Neural Correlates of Tonality in Music

Simon Durrant
ICCMR
University of Plymouth
Plymouth, PL4 8AA, UK
simon.durrant@plymouth.ac.uk

David R. Hardoon
Dept. of Computer Science
University College London
London, WC1E 6BT, UK
D.Hardoon@cs.ucl.ac.uk

Eduardo R. Miranda
ICCMR
University of Plymouth
Plymouth, PL4 8AA, UK
eduardo.miranda@plymouth.ac.uk

John Shawe-Taylor
Dept. of Computer Science
University College London
London, WC1E 6BT, UK
jst@cs.ucl.ac.uk

André Brechmann
Speziallabor Nicht-Invasive Bildgebung
Leibniz Institute for Neurobiology
39118 Magdeburg, Germany
andre.brechmann@ifn-magdeburg.de

Henning Scheich
Leibniz Institute for Neurobiology
39118 Magdeburg, Germany
Henning.Scheich@ifn-magdeburg.de

Abstract

The paper presents fMRI results from experiments of subjects listening to musical stimuli. In this study we examine the neural correlates of tonality by presenting a set of stimuli with key changes of different distances along the circle-of-fifths, along with atonal control stimuli. Results are presented using both conventional statistical analysis across subjects, together with experiments using support vector machines. We find that a number of areas are significantly more active for tonal than atonal processing, and further that the response is significantly stronger in some of these areas for more distant key changes.

1 Introduction

Music has attracted a wide spread of research into how and why do people create, perform and listen to music. In the following paper we focus on the study of tonality, which is a major topic in music theory [1] and music psychology [2], investigating tonal processing, and in particular the effect of distance along the circle-of-fifths [3] for key changes within the stimuli. Tonal melodies are complex structures, and cognitive processing reflects this structure [4]. Analysing the neural activity associated with a melody is therefore particularly useful for determining which areas control higher order sequence processing [5].

The present study seeks to explore the effects of tonality in the brain. Previous work in the area of tonality [6] made strong claims about the possibility of identifying a tonal map in the rostromedial prefrontal cortex. These results have not been reproduced to our knowledge. We set ourselves a somewhat complementary goal of testing whether differences between tonal and atonal stimuli can be detected, as well as differences correlating with the distance along the circle-of-fifths as the stimuli change keys.

In addition to being a study on specific aspects of music, we are also interested in testing different forms of analysis. In general machine learning methods have been recently applied to fMRI analysis [7, 8, 9, 10, 11, 12, 13, 14, 15] to analyse the relationship between stimulus categories and fMRI responses. Higher order cognitive effects are known to be difficult to detect in fMRI due to the presence of confounds and the nonlinear mapping between cognition and BOLD response, so fMRI data from a music task that focuses on such effects provides a significant challenge to machine learning algorithms. As well as the conventional General Linear Model (GLM) analysis, therefore, we tested a machine learning approach (support vector machine) in the analysis of our data, in order to see whether or not it was able to find meaningful patterns in this data.

This paper is structured as follows; Section 2 discusses the material and methods used in our experiment. This contains an elaboration on the experiment protocol, as well as providing basic information as to the participating subjects and acquisition of data. Section 2 ends with a description of the pre-processing applied to the fMRI data. In Section 3, we give the results for the conventional (GLM) analysis and discuss the results of the machine learning (SVM) approach. Finally, Section 4 brings the paper to a close with a discussion on the work conducted and the analysis arising from it.

2 Materials and Methods

2.1 Experimental Design and Stimuli

The experiment was concerned with the tonality of short musical sequences. In particular, the focus of interest was the effect of relative tonality (the relationship between musical keys). Each stimulus consisted of 16 isochronous events lasting 500ms each (with each stimulus therefore lasting 8s), with no gaps in between; each event consisted of four simultaneous tones forming a chord recognised in Western tonal music theory. The stimuli were created using the MIDI protocol, and rendered into audio files using a piano sound patch from the Roland Sound Canvas© digital samples. The stimuli were divided between tonal stimuli, which were designed to create a clear sense of key, and atonal stimuli that were designed to create no clear sense of key, by the ordering of the chords, which were nevertheless equally consonant at the individual chord level in both types of stimuli. In order to verify this sense of key, the MIDI toolbox [16] was used to test the stimuli with the psychologically-derived Krumhansl-Kessler key-finding algorithm [17], with $\mu_T = 0.93931$ and $\sigma_T = 0.038562$ for the tonal stimuli, and $\mu_A = 0.65857$ and $\sigma_A = 0.13196$ for the atonal stimuli, where μ is the mean of the strongest key certainty ratings (ranging between 0 and 1), and σ is the standard deviation of these ratings. Altogether, eight different tonal stimuli were created, and twenty-four atonal stimuli. For a single run, stimuli were ordered into twenty-four groups of three stimuli with no gaps between stimuli or groups. The first stimulus in each group was always a tonal stimulus presented in the home key of C major, the second was always a tonal stimulus that could either be in the distant key of F# major (first condition), the closely-related key of G major (second condition), or the same key of C-major (third condition). The third stimulus in each group was always an atonal stimulus (fourth condition), which also reset the listener's sense of key.

As a result of the contiguity of groups, the first stimulus in each group followed the atonal stimulus in the previous group (except for the initial group), which was therefore defined as the initial (fifth) condition. The first and second conditions therefore define changes from one key to another (distant or close). The third condition defines no change of key. The fourth condition defines no key present, and the fifth condition defines a change of no key back to a sense of key. The stimuli were ordered such that all tonal stimuli were used an equal number of times, and conditions appeared in all permutations equally in order to control for order effects. The behavioural task for subjects was to click the left mouse button when they heard a change from one key to another (conditions one and two), and right-click the mouse button when they heard a change from no key to a key (condition five), in order to concentrate their attention on the tonal structure of the stimulus stream. The task was explained as clicking in response to a change (since nonmusicians would not know what is meant by a key), and a short training session prior to scanning was used to ensure that subjects understood which type of change was being referred to. The behavioural results indicated that subjects clearly understood and were able to carry out the task.

2.2 Subjects

We tested sixteen right-handed subjects with normal hearing (9 female, 7 male; age $19-31$) none of whom had received any formal musical education. All subjects gave written informed consent to the study, which was approved by the Ethics Committee of the University of Magdeburg.

2.3 Data Acquisition

Functional magnetic resonance imaging data was acquired at the Leibniz Institute of Neurobiology (Magdeburg, Germany) on a Siemens Trio (Erlangen, Germany) 3T MRI scanner equipped with an eight channel head coil. Functional volumes were collected using echo planar imaging (EPI) with the following parameters: TE=30ms; TR=2000ms; interslice time: 62ms; slice thickness: 3mm; slice gap thickness: 0.3mm; inplane resolution: 3mm$\times$3mm (giving 3mm$\times$3mm$\times$3mm cubic isovoxels); number of slices: 32; FA: $80°$; FOV: 192mm$\times$192mm; matrix size: 64×64. Stimulus delivery and scanning coordination was controlled with the Presentation$^{©}$ software (Neurobehavioural Systems Inc, Albany, USA) using a custom-written script. The perceived scanner noise was attenuated by earplugs (24 dB) and ear muffs ($20 - 30$ dB) in which MRI compatible electrodynamic headphones were integrated [18]. The loudness of the stimuli was individually adjusted to a comfortable level. Each stimulus block lasted 8s (4 volumes) and was immediately followed by the next stimulus block. Two experimental runs were carried out during the session, with 20s (10 volumes) before each run, and after the final run, to provide a baseline condition. Altogether, each session therefore consisted of 606 functional volumes, as well as anatomical data collection and dummy runs for scanner alignment. Subjects were also given an initial scan-free practice period on stimuli not used in the functional data collection in order to ensure that they understood the task.

2.4 Pre-Processing

Functional data in the original space were co-registered with in-plane anatomical data and preprocessed with 3D motion correction using trilinear interpolation, slice scan time correction using sinc interpolation, linear trend removal, high-pass temporal filtering with a cut-off of 3 cycles throughout the time course, equivalent to 0.00236593Hz, and Gaussian spatial smoothing with a FWHM of 4mm. A correlation analysis was performed in the original data space following pre-processing to verify its effectiveness. They were then co-registered with a high-quality anatomical data set and transformed into AC-PC space using using cubic spline interpolation and subsequently into Talairach space with trilinear interpolation. These data were then used separately for both the conventional analysis and the machine learning analysis.

3 Experimental Results

3.1 Conventional fMRI Analysis

The conventional analysis was performed with a General Linear Model (GLM) using Brain-Voyager QX© (Brain Innovation B.V.,Maastricht). Second-level (group) analyses for all the results reported below were carried out using a two-step procedure. A grey-matter cortical mask was applied and a fixed effects analysis within voxels with global Bonferroni correction carried out to identify areas that are implicated in the processing of tonal sequences for the group (and by inference for at least a proportion of the underlying population). Only clusters that were strongly significant after this stringent correction procedure ($p<0.05$, corrected) were considered further. A more specialised tonality cortical mask was created based on these clusters (separately for each GLM contrast of interest), and a random effects analysis (now also including between-subjects variance) carried out, and only those voxels that survived ($p<0.05$, corrected at the mask level for a 5% false discovery rate) were considered further. Finally, a minimum cluster size of 500 voxels was imposed in order to allow only large activations. All active clusters over the baseline after this procedure are reported in the results below. The purpose of this approach is to identify regions involved in tonality within the group (fixed effects analysis), and to determine the level of inference possible to the underlying population (random effects analysis), with a stringent multiple comparisons correction procedure and cluster size limit imposed in order to avoid artefacts and susceptibility errors.

3.1.1 Tonal vs Atonal

Location	Tal X	Tal Y	Tal Z	Cluster Size
Right Precentral Gyrus (BA 4)	49	-10	45	1872
Left Medial Frontal Gyrus (BA 6)	-1	-4	58	742
Left Precentral Gyrus (BA 4)	-49	-10	42	1257

Table 1: Anatomical results of a GLM analysis contrasting tonal and atonal conditions.

Condition	49 -10 45	-1 -4 58	-49 -10 42
distant	0.739278	1.2853	1.06331
close	0.785971	1.23277	1.18109
same	0.864002	1.3916	1.25538
none	0.694907	1.1751	1.0339
initial	0.914343	1.35245	1.26462

Table 2: GLM beta statistics for each experimental condition for the three active clusters, identified here by their Talairach coordinates.

The contrast examined here was for the initial tonal stimuli vs the atonal stimuli, in order to avoid any confounds related to key changes. The results can be seen in table 1. There are three major active clusters. It is well established that left frontal regions are involved in hierarchical pattern processing, including music processing [5], and is necessary for the cognition of tonality, so it is not surprising to find activation in the left medial frontal gyrus (BA6; the supplementary motor area). A clue to the specific role of this area in tonality processing comes from the strong bilateral activation of the precentral gyrus (see figure 1).

One possibility is that this is simply motor activity related to the mouse click showing a change from no key to key, with no mouse click required when moving from tonal to atonal stimuli. However, an analysis with the contrast same (condition 3) vs atonal (condition 5) (neither of which requires a mouse click) was performed, and the same results were found, so motor activity is not an adequate explanation here. The regression beta statistics (table 2) show clearly that this area has a strong association with processing stable tonal stimuli. There is a long tradition of musicians and music psychologists investigating the link between music and motion [19], and with these three active clusters all involved in motor processing, we have found what we believe is evidence for this link at a neural level.

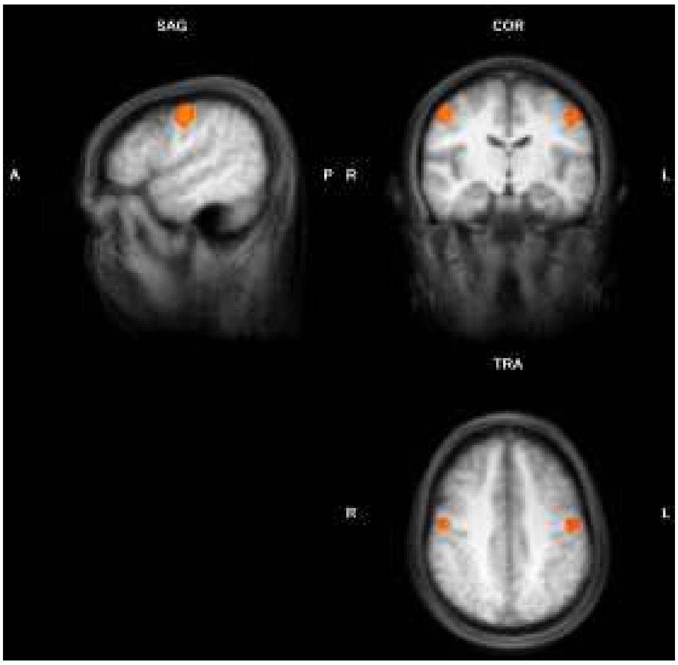

Figure 1: Bilateral activation of the precentral gyrus for tonal vs atonal stimuli. All active clusters preferentially favour tonal stimuli.

3.1.2 Key Changes

Location	Tal X	Tal Y	Tal Z	Cluster Size
Right Transverse Temporal Gyrus (BA 41)	51	-17	10	1023
Right Insula (BA 13)	36	17	13	948
Right Lentiform Nucleus	24	-1	1	750
Right Caudate	14	-4	22	1443
Left Anterior Cingulate (BA 32)	-1	41	11	2574
Left Cingulate Gyrus (BA 24)	0	-16	35	786
Left Superior Frontal Gyrus (BA 8)	-12	50	36	2241
Left Transverse Temporal Gyrus (BA 41)	-51	-18	11	981

Table 3: Anatomical results of a GLM analysis contrasting conditions with and without a key change. All active clusters preferentially favour key change stimuli.

We examined the difference between the two conditions containing a key change, and the third condition which remained in the same key. A balanced GLM contrast between these was performed and the results are shown in table 3. A diverse network of activation is present, of which two features are notable. First is the strong presence of medial structures, in particular cingulate cortex and caudate nucleus. Second is the presence of preferential activation for key changes in bilateral auditory cortex (transverse temporal gyrus). The activation curves in these areas show strongest activity for the distant key changes, slightly less (but still significant) activity for the close key changes, and much less activity for no key changes. It should be emphasised that this occurred across a variety of different stimuli (all of equal amplitude and with very similar basic auditory features, such as envelope and broad spectral content), and that both left and right auditory cortex showed very similar response curves (see figure 2), highlighting the robust nature of this finding. This suggests that these auditory cortical areas may not be limited to low level individual tone processing, but might also be involved in some higher-order sequence processing. This intriguing result merits further investigation.

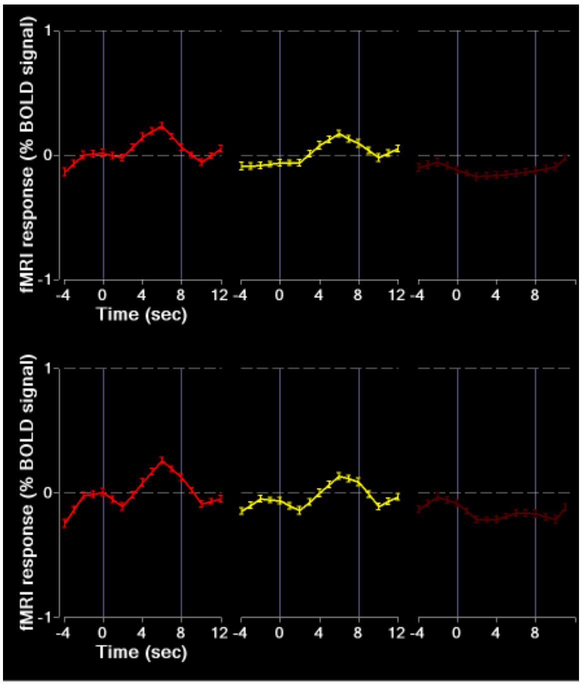

Figure 2: Activation curves in right (top) and left (bottom) auditory cortices, for changing to a distant key (left), changing to a close key (middle) and remaining in the same key (right). Times are shown relative to the start of a given stimulus block; the curves represent averages over all blocks in a given condition.

3.1.3 Machine Learning fMRI Analysis

In order to test the ability of machine learning analysis to detect patterns in fMRI data relating to higher order cognitive tasks, we adopted a machine learning framework of Support Vector Machines (SVM) [20, 21] for a leave-subject-out analysis, i.e. we learn a discriminatory task on the combined data of 15 subjects and test on the remaining subject. This procedure is repeated for all the subjects and averaged for all. For comparison with the GLM analysis we confined ourselves to the two following tasks: tonal (condition 5) versus atonal (condition 4) and key change (conditions 1 & 2) versus no key change (condition 3) stimuli. The overall aim is to observe whether given the training subjects, in either of the tasks, is it possible to generalise the task for a new subject.

We used a linear kernel SVM that allows direct extraction of the weight vector as an image (i.e. the discriminating spatial pattern). A parameter C, that controls the trade-off between training errors and smoothness was fixed at $C = 1$ for all cases (default value).[1] For both of the experiments we used the same grey matter cortical masks that were applied in the GLM analysis. These were combined to give a meta grey matter mask across all individuals. In addition to the preprocessing that was performed on the data prior to the GLM analysis (which was also adopted for this machine learning analysis) we further preprocessed the data by subtracting from each subject the first scan representing silence as to remove baseline 'noise' in the data.

In summary, we obtained a 54% accuracy for tonal versus atonal discrimination, with the detection of tonality and atonality each obtaining an accuracy of 55% and 49% respectively. The key change versus no key change discrimination task obtained an accuracy of 54%, which comprised of 82% for detecting key change and 25% for no key change. It seems that even with masking and the subtraction of the initial scan the difference between the conditions in each case remains minimal, with the result that it is extremely difficult for the learning algorithm to find a hyperplane that will accurately distinguish between them.

4 Discussion

Using a conventional analysis, we have found clear activation in motor areas in response to tonal (as opposed to atonal) stimuli that cannot be attributed to actual motor activity. Looking more closely at the relative distance of key changes within a tonal context, we have found that activity in auditory cortical areas is associated with this distance, suggesting that these areas may not be limited to low-level single tone processing as previously thought. These unusual findings were detected with a robust analytical procedure and a controlled experimental design, and suggest the need for further research in this area.

The relatively low classification performance of the machine learning algorithm we attribute to two related factors. Firstly, the task itself is one focusing on higher order cognitive functions (the cognition of tonal structure), which are notoriously difficult to distinguish on the basis of BOLD (blood-oxygen level dependent) response (the standard fMRI signal). Secondly, inter-subject variance was quite high, which means that a group-level classification performance is liable to perform badly because even with a leave-one-out approach, the training data and test data are potentially quite different. As such, we would suggest that in these situations, regression analysis might be more beneficial, and that machine learning framework for fMRI focusing on regression rather than classification needs to be developed.

Acknowledgements

This work was supported by the EPSRC project 'Learning the Structure of Music' (LeStruM)[2], EP-D063612-1. Also we would like to thank Janaina Mourão-Miranda for helpful advice. This publication only reflects the authors views.

[1] The LibSVM toolbox for Matlab was used to perform the classifications http://www.csie.ntu.edu.tw/∼cjlin/libsvm/.

[2] LeStruM project website http://www.lestrum.org/ .

References

[1] Walter Piston and Mark Devoto. *Harmony (5'th ed)*. New York: Norton, 1987.

[2] Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. OUP, Oxford, 1991.

[3] Roger N. Shepard. Structural representations of musical pitch. In Diana Deutsch, editor, *The Psychology of Music*, pages 344–390. OUP, 1982.

[4] Eugene Narmour. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press, 1991.

[5] Mary Tervaniemi. Musical sound processing: EEG and MEG evidence. *The Cognitive Neuroscience of Music*, pages 294–309, 2003.

[6] Petr Janata, Jeffrey L. Birk, John D. Van Horn, Marc Leman, Barbara Tillmann, and Jamshed J. Bharucha. The cortical topography of tonal structures underlying western music. *Science*, 290:2167–2170, 2002.

[7] D. D. Cox and R. L. Savoy. Functional magnetic resonance imaging (fmri) 'brain reading': detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19:261–270, 2003.

[8] T. A. Carlson, P. Schrater, and S. He. Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience*, 15:704–717, 2003.

[9] X. Wang, R. Hutchinson, and T. M. Mitchell. Training fmri classifiers to detect cognitive states across multiple human subjects. In *In Proceedings of the 2003 Conference on Neural Information Processing Systems*, 2003.

[10] T.M. Mitchell, R. Hutchinson, R.S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine Learning*, 1-2:145–175, 2004.

[11] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu. Support vector machines for temporal classification of block design fmri data. *NeuroImage*, 26:317–329, 2005.

[12] J. Mourao-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and S. Stetter. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. *NeuroImage*, 28:980–995, 2005.

[13] J. D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, 8:686–691, 2005.

[14] C. Davatzikos, K. Ruparel, Y. Fan, D. G. Shen, M. Acharyya, J. W. Loughead, R. C. Gur, and D. D. Langleben. Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *NeuroImage*, 28:663–668, 2005.

[15] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *PANAS*, 103:3863–3868, 2006.

[16] T. Eerola and P. Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyvskyl: Kopijyv, Jyvskyl, Finland. Available at http://www.jyu.fi/musica/miditoolbox/, 2004.

[17] C. L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89:334–368, 1982.

[18] F. Baumgart, T. Kautisch, C. Tempelmann, B. Gaschler-Markefski, C. Tegeler, F. Schindler, D. Stiller, and H. Scheich. Electrodynamic headphones and woofers for application in magnetic resonance imaging scanners. *Med. Phys*, 25:2068–2070, 1998.

[19] Eric F. Clarke. Meaning and the specification of motion in music. *Musicae Scientiae*, 5:213–234, 2001.

[20] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

[21] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.