Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Sparse Multiview Methods for Classification of Musical Genre from Magnetoencephalography Recordings

Tom Diethe*[1], Gabi Teodoru#[2], Nick Furl*[3], John Shawe-Taylor*[4]

*Department of Computer Science, University College London, UK
#Gatsby Computational Neuroscience Unit, University College London, UK
[1]t.diethe@cs.ucl.ac.uk, [2]n.furl@cs.ucl.ac.uk, [3]gabiteodoru@gmail.com, [4]jst@cs.ucl.ac.uk

ABSTRACT

Classification of musical genre from audio is a well-researched area of music research. However to the authors' knowledge no studies have been performed that attempt to identify the genre of music a person is listening to from recordings of their brain activity. It is believed that with the appropriate choice of experimental stimuli and analysis procedures, this discrimination is possible. The main goal of this experiment is to see whether it is possible to detect the genre of music that a listener is attending to from brain signals. The present experiment is focuses on Magnetoencephalography (MEG), which measures magnetic fields produced by electrical activity in the brain. We show that classification of musical genre from brain signals alone is feasible, but unreliable. We show that though the use of sparse multiview methods, such as Sparse Multiview Fisher Discriminant Analysis (SMFDA), we are able to reliably discriminate between different genres.

I. BACKGROUND

Classification of musical genre from raw audio files is a well researched area of music research. The Music Information Retrieval Evaluation eXchange (MIREX) is a yearly competition in a wide range of machine learning applications in music. MIREX 2005 included a genre classification task, the winner of which [1] was an application of the multiclass boosting algorithm AdaBoost.MH [2]. Linear Programming Boosting (LPBoost) [3] was shown to be a more appropriate algorithm for this application due to the higher degree of sparsity in the solutions [4]. That study also highlighted one of the main problems for audio identification tasks, namely the choice of an appropriate dataset. This is interesting from a cognitive perspective, as genre classification may represent both low- and high-order cognitive processes. To the authors' knowledge no studies have been performed that attempt to discern which genre of music a person is listening to on the basis of electrophysiological recordings of their brain activity. It is believed that with the appropriate choice of dataset as experimental stimuli (namely well chosen highly representative samples of specific genres) and appropriate analysis procedures, this discrimination is possible.

The analysis procedures employed in this study are based on those used for fMRI using standard GLM and SVM/KCCA methods [5], and methods used for analysis of EEG using KCCA as a semantic dimensionality reduction method prior to classification [6]. We begin with performing genre classification on the audio source only, as outlined in [4], except that in this study we use features derived from the midi versions of the audio files rather than raw audio files. The reasons for the are twofold. Firstly, the features of interest are more readily available from the midi, as we have direct access to the pitch values and note durations of the musical sequences. Secondly, the nature of the stimuli means that there is no timbral information available. Most of the features used in previous studies such as [1, 4] are based on short-term spectral information, most of which are strongly picking out timbral features.

Following this, features are derived from the MEG data using spectral methods common to the neuropsychological literature, after which machine learning algorithms are used to classify these features according to genre. We then apply multiview methods, following on from [5, 6], which attempt to use the stimuli themselves as another view in the classification of the brain signals. We improve on these methods through the use of Sparse Multiview Fisher Discriminant Analysis (SMFDA) [7]. The key difference between this and previous approaches is that SMFDA uses label information to find informative projections of each view into a shared space, which are more appropriate in supervised learning settings. In addition, SMFDA seeks to find sparse solutions by using L_1 optimisation, which is known to approximate the optimal sparse L_0 solution. This in turn is a form of regularisation that prevents overfitting in high dimensional feature spaces. Sparsity of solutions is important in this setting as the feature set constructed from the MEG data is extremely high dimensional, with a low signal-to-noise ratio.

From [8], the optimisation for SMFDA is given by,

$$\min_{\boldsymbol{\alpha}_j, b, \boldsymbol{\xi}, \boldsymbol{\zeta}_j} \quad H(\boldsymbol{\xi}, \boldsymbol{\zeta}_j) + \lambda P(\boldsymbol{\alpha}_j), \qquad j = 1, \ldots, k \tag{1}$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{K}_j \boldsymbol{\alpha}_j + \mathbf{1} b = \mathbf{y} + \boldsymbol{\xi} + \boldsymbol{\zeta}_j \qquad j = 1, \ldots, k \\ & \mathbf{1}_i' \boldsymbol{\xi} = 0 \qquad\qquad\qquad\qquad i = 1, 2 \end{aligned} \tag{2}$$

The natural choices for the regularisation function $P(\boldsymbol{\alpha}, \boldsymbol{\beta})$ would either be the l_2-norm of the dual weight vectors, i.e. $P(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\alpha}\|_2^2$, or the l_2-norm of the primal weight vector $P(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\alpha}' \mathbf{K}_a \boldsymbol{\alpha} + \boldsymbol{\beta}' \mathbf{K}_b \boldsymbol{\beta}$. However more interesting is the l_1-norm of the dual weight vector, $P(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1$, as this choice leads to sparse solutions due to the fact that the l_1-norm can be seen as an approximation to the l_0-norm.

We can also follow [9] and remove the assumption of a Gaussian noise model, resulting in different loss functions on the slacks $\boldsymbol{\xi}$. For example, if we choose a Laplacian noise model we can simply replace $\|\boldsymbol{\xi}\|_2^2$ with $\|\boldsymbol{\xi}\|_1$ in the objective function. The advantage of this is if the l_1-norm regulariser from above is chosen, the resulting optimisation is a linear programme, which can be solved efficiently using methods such as column generation.

II. AIMS

The main goal of this experiment is to see whether it is possible to detect the genre of music that a listener is attending to from brain signals. The present experiment is focuses on Magnetoencephalography (MEG), which is an imaging technique used to measure the magnetic fields produced by electrical activity in the brain. The data is from an experiment conducted at the Functional Imaging Laboratory of University College London.

III. METHOD

A. Data Acquisition and Preprocessing

MEG recordings from 2 participants are from a 275-channel CTF system with SQUID-based axial gradiometers at a sampling rate of 1200 *Hz*. Sensors were automatically rejected whose mean power were beyond a static threshold, and trials were rejected in which there was a "sensor jump". The data is filtered using least-squares FIR filters: low pass at 100 *Hz*; notch filter at 49-51 *Hz*. The data is then split into epochs and then downsampled to 200 *Hz*.

B. Design

Stimuli 9 seconds long, with an ISI of 2 seconds during which behavioural responses are collected. The behavioural tasks is identification of genre. Participants presented four blocks of 20 stimuli.

C. Stimuli

The independent variable was the genre of the musical piece, with 4 levels. Each stimulus was 9 seconds in duration, with an inter-stimulus-interval of 2 seconds within which participants gave their responses for the behavioural task. The behavioural task was identification of genre. Participants were presented four blocks of 20 stimuli, with a break between each block. Blocks were randomized to ensure that practice and fatigue effects are accounted for.

The following genres were included in the experiment: *Classical, Jazz, Ragtime, Pop*. In order to avoid confounding factors of spectral or timbral properties of the pieces within each genre being the main criteria of discrimination, all pieces are based on a single instrument, the piano. The stimuli were sourced and selected as MIDI files from various sources, and then rendered to WAVE format using a single instrument and normalized according to peak amplitude. Most of the excerpts in the *Pop* category were solo piano introductions. The experimental stimuli were validated *a-priori* firstly by classification of genre from the MIDI files using the analysis procedures described by [1, 3] and secondly by examination of the behavioural results.

D. Feature Extraction from Audio

Following of from [1, 4], the general approach to genre classification taken was to create a large set of features from the audio, and then use a sparse boosting algorithm (LPBoost) which effectively performs feature selection during the classification stage. Since we are using midi files rather than raw audio, we are able to take advantage of a range of features that are readily derivable from the midi. The features that we use are as follows (numbers in parentheses indicate the dimensionality of that feature):

Table 1: MIDI features used for genre classification

Feature	**Dimensionality**
Meter features	
Tempo	1
Meter	1
Proportion of concurrent onsets	1
Note density	1
Autocorrelation of onset times	33
Melodic features	
Ambitus (melodic range)	1
Tonal features	
Pitch class profiles	12
Distribution of pitch classes (DPC)	12
Krumhansl-Kessler (KK) key estimation	1
Correlation of DPC to KK profiles	24
Mean & standard deviation of KK profiles	2
Statistical features	
Entropy	1
Distribution of note durations	9
Number of notes	1
Total	101

For extraction of the features we used the midi matlab toolbox of Eerola and Toiviainen [10]. These features are then concatenated to produce a single feature vector of length 101.

E. Feature Extraction from Brain Signals

After preprocessing, the data from each trial were split into 3 segments, representing the first, middle and last 3 seconds of each stimulus presentation. Each of these segments were then used as an example for classification. We then performed dimensionality reduction using both Principal Components Analsyis (PCA) and Independent Components Analysis over the channels, two create two sets of 10 "virtual electrodes". The segments were flattened to form a feature vector of length [20×1800] for each example.

IV. RESULTS

A. Classification of Genre by Participants

Table 2 shows the confusion matrix of the behavioural performance of the subjects. The order of the genres is *classical*, *jazz*, *pop*, *rag*. The true labels are on the rows. First we present results of the behavioural task of the participants. The overall error 0.15 (i.e. 85% classification success). Note that for 4 classes a random classifier would achieve 25%, so this is significantly above chance. This appears to validate the stimuli, and is similar to (or above) levels of accuracy reported elsewhere (see [11] for a review).

Table 2: Confusion matrix for classification of genre by participants. True labels are in rows, estimates in columns.

	classical	jazz	pop	rag	Error
classical	48	1	5	6	0.20
jazz	2	51	4	3	0.15
pop	8	1	49	2	0.18
rag	2	2	1	55	0.08
average					0.15

B. Classification of Genre from Audio Features

Using the feature set generated from the midi stimuli, we applied LPBoost [3] using decision stumps as the weak learners as per [1], which results in 6262 weak learners for the algorithm. In order to boost classification performance we split the files into 3 parts, and then took the sum of the classification functions for each of the 3 parts before normalising and classifying. The overall 4-fold cross-validation error is 0.05 (i.e. 95% classification success). This further validates the stimuli, and shows that the methods are appropriate.

Tracing back from the chosen weak learners (of which there were 114/6262), it is possible to see which features were chosen. Interestingly a wide spread of the features were used (52 of the vector of length 101). The only blocks of features not used at all were: *KK key estimation*, *Mean of KK profile*, *Onset autocorrelation*. The key advantage of the LPBoost method is that you can throw as many features as possible at it and it will only pick the useful ones, as it is a sparse method. This means that the same method can be applied to a variety of classification tasks, the algorithm effectively peforming feature selection and classification simultaneously.

Figure 1 shows a spider diagram of the overall confusion matrix resulting from classification of genre using audio. This diagram demonstrates that the performance of the classification algorithm is similar across all four genres, with no particular bias towards confusion between any of the genres. The exception is *rag*, for which the performance is genrally improved. This can be explained by the fact that the genre is generally more homogenous, and also less derivative of the other genres. In each of the other genres examples can be found which are in some way similar to one of the other genres.

2009/spiderconfusion1.png

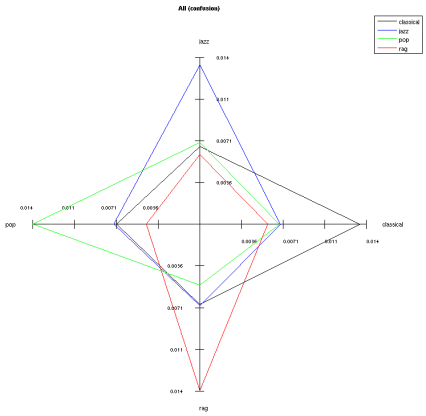

Figure 1: Spider diagram of the overall confusion matrix resulting from classification of genre using audio.

C. Classification of Genre from MEG Features

Using the feature set generated from the MEG data, we constructed linear kernels and applied Kernel Fisher Discriminant Analysis (KFDA) [12]. As with the classification of genre from audio features, we split the files into 3 parts, and then took the sum of the classification functions for each of the 3 parts before normalising and classifying. The overall 4-fold cross-validation error is 0.71 for participant 1 and 0.70 for participant 2 (i.e. 29% and 30% classification success respectively). Note that this is still some way above chance level (25%) but far from reliable.

D. Classification of Genre using both Data Sources

Using the feature sets generated from the MIDI data and the MEG data, we constructed linear kernels and applied Sparse Multiview Fisher Discriminant Analysis (SMFDA) [8]. We used 4-fold cross validation for the selection of parameters. Since we are using the sparse version of multiview FDA, we are able to set the regularisation parameter using a heuristic method to a small value (1e-3) as it has little effect. As with the classification of genre from audio features, we split the files into 3 parts, and then took the sum of the classification functions for each of the 3 parts before normalising and classifying. The overall 4-fold cross-validation error is 0.65 for participant 1 and 0.63 for participant 2 (i.e. 35% and 37% classification success respectively). In itself these classification results are not so impressive, but the side benefit is that we can use the weights of the classifier over the MEG features to then calculate the the brain regions involved in classification of musical genre.

V. CONCLUSIONS

In this study we have show that classification of musical genre from brain signals alone is feasible, but unreliable. We show that though the use of sparse multiview methods, such as SMFDA, we are able to improve the discrimination between different genres.

The procedures [5, 7] both incorporate information from the stimuli themselves to improve classification performance. We extend these through the use of Sparse Multiview Fisher Discriminant Analysis (SMFDA) [8]. The key difference is that SMFDA uses label information to find informative projections. It is also important that the method is sparse, as the MEG data is extremely high dimensional.

The key ingredient in the approach of this paper 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 has enormous promise in a wide range of signal processing and time series data analysis tasks.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the EPSRC grant agreement EP-D063612-1, *"Learning the Structure of Music"*.

References

[1] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and K. Balázs. Aggregate features and ADABOOST for music classification. *Machine Learning*, 65 (2-3):473–484, 2006.

[2] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37:297–336, 1999.

[3] Ayhan Demiriz, Kristin P. Bennett, and John Shawe-Taylor. Linear programming boosting via column generation. *Journal of Machine Learning Research*, 46(13):225–254, 2002.

[4] T. Diethe and J. Shawe-Taylor. Linear programming boosting for classification of musical genre. Technical report, Presented at the NIPS 2007 workshop Music, Brain & Cognition, 2007.

[5] Simon Durrant, David R. Hardoon, Eduardo R. Miranda, John Shawe-Taylor, and André Brechmann. Neural correlates of tonality in music. Technical report, Presented at the NIPS 2007 workshop Music, Brain & Cognition, 2007.

[6] T. Diethe, S. Durrant, J. Shawe-Taylor, and H. Neubauer. Detection of changes in patterns of brain activity according to musical tonality. In *Proceedings of IASTED 2009 Artificial Intelligence and Applications*, 2009.

[7] T. Diethe, S. Durrant, J. Shawe-Taylor, and H. Neubauer. Semantic dimensionality reduction for the classification of EEG according to musical tonality. Technical report, Presented at the NIPS 2008 workshop Learning from Multiple Sources, 2008.

[8] T. Diethe, D.R. Hardoon, and J. Shawe-Taylor. Multiview fisher discriminant analysis. Technical report, Presented at the NIPS 2008 workshop Learning from Multiple Sources, 2008.

[9] S. Mika, G. Rätsch, and K.-R. Müller. A mathematical programming approach to the kernel fisher algorithm. In T.K. Leen, T.G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13, pages 591–597, 2001.

[10] T. Eerola and P. Toiviainen. Midi toolbox: Matlab tools for music research. Jyväskylän yliopisto, ISBN 951-39-1796-7. URL: http://www.jyu.fi/musica/miditoolbox/, 2004.

[11] A. Meng. *Temporal Feature Integration for Music Organisation*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2006. Supervised by Jan Larsen and Lars Kai Hansen, IMM.

[12] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, U.K., 2004.