C-J (2002): A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* 13(2):415–425.

[39] Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, Sclaug G (2009): Musical training shapes structural brain development. *Journal of Neuroscience* 29(10):3019–3025. DOI: 10.1523/JNEUROSCI.5118-08.2009

[40] Im K, Lee J-M, Lee J, Shin Y-W, Kim IY, Kwon JS, Kim SI (2006): Gender difference analysis of cortical thickness in healthy adults with surface-based methods. *NeuroImage* 31(1):31–38. DOI: 10.1016/j.neuroimage.2005.11.042

[41] James CE, Oechslin MS, Ville van de D, Hauert CA, Descloux C, Lazeyras F (2012): Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. *Brain Structure and Function* 219(1):353–366. DOI: 10.1007/s00429-013-0504-z

[42] Kohavi R (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artifical Intelligence*. P 1137–1143.

[43] Koster-Hale J, Saxe R, Dungan J, Young LL (2013): Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America* 110(14): 5648–5653. DOI: 10.1073/pnas.1207992110

[44] Landis JR, Koch GG (1977): The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. DOI: 10.2307/2529310

[45] Lu Y, Paraskevopoulos E, Herholz SC, Kuchenbuch A, Pantev C (2014): Temporal processing of audiovisual stimuli is enhanced in musicians: evidence from magnetoencephalography (MEG). *PLoS ONE* 9(3):e90686. DOI: 10.1371/journal.pone.0090686

[46] Luders E, Gaser C, Jancke L, Sclaug G (2004): A voxel-based approach to gray matter asymmetries. *NeuroImage* 22(2):656–664. DOI: 10.1016/j.neuroimage.2004.01.032

[47] Maguire EA, Woollett K, Spiers HJ (2006): London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus* 16(12):1091–1101. DOI: 10.1002/hipo.20233

[48] Mechelli A, Crinion JT, Noppeney U, O'Doherty J, Ashburner J, Frackowiak RS, Price CJ (2004): Neurolinguistics: Structural plasticity in the bilingual brain. *Nature* 431:757.

[49] Merrett DL, Peretz I, Wilson SJ (2013): Moderating variables of music training-induced neuroplasticity: A review and discussion. *Frontiers in Psychology* 4:606. DOI: 10.3389/fpsyg.2013.00606

[50] Miendlarzewska EA, Wiebke JT (2013): How musical training affects cognitive development: rhythm, reward and other modulating variables. *Frontiers in Neuroscience* 7:279. DOI: 10.3389/fnins.2013.00279

[51] Mongelli V, Dehaene S, Vinckier F, Peretz I, Bartolomeo P, Cohen L (2017): Music and words in the visual cortex: The impact of musical expertise. *Cortex* 86:260–274. DOI: 10.1016/.j.cortex.2016.05.016

[52] Moore E, Schaefer RS, Bastin ME, Roberts N, Overy K (2014): Can musical training influence brain connectivity? Evidence from diffusion tensor MRI. *Brain Sciences* 4(2):405–427. DOI: 10.3390/brainsci4020405

[53] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015): Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104:398–412. DOI: 10.1016/j.neuroimage.2014.10.002

[54] Münte TF, Altenmüller E, Jäncke L (2002): The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience* 3:473–478

[55] Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011): Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology* 21:1641–1646. DOI: 10.1016/j.cub.2011.08.031

[56] Panizzon MS, Fennema-Notestine C, Eyler LT, Jernigan TL, Prom-Wormley E, Neale M, Jacobson K, Lyons MJ, Grant MD, Franz CE, Xian H, Tsuang M, Fischl B, Seidman L, Dale A, Kremen WS (2009): Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex* 19:2782–2735. DOI: 10.1093/cercor/bhp026

[57] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011): Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

[58] Pletikos M, Sousa AMM, Sedmak G, Meyer KA, Zhu Y, Cheng F, Li M, Kawasawa YI, Sestan N (2014): Temporal specification and bilaterality of human neocortical topographical gene expression. *Neuron* 81:321–332. DOI: 10.1016/j.neuron.2013.11.018

[59] Raznahan A, Shaw P, Lalonde F, Stockman M, Wallace GL, Greenstein D, Clasen L, Gogtay N, Giedd JN (2011): How does your cortex grow? *The Journal of Neuroscience* 31(19):7174–7177. DOI: 10.1523/jneurosci.0054-11.2011

[60] Ressel V, Pallier C, Ventura-Campos N, Diaz B, Roessler A, Ávila C, Sebastián-Gallés N (2012): An effect of bilingualism on the auditory cortex. *Journal of Neuroscience* 32(47):16597–16601. DOI: 10.1523/JNEUROSCI.1996-12.2012

[61] Reybrouck M, Brattico E (2015): Neuroplasticity beyond sounds: Neural adaptations following long-term musical aesthetic experiences. *Brain Sciences* 5(1):69–91.
DOI: 10.3390/brainsci5010069

[62] Rodrigues AC, Loureiro MA, Caramelli P (2013): Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition* 82(3):229–235. DOI: 10.1016/j.bandc.2013.04.009

[63] Saari P, Burunat I, Brattico E, Toiviainen P (2018): Decoding musical training from dynamic processing of musical features in the brain. *Scientific Reports* 8:708. DOI: 10.1038/s41598-018-19177-5
[64] Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RSR, Busa E, Morris JC, Dale AM, Fischl B (2004): Thinning of the cerebral cortex in aging. *Cerebral Cortex* 14(7):721–730. DOI: 10.1093/

cercor/bhh032

[65] Salti M, Monto S, Charles L, King JR, Parkkonen L, Dehaene S (2015): Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife* e05652. DOI: 10.7554/eLife.05652

[66] Sampaio-Baptista C, Scholz J, Jenkinson M, Thomas AG, Filippini N, Smit G, Douaud G, Johansen-Berg (2014): Gray matter volume is associated with rate of subsequent skill learning after a long term training intervention. *NeuroImage* 96:158–166. DOI: 10.1016/j.neuroimage.2014.03.056

[67] Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BTT (2018): Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex* 28:3095–3114

[68] Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A (2005): Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience* 8(9):1241–1247. DOI: 10.1038/nn1530

[69] Scholz J, Klein MC, Behrens TE, Johansen-Berg H (2009): Training induces changes in white-matter architecture. *Nature Neuroscience* 12(11):1370–1371. DOI: 10.1038/nn.2412

[70] Schön D, Besson M (2003): Audiovisual interactions in music reading. A reaction times and event-related potentials study. *Annals of the New York Academy of Sciences* 999:193–198. DOI: 10.1196/annals.1284.028

[71] Shimojo S, Shams L (2001): Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology* 11:505–509. DOI: 10.1016/S0959-4388(00)00241-5

[72] Sowell ER, Peterson BS, Kan E, Woods RP, Yoshii J, Bansal R, Xu D, Zhu H, Thompson PM, Toga AW (2007): Sex differences in cortical thickness mapped in 176 healthy individuals between 8 and 87 years of age. *Cerebral Cortex* 17(7):1550–1560. DOI: 10.1093/cercor/bhl066

[73] Steele CJ, Scholtz J, Douaud G, Johansen-Berg H, Penhune VB (2012): Structural correlates of skilled performance on a motor sequence task. *Frontiers in Human Neuroscience* 6:289. DOI: 10.3389/fnhum.2012.00289

[74] Thambisetty M, Wan J, Carass A, An Y, Pierce JL, Resnick SM (2010): Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage* 52(4):1215–1223. DOI: 10.1016/j.neuroimage.2010.04.258

[75] Thomas C, Baker CI (2013): Teaching an adult brain new tricks: A critical review of evidence for training-dependent structural plasticity in humans. *NeuroImage* 73:225–236. DOI: 10.1016/j.neuroimage.2012.03.069

[76] Toga AW, Thompson PM, Sowell ER (2011): Mapping brain maturation. *Trends in Neurosciences* 29(3):148–159. DOI: 10.1016/j.tins.2006.01.007

[77] Vaquero L, Hartmann K, Ripollés P, Rojo N, Sierpowska J, Francois C, Càmara E, Tijmen van Vugt F, Mohammadi B, Samii A, Münte TF, Rodríguez-Fornells A, Altenmüller E (2015): Structural neuro-

plasticity in expert pianists depends on the age of musical training onset. *NeuroImage* 26:106–119. DOI: 10.1016/j.neuroimage.2015.11.008

[78] Vuoksimaa E, Panizzon MS, Chen C-H, Fiecas M, Eyler LT, Fennema-Notestine C, Hagler DJ Jr, Franz CE, Jak AJ, Lyons MJ, Neale MC, Rinker DA, Thompson WK, Tsuang MT, Dale AM, Kremen WS (2016): Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *NeuroImage* 129:356–366. DOI: 10.1016/j.neuroimage.2016.01.049

[79] Waldert S, Preissl H, Demandt E, Braun C, Birbaumer N, Aertsen A, Mehring C (2008): Hand movement direction decoded from MEG and EEG. *Journal of Neuroscience* 28(4):1000–1008. DOI: 10.1523/JNEUROSCI.5171-07.2008

[80] Walhovd KB, Fjell AM, Giedd J, Dale AM, Brown TT (2016): Through thick and thin: a need to reconcile contradictory results on trajectories in human cortical development. *Cerebral Cortex* 27(2):1472–1481. DOI: 10.1093/cercor/bhv301

[81] Weisberg SM, Newcombe NS, Chatterjee A (2019): Everyday taxi drivers: Do better navigators have larger hippocampi? *Cortex* 115:280–293. DOI: 10.1016/j.cortex.2018.12.024

[82] Wenger E, Brozzoli C, Lindenberger U, Lövdén M (2017): Expansion & renormalization of human brain structure during skill acquisition. *Trends in Cognitive Sciences* 21(12):930–939. DOI: 10.1016/j.tics.2017.09.008

[83] Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V (200): Feature selection for SVMs. In: *Advances in Neural Information Processing Systems* 13. MIT Press. P 668–674.

[84] Winkler AM, Greve DN, Bjuland KJ, Nichols TE, Sabuncu MR, Håberg AK, Skranes J, Rimol LM (2018): Joint analysis of cortical area and thickness as a replacement for the analysis of the volume of the cerebral cortex. *Cerebral Cortex* 28(2):738–749. DOI: 10.1093/cercor/bhx308

[85] Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasti P, Locascio JJ, Bakkour A, Kouwe van der A, Triantafyllou C, Corkin S, Dickerson BC (2009): Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence, voxel geometry, and parallel imaging. *NeuroImage* 44(4):1324–1333. DOI: 10.1016/j.neuroimage.2008.10.037

[86] Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007): Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10:420–422. DOI: 10.1038/nn1872

[87] Zatorre RJ, Fields RD, Johansen-Berg H (2012): Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature Neuroscience* 15:528–536. DOI: 10.1038/nn.3045. DOI: 10.1038/nn.3045

Table 1: Participant demographics. Handedness missing for two participants.

Group	N	Age			Gender		Handedness			Years training	
		Mean	SD	Range	Male	Female	Left	Right	Ambidext.	Mean	SD
Non-mus	80	28.61	8.15	19–53	35	50	6	78	1	2.15	3.81
Musician	35	27.53	7.19	18–45	16	14	2	28		13.00	7.62
Total	115				51	64	8	106	1		

Table 2: SVM classification results. Sample size was $N = 115$ participants.

	Non-musician	Musician	Total
Non-musician	53.9 %	20 %	73.9 %
Musician	9.6 %	16.5 %	26.1 %
Total	63.5 %	36.5 %	100 %

Table 3: Variable selection results

Region	Lobe	#	Musicians	Non-mus	Diff
Left hemisphere					
Middle frontal gyrus	Frontal	10	2.57	2.63	-0.060
Middle frontal sulcus	Frontal	8	2.165	2.218	-0.053
Paracentral lobule and sulcus	Parietal	10	2.285	2.38	-0.095
Intraparietal sulcus and trans. parietal sulci	Parietal	10	2.118	2.2	-0.072
Angular gyrus	Parietal	9	2.608	2.679	-0.071
Superior parietal lobule	Parietal	7	2.369	2.424	-0.055
Posterior transverse collateral sulcus	Occipito-temporal	10	2.003	2.115	-0.112
Cuneus	Occipital	10	1.817	1.891	-0.074
Middle occipital gyrus	Occipital	10	2.498	2.592	-0.094
Superior occipital gyrus	Occipital	10	2.144	2.245	-0.101
Superior occ. sulcus and transv. occ. sulcus	Occipital	10	2.033	2.12	-0.087
Anterior occipital sulcus and preoccipital notch	Occipital	10	2.188	2.273	-0.085
Middle occipital sulcus and lunatus sulcus	Occipital	8	1.974	2.035	-0.061
Parieto-occipital sulcus	Occipital	9	2.189	2.262	-0.073
Right hemisphere					
Angular gyrus	Parietal	10	2.675	2.752	-0.077
Parieto-occipital sulcus	Parieto-occipital	10	2.21	2.299	-0.089
Medial occ.-temp. sulcus and lingual sulcus	Occipito-temporal	10	2.417	2.485	-0.065
Inferior occipital gyrus and sulcus	Occipital	10	2.506	2.613	-0.107

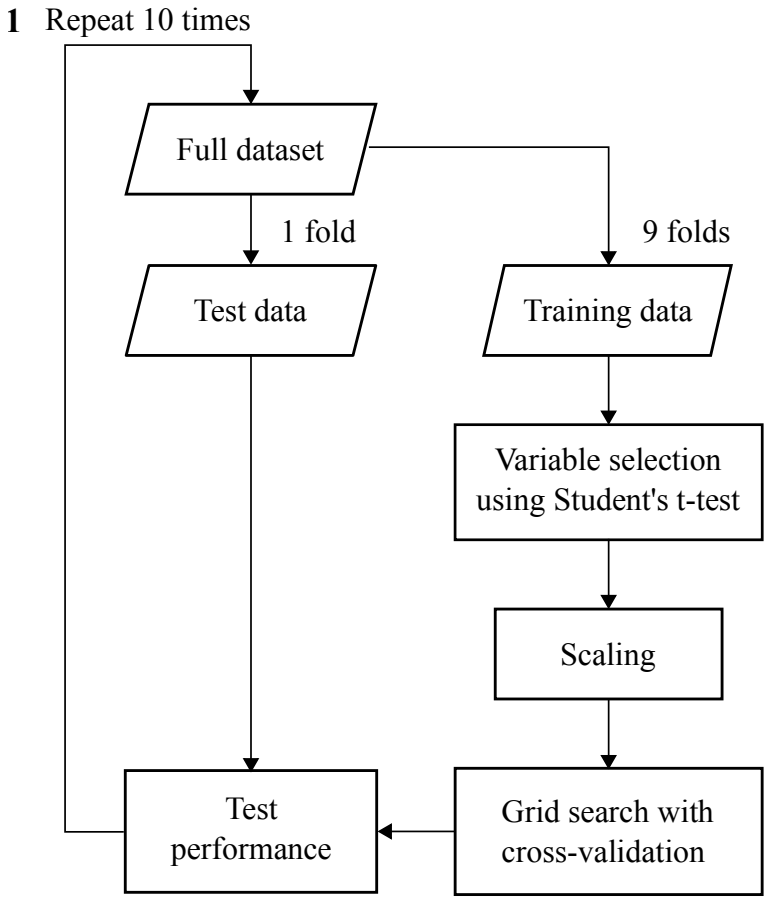

Figure 1: Nested cross-validation using ten-fold cross-validation as the outer loop. The model was constructed using 9 folds of the data and tested with the remaining one. The process was repeated 10 times, so that each fold acted as the test data once. The SVM parameters were selected using a grid search with a second cross-validation loop.

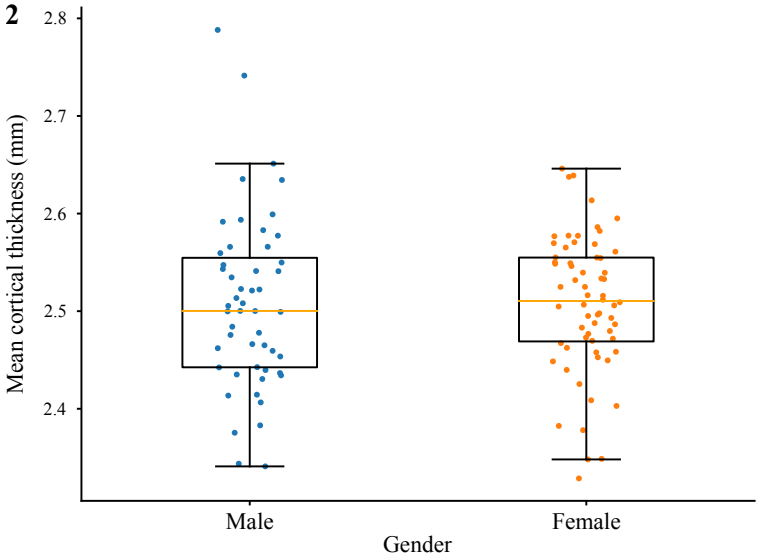

Figure 2: Mean cerebral cortical thickness (mm) as a function of age (years). Age explained 26 % of variation in the mean cortical thickness.

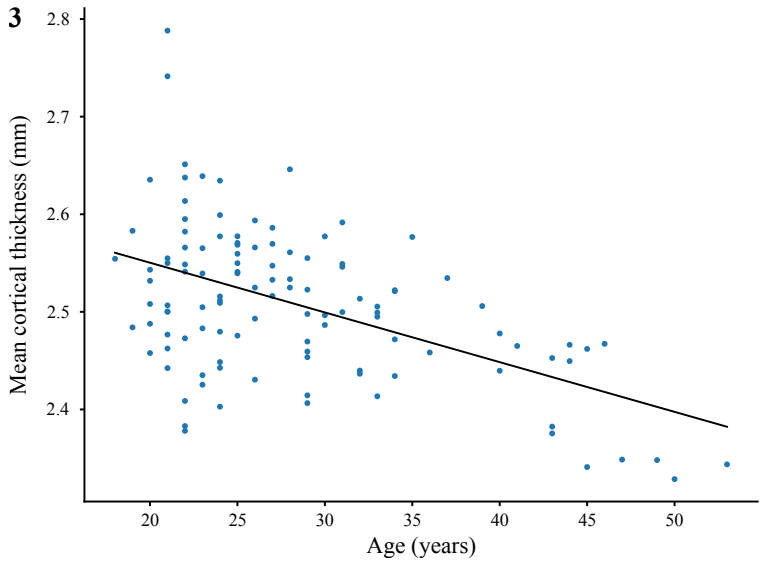

Figure 3: The mean cerebral cortical thickness was equal in males and females.

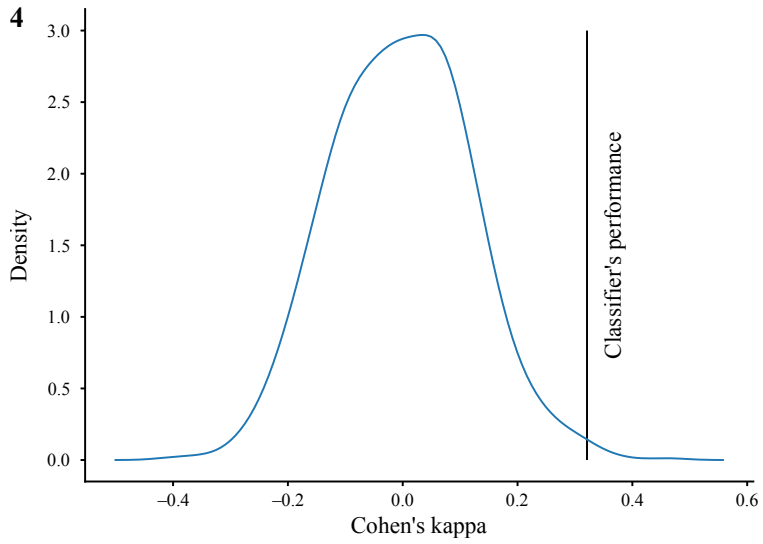

Figure 4: The permutation-test results over 1000 iterations. The density estimate shows the distribution of the Cohen's kappa statistic when the group labels "musician" and "non-musician" were randomly permuted and the machine learning repeated. The black vertical bar indicates the actual classifier performance.

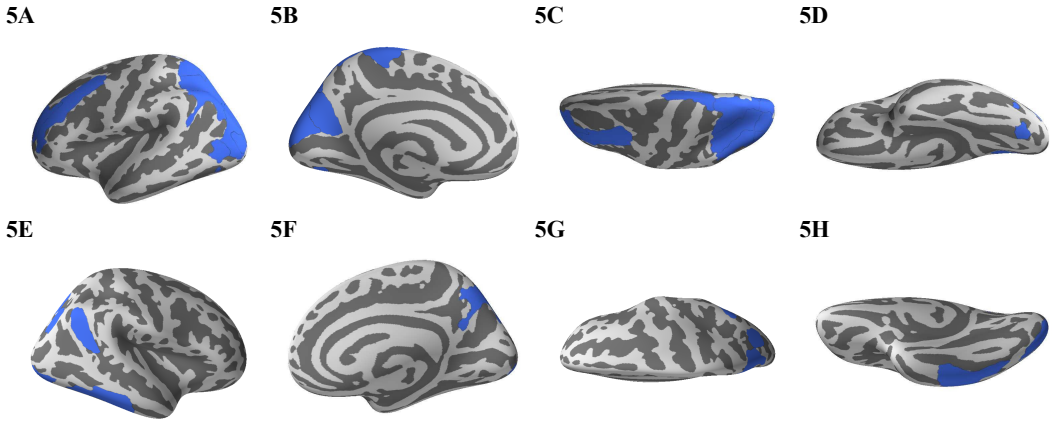

Figure 5: The feature selection results of the linear SVM. The regions that were selected on more than half of the cross-validation iterations are shown in blue. **A** lateral view of the left hemisphere; **B** medial view of the left hemisphere; **C** dorsal view of the left hemisphere; **D** ventral view of the left hemisphere; labels from **E** to **H** show the right hemisphere.