DETECTION OF CHANGES IN PATTERNS OF BRAIN ACTIVITY ACCORDING TO MUSICAL TONALITY

Tom Diethe
Department of Computer Science
University College London
London, UK
email: t.diethe@cs.ucl.ac.uk

Simon Durrant
School of Psychological Sciences
The University of Manchester
Manchester, UK
email: Simon.Durrant@manchester.ac.uk

John Shawe-Taylor
Department of Computer Science
University College London
London, UK
email: j.shawe-taylor@cs.ucl.ac.uk

Heinrich Neubauer
Leibniz Institute for Neurobiology
Magdeburg, Germany
email: heinrich.neubauer@ifn-magdeburg.de

ABSTRACT
A common structural element of Western tonal music is the change of key within a melodic sequence. The present paper examines data from a set of experiments that were conducted to analyse human perception of different modulations of key. EEG recordings were taken of participants who were given melodic sequences containing changes in key of varying distances, as well as atonal sequences, with a behavioural task of identifying the change in key. Analysis of EEG involved derivation of 122120 separate dependent variables (features), including measures such as inter-electrode spectral power, coherence, and phase. The paper presents a novel method of performing semantic dimension reduction that produces a representation enabling high accuracy identification of out-of-subject tonal verses atonal sequences.

KEY WORDS
Machine Learning, KCCA, EEG, Music, Tonality

1 Introduction

The analysis of brain scans with a view to accurately identifying the semantic processing of the subject has received increasing attention recently [1]. Analysis of human perception of musical tonality has also received some attention, although these studies are mainly concerned with evoked potential (ERP) paradigms (e.g. [2, 3]) or localisation of activity from fMRI (e.g. [4, 5]).

In the present study we are concerned with distinguishing between different tonal and atonal stimuli through long-term changes in observed electroencephalography (EEG) recordings of the subjects. EEG is the measurement of electrical activity produced by the brain as recorded from electrodes placed on the scalp. In the present study we will use single EEG trials as inputs to machine learning algorithms. We believe that EEG is ideally suited to studying musical stimuli due to its high temporal resolution. However EEG recordings are notoriously noisy and making reliable cross-subject predictions has proved difficult even for simple tasks. Indeed we will see that a naive application of support vector machines (SVMs) to the collected signals is unable to make out-of-subject predictions much better than chance, although within-subject predictions were possible. The contribution of this paper lies in both the application and the method. Firstly this is a new EEG dataset from a novel experimental design. Secondly the method for classification of EEG signals using the additional information from the pitch kernels is novel. Finally, the key contribution of the paper will be the demonstration of a novel semantic dimension reduction method that makes use of a complex description of the stimuli to identify key dimensions in the space of signals that highly correlated with the stimulus. Using even a simple nearest neighbour classifier in this semantic space can achieve very high accuracy in both within-subject and out-of-subject prediction.

The goal is then to discover statistical relationships between musical structure and EEG recordings of participants to the same music. The proposed analysis to achieve this is based on the premise that the brain represents structural elements of the auditory signal that it receives through shifting patterns of activity. This activity may take many forms, ranging from generalised changes in activity in certain brain regions to more complex relationships. By taking a multivariate approach to the signal processing of the EEG signal, it is possible to analyse a wide range of such relationships. As such pairwise electrode comparisons, which provide an indication of communication between brain regions, are of paramount importance. The analysis to date has included pairwise statistics such as cross power and coherence. Cross phase is another interesting statistic that is investigated, as it indicates that there may be an increase (or decrease) in synchrony between brain regions. The collection of statistics derived from the EEG analysis procedure will then be compared with the features derived from the audio recordings in order to seek common patterns.

In order to use the additional information from the stimuli to aid classification, the stimuli must be encoded through a kernel designed to capture the melodic and harmonic structure of a musical score available in a simple midi format. The kernel chosen for the present study captures distances between musical sequences embedded into pitch-class space, which is ideal for separating sequences according to their tonality and key structure.

Section 2 gives details of the setup and protocol of the experiment upon which the analysis was performed, including details of the EEG data preprocessing . Section 3 gives details about the process of the multivariate signal processing techniques used to extract features from the EEG data for classification. Section 4 describes the machine learning analysis approaches taken, including conventional SVM analysis as well as a semantic dimension reduction method based on KCCA.

2 Experiment

The data under examination in this paper was produced by an EEG experiment conducted in partnership with the University of Magdeburg. The principal hypothesis was that neural patterns should reflect relative changes in the key of music that a listener is attending to. In order to examine this, a series of stimuli (chord sequences) were constructed and ordered such that there were the following five experimental conditions:

1. Distant key
2. Close key
3. Same key
4. No key (atonal)
5. Initial

2.1 Participants

16 right-handed participants (9 female, 7 male), aged 19 to 31, with normal hearing took part in the experiment. None had received any formal musical education. All participants gave written informed consent to the study, which was approved by the ethics committee of the University of Magdeburg.

2.2 Experimental Design

The stimuli consist of sequences of chords, with each stimulus in a single key (or no key). All sequences consist of 16 chords with onsets at 500ms intervals and with duration filling the entire 500ms, giving a total length of 8s. The experimental conditions are defined by contiguous stimulus triplets with changes in relative key (listed below). Relative key is established by tonal stimuli, and reset by atonal stimuli. Stimuli from the first three conditions are followed by a stimulus from condition four as a contrast and a reset of relative tonality. 48 stimuli required altogether, all chordal (in root position), of which 32 are tonal and 16 atonal. Tonal stimuli to be transposed as required to fulfil experimental role. First stimulus in each tonal pair is to be in C major, to eliminate any long-term tonality effects (or at least to take advantage of them); second is in either F# major (condition 1), G major (condition 2) or C major (condition 3). In total there were 48 ‘initial’, 48 ‘atonal’, 16 ‘close’, 16 ‘distant’ and 16 ‘same’ trials per participant, giving a total of 144 trials.

Ordering Principles:-

(1) Each condition should appear an equal number of times (2) Each different melody type (a,b etc.) should appear an equal number of times (3) The three conditions should appear in each permutation (to minimise condition order effects) (4) Each different melody type should be used once for each of the three main conditions (to minimise individual melody effects) (5) Each tonal pair in the conditions should use the same stimulus (6) Each tonal pair should be followed by a unique atonal stimulus to reset tonality (and provide a control condition) (7) Same order for each run and for each subject (for direct comparison in subsequent analysis)

2.3 EEG Measurements

EEG recordings were acquired at the Leibniz Institute for Neurobiology (Magdeburg, Germany). 64 unipolar channels, including 2 Electrooculogram (EOG) channels and one nose reference electrode were recorded at a sampling of $500Hz$ and a resolution of $0.1\mu V$. Across all participants the voltage range was $3.2767mV$ and the impedance was less than $5k\Omega$. The music was played to the participants using a Terratec EWX 24/96 soundcard, Black Cube Linear Science amplifier by Lehmann Audio (www.lehmannaudio.de), and Eartone 3A Insert Earphones 50Ω using binaural presentation. The volume of the amplifier was at notch 6. Stimulus delivery and scanning coordination were controlled with $Presentation^{©}$ software (Neurobehavioural Systems Inc, Albany, USA) using a custom-written script.

2.4 Data Preprocessing

Muscular activity related to eye movements and eye blinks alter the electromagnetic fields around the eyes and typically introduce artefacts into the EEG, especially in frontal regions. A number of algorithms have been proposed to correct for EOG artefacts, which all correct for EOG artefacts by subtracting a proportion of one or more EOG channels from the EEG channels. A study by [6] evaluated four correction techniques by correcting blinks, vertical and horizontal eye movements from 26 subjects. The study concluded that in the absence of specific calibration protocols, the method described by [7], based on multiple regression, was the best solution. The approach taken by [8] was based on the algorithm suggested by [7], with modifications described in [9]. This latter method was chosen for the present study.

Prior to time-frequency analysis, the data was filtered using two-way least-squares FIR filtering. Digital filters: $0.2Hz$ low pass filter. $100Hz$ high pass filter. The $50Hz$

component of the signal was removed using a notch filter between $49Hz$ and $51Hz$ due to AC mains signal.

The electrodes were then re-referenced using the nose electrode.

3 Feature Extraction

The data from the 64 channel EEG system at 500Hz sampling rate was imported as a single matrix such that the format was [channels x frames]. The data was segmented into 8 second epochs, giving 144 epochs per subject. These epochs have a one-to-one correspondence with the experimental stimuli. This results in a data matrix of shape [channels x frames x epochs].

3.1 Time-Frequency Analysis

A multitaper spectrum is produced by averaging multiple windowed FFTs generated with a set of orthogonal data tapering windows known as discrete prolate spheroidal sequences (DPSS) or Slepian functions. Since each of the windows in a specific sequence is uncorrelated, an unbiased average spectrum can be produced. A multitaper spectrum offers no greater frequency resolution than a single tapered spectrum. In fact, the spectral peaks resulting from the algorithm have a flat-topped envelope shape which makes the central frequency determination more difficult. What is gained is a reduced-variance spectral estimator that retains a high dynamic range. [10]

Using DPSS, inter-channel coherence, cross phase and cross power were computed, for all pair-wise combinations of channels, excluding the EOG electrodes and nose reference electrode. Cross power simply refers to the ratios of the power within each of the frequency bandwidths. The coherence function measures the correlation between two signals as a function of the frequency components they contain, and is therefore a correlation spectrum [11, 12]. It determines the likelihood of two stochastic signals arising from the same generating process.

This differs from the cross-correlation function, which involves calculating Pearson product-moment correlation coefficients for the two signals at various displacements of sampling interval. Quantitative analysis [12] has shown that the cross-correlation sometimes fails in situations where coherence does not, as well as being more expensive to compute. Complementary to the computation of the coherence spectrum is the phase spectrum, which indicates the phase relationship between two signals as a function of frequency - information that is lost using ordinary spectral methods. An important feature of all of these methods is that they are independent of amplitude, as the amplitudes of electrodes are known to vary greatly both within and between recording sessions.

The resulting 256 fourier coefficients for each of the measures were divided into bands, providing estimates of spectral power within the following recognised frequency bandwidths:

- delta (0.3-3.9Hz)
- theta (4-7.9Hz)
- alpha (8-13Hz)
- beta1 (13-19Hz)
- beta2 (20-30Hz)

In addition,

- low gamma (30-42Hz)
- 40Hz (38-42Hz)
- mid gamma (43-63Hz)
- high gamma (64-100Hz)
- general gamma (30-100Hz)
- global (0.01-100Hz)

bandwidths were computed. There were 61 channels after removal of EOG/nose reference. After calculating the DPSS, the data were split into the 11 bands listed above. For each measure we calculated the mean and variance (i.e. Gaussian fitting): There were then 61x11x2 PSD values; 61x61x11x2 Cross-PSD values, 61x61x11x2 coherence values; 61x61x11x2 cross-phase values. However clearly only the upper diagonal parts of the inter-electrode comparison matrices are needed. These features were then concatenated to create a large feature vector of length 122120 for classification.

4 Results

4.1 SVM Analysis

The data was standardised across the features to obtain “standard normal” random variables with mean 0 and standard deviation 1. The data for each subject was split randomly into 75% train, 25% test and then concatenated to form the full training and test sets. The same random split was applied for all of the analysis. Classification was performed using the SVM-Light Support Vector Machine implementation [13] with linear, rbf and laplace kernels (where the laplace kernel is the same as the rbf kernel except that the 2-norm is replaced with a 1-norm). 5-fold cross validation was performed on the training set to discover best setting of the C and sigma parameters, before testing on the separate test set. Table 1 shows the test errors for the SVM classifier on the split of the data described above. The significance of the classifier was evaluated using the upper bound of the cumulative distribution function of the binomial distribution of a random classifier, calculated as follows:

$$p \leq \exp\left(-2\frac{(np-k)^2}{n}\right) \tag{1}$$

where n is the number of trials (test examples), p is the probability of success (0.5 for a random classifier) and k is the test error of the classifier.

Table 1. Test errors for within-subject SVM classification. ** denotes significance at the $p < 0.001$ level (see text)

Test	Linear	RBF	Laplace
Tonal vs Atonal	0.2298**	**0.1175****	0.2742**
Close vs Distant	0.3125**	**0.2422****	0.4375
Same vs Distant	0.2656**	0.2344**	**0.2109****
Same vs Close	0.2031**	**0.1641****	**0.1641****

Table 2 shows the leave-one-out test error for each of the participants using a linear kernel. In this test the data from 15 of the participants is used as the training set and the data from the remaining participant is used as the testing set. This is a much more difficult test, in the sense that we are now trying to learn features that can generalise from one set of brains to a new brain. It is therefore not surprising that with a subject pool of only 16 participants the classification errors are close to chance for most subjects. Results (not given) for the rbf and laplace kernels were not significantly different. It is interesting to note that the distinction between "close" and "distant" gives the best classification results rather than tonal vs atonal. As such it appears that conditions with key changes result in more consistent prediction across brains than those for processing atonal music.

Table 2. Test errors for tonal vs atonal leave-one-out SVM classification using a linear kernel. The numbers in parentheses represent the number of test examples. None of the test errors reached significance at the $p < 0.01$ level

Subject	Tonal v atonal (96)	Close v distant (32)	Same v distant (32)	Same v close (32)
1	0.4583	0.3438	0.3125	0.3750
2	0.4947	0.4688	0.3438	0.4375
3	0.4688	0.3438	0.3750	0.4062
4	0.4688	0.4375	0.5000	0.4062
5	0.4896	0.4688	0.5000	0.4062
6	0.5000	0.5000	0.4375	0.4688
7	0.4583	0.4688	0.4375	0.4375
8	0.4896	0.3750	0.3438	0.5000
9	0.4896	0.4375	0.5000	0.5000
10	0.4688	0.4062	0.3750	0.4688
11	0.4792	0.3438	0.5000	0.4688
12	0.4792	0.3125	0.4062	0.5000
13	0.4583	0.3750	0.4375	0.5000
14	0.5000	0.3750	0.5000	0.4062
15	0.4688	0.4375	0.3125	0.4688
16	0.5000	0.3125	0.3750	0.5000
mean	0.4795	0.4004	0.4160	0.4531
median	0.4792	0.3906	0.4219	0.4688

4.2 KCCA Analysis

Various methods have been proposed for searching for common patterns between two sets of signals, including kernel canonical correlation analysis (KCCA), which can be viewed as a generalised form of kernel independent components analysis [14]. Canonical correlation analysis (CCA) is a technique to extract common features from a pair of multivariate data. KCCA is a nonlinear version of this technique which allows nonlinear relations to be found between multivariate variables effectively [15].

$$\rho = \max_{\alpha,\beta} \frac{\alpha' \mathbf{K}_x \mathbf{K}_y \beta}{\sqrt{\alpha' \mathbf{K}_x^2 \alpha \cdot \beta' \mathbf{K}_y^2 \beta}} \tag{2}$$

For this analysis it was necessary to calculate kernels on the musical stimuli. For simplicity of analysis, the only distinction being examined in this section is tonal vs atonal, as the experimental setup does not lead to a simple calculation of relative pitch for stimuli that were presented following silence. The midi audio files used to generate the experimental stimuli were first embedded into pitch class space. Pitch class space [16] is the circular (quotient) space with the result that differences between octave-related pitches are ignored. In this space, there is no distinction between tones that are separated by an integral number of octaves. The pitch class vectors for each stimulus were then formed into kernels using a squared exponential kernel. As a sanity check, running an svm on these gives a test error of 0.0261, showing that this kernel representation is valid. Perfect classification was not achieved as there appear to be outlier stimuli, i.e. atonal sequences that appear tonal in this represenation.

For the purposes of the KCCA analysis, a linear kernel is used for the EEG, as the dimensionality of the RBF kernel in this case is too high. Both kernels were projected into Gram-Schmidt space using the partial Gram-Schmidt decomposition outlined in [15]. The precision parameter was set to 0.3 using a heuristic method. The use of this decomposition results in an implicit regularisation, and as such the KCCA regularisation parameter was set to zero. Experimentation with different values of this parameter did not show any improvement in results.

The kernels from each view were then projected into the shared feature space using the top 100 resulting KCCA directions. The test kernel for the EEG was also projected into this space, and then normalised such that the L2-norm of vector was 1. Using the 100 largest correlation values and the corresponding projections, the labels given by the corresponding example in the music kernel were used as the classification. The reported errors are then the mean of the differences between these labels and the true test labels. This method is an extension of mate-based retrieval [17], and is given algorithmically below.

The classification results using the PNN classification approach are given in table 3. It can be seen that this method is able to classify between the tonal and atonal experimental conditions almost perfectly. As a comparison, we trained an SVM on the projection of the EEG data into the shared feature space, using a linear kernel and 5-fold cross validation to select the C parameter. The results show that the PNN method performs competitively with the SVM, whilst being essentially an unsupervised method. It

Algorithm 1 Projected Nearest Neighbour (PNN) Classification

1: Given Kernels from each view $\mathbf{K}_x$ and $\mathbf{K}_y$, dual weight vectors α and β from KCCA, training labels $\mathbf{y}$, and vectors of train and test indices $\mathbf{i}$ and $\mathbf{j}$ respectively

2: Compute the projection of the training kernel of the first view

$$P_x \leftarrow \mathbf{K}_x[\mathbf{i}, \mathbf{i}]\alpha$$

3: Compute the projection of the train-test kernel of the second view:

$$P_y \leftarrow \mathbf{K}_y[\mathbf{j}, \mathbf{i}]\beta$$

4: Compute the covariance matrix of the projections:

$$C_{xy} \leftarrow P_x P_y'$$

5: Find the indices of the maximal values of each column:

$$\mathbf{z}_j = \arg\max_i \left(C_{xy}[i, j]\right) \quad \text{for } j = 1, \ldots, k$$

where k is the cardinality of $\mathbf{j}$

6: Select the labels of the training examples of those indices as the predictions:

$$\hat{\mathbf{y}} \leftarrow \mathbf{y}[\mathbf{z}]$$

is also much more computationally efficient as there are no parameters to tune.

Table 3. Test errors for within-subject classification for Tonal vs Atonal using KCCA projected nearest neighbour (PNN) and SVM classification. ** denotes significance at the $p < 0.001$ level

Classifier	# Train	# Test	Linear
KCCA + PNN	1152	383	0.0183**
KCCA + SVM (linear)	1152	383	0.0157**

Table 4 shows the leave-one-out test error for each of the participants using the PNN classification approach, along with the SVM trained on the projection of the EEG data into the shared features space, again using a linear kernel and 5-fold cross validation to select the C parameter. The results show that the PNN method performs competitively with the SVM, whilst both signficantly outperform the naive SVM approach (see table 2).

5 Discussion

The results demonstrate that using standard modern signal processing and machine learning techniques with careful manipulation of the data can enable us to differentiate between certain patterns of brain activity. Coherence analysis and other types of cross-spectral analysis may be used to identify variations which have similar spectral properties (high power in the same spectral frequency bands) if the variability of two distinct time series is interrelated in the spectral domain. In the present study we were able to reliably distinguish between whether a listener was attending to tonal or atonal music. This can be considered to be a task of high-order cognitive processing, rather than a simple sensory input task. As the differentiation was based on properties of the EEG over relatively long timespans (i.e. the length of an epoch, or 8 seconds), this is clearly not due to simple evoked potentials, but instead represents a more fundamental change in the pattern of processing over time.

Table 4. Test errors for leave-one-subject-out KCCA projected nearest neighbour classification. * and ** denote significance at the $p < 0.01$ and $p < 0.001$ level respectively

Participant	KCCA + PNN	KCCA + SVM (linear)
1	0.2708	0.1667**
2	0.2737	0.2421
3	0.3125	0.2500
4	0.2083*	0.1667**
5	0.4062	0.2500
6	0.2500	0.2500
7	0.5625	0.1667**
8	0.2500	0.2500
9	0.2708	0.2500
10	0.1667**	0.1667**
11	0.7396	0.2500
12	0.2500	0.2500
13	0.1562**	0.1667**
14	0.3542	0.2500
15	0.2500	0.2500
16	0.4688	0.1667**
mean	0.3244	0.2183
median	0.2708	0.2500

Further analysis using KCCA demonstrated that through the use of unsupervised methods it is possible to significantly improve the classification accuracy. We developed a new classification method using the shared semantic space given by projections from KCCA weight vectors together with a nearest neighbour method. This was able to distinguish between the tonal and atonal experimental conditions with a high degree of accuracy. The PNN method is not only nonparametric, it is basically free given the labels on the second view. Since this is common for the experimental paradigm presented (i.e. not limited to analysis of music).

We also showed that an SVM trained on projected data performed extremely well. We attribute the success of both of these methods as being due to the KCCA projections acting as a data cleaning step, in which a form of semantic dimensionality reduction is occurring. As the musical stimuli are sufficiently distinct between conditions, the additional information extracts the directions correlated with the differing experimental conditions. The key ingredient in the approach is the introduction of a clean source of data that encodes a complex description of the experience of the subject. We believe that this approach to information extraction has enormous promise in a wide range of signal processing and time series data analysis tasks.

We were also able to reliably discriminate between subtler discriminations in the task of the listener, such as distinguishing a move from one key to a close or distant key. However the results were not as convincing as for the tonal-atonal distinction. There are several possible reasons for this. Firstly, there were fewer examples of these events by a factor of 3, which on its own increases the difficulty in learning. Secondly, the cognitive task is clearly much more subtle than the tonal vs atonal case, and as such the changes in patterns of activity are likely to be much more subtle, although this is of course speculative. Finally, the type of relationship between the patterns of activity in this case may be qualitatively rather than quantitatively different, meaning that the signal processing techniques employed were unable to detect them (as opposed to the learning algorithm). Further experiments with larger datasets (more repetitions or more participants) could provide the answers to these questions.

EEG data is notoriously noisy and unreliable, so it is extremely encouraging that it is possible to generate reliable discriminations using fully automatic procedures. It is usual to perform artefact rejection by hand during the preprocessing stage, as well as other manual techniques. The present study used automatic techniques at every stage of the process (preprocessing, feature extraction, data treatment, and classification). The methods presented demonstrate the ability to reliably discriminate between brain signals associated with different sequences of music in both within-subject and out-of-subject paradigms.

References

[1] J. Mourao-Miranda, A.L.W. Bokde, C. Born, H. Hampel, and M. Stetter. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional mri data. *NeuroImage*, 4:980–995, 2005.

[2] S. Koelsch, T. Gunter, E. Schröger, and A. D. Friederici. Processing tonal modulations: An ERP study. *Journal of Cognitive Neuroscience*, 15:8:11491159, 2003.

[3] S. Koelsch and W. Siebel. Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9:578–584, 2006.

[4] P. Janata, J.L. Birk, J.D. Van Horn, M. Leman, B. Tillmann, and J.J. Bharucha. The cortical topography of tonal structures underlying western music. *Science*, 298:2167–2170, 2002.

[5] S. Durrant, D.R. Hardoon, A. Brechmann, J. Shawe-Taylor, E. Miranda, and H. Scheich. GLM and SVM analyses of neural response to tonal and atonal stimuli: New techniques and a comparison. *Special Issue on Music, Brain and Cognition*, In press.

[6] R.J. Croft, J.S.Chandler, R.J.Barry, and N.R.Cooper. EOG correction: A comparison of four methods. *Psychophysiology*, 42:16–24, 2005.

[7] G. Gratton, M. G. H. Coles, and E. Donchin. A new method for off-line removal of ocular artifact. *Electroencephaolography and Clinical Neurophysiology*, 55:468–484, 1983.

[8] C.W. Pleydell-Pearce, S.E. Whitecross, and B.T. Dickson. Multivariate analysis of EEG: Predicting cognition on the basis of frequency decomposition, inter-electrode correlation, coherence, cross phase and cross power. In *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS03)*, 2003.

[9] M.A. Conway, C.W. Pleydell-Pearce, and S.E. Whitecross. The neuroanatomy of autobiographical memory: A slow cortical potential study of autobiographical memory retrieval. *Journal of Memory and Language*, 45:493524, 2001.

[10] D.B. Percival and A.T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge University Press, 1993.

[11] J.C. Shaw. An introduction to the coherence function and its use in EEG signal analysis. *Journal of Medical Engineering Technology*, 5:279–288, 1981.

[12] J.C. Shaw. Correlation and coherence analysis of the EEG: a selective tutorial review. *International Journal of Psychophysiology*, 1:255–266, 1984.

[13] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

[14] F. Bach and M. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

[15] D.R. Hardoon, S.Szedmak, and J.Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[16] D. Deutsch and J. Feroe. The internal representation of pitch sequences in tonal music. *Psychological Review*, 88:503–522, 1981.

[17] D.R. Hardoon and J. Shawe-Taylor. KCCA for different level precision in content-based image retrieval. In *Proceedings of Third International Workshop on Content-Based Multimedia Indexing*, 2003.