GENERATION OF STIMULUS FEATURES FOR ANALYSIS OF FMRI DURING NATURAL AUDITORY EXPERIENCES

Valeri Tsatsishvili[1], Fengyu Cong[1,2], Tapani Ristaniemi[1], Petri Toiviainen[3], Vinoo Alluri[1,3], Elvira Brattico[4], Asoke Nandi[5,1]

[1] Department of Mathematical Information Technology, University of Jyväskylä, Finland
[2] Department of Biomedical Engineering, Dalian University of Technology, China
[3] Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland
[4] Cognitive Brain Research Unit, Institute of Behavioral Sciences, University of Helsinki, Finland
[5]Department of Electronic and Computer Engineering, Brunel University, UK

ABSTRACT

In contrast to block and event-related designs for fMRI experiments, it becomes much more difficult to extract events of interest in the complex continuous stimulus for finding corresponding blood-oxygen-level dependent (BOLD) responses. Recently, in a free music listening fMRI experiment, acoustic features of the naturalistic music stimulus were first extracted, and then principal component analysis (PCA) was applied to select the features of interest acting as the stimulus sequences. For feature generation, kernel PCA has shown its superiority over PCA in various applications, since it can implicitly exploit nonlinear relationship among features and such relationship seems to exist generally. Here, we applied kernel PCA to select the musical features and obtained an interesting new musical feature in contrast to PCA features. With the new feature, we found similar fMRI results compared with those by PCA features, indicating that kernel PCA assists to capture more properties of the naturalistic music stimulus.

Index Terms— kernel PCA, ICA, Polynomial kernel, naturalistic music, fMRI

1. INTRODUCTION

Traditionally, fMRI experiments have been conducted in controlled environment where stimulus sequences or onset and offset times are strictly defined. Typically, stimuli are simplified or artificially generated to isolate features of interest as much as possible. It has been questioned whether the results of such controlled experiments are generalisable to much more complex real-world experiences [1-3]. Consequently, interest towards studying brain activations in real world experiences, involving natural continuous stimuli, is quickly growing [1, 2, 4, 5]. In such real-world experimental setups where brain responds to continuous stream of complex stimulus, we need to extract the features to segregate neural responses to various concurrently occurring stimulus events, which might be difficult for certain type of stimuli. Moreover, conventional analysis methods that rely on block or event-related experimental design are not easily applicable in such naturalistic paradigm [3, 6]. Recently, several approaches have been reported that overcome limitations of traditional analysis methods. Hasson et al. [1] proposed pairwise inter-subject correlations and reverse correlation method for analyzing fMRI during free watching movie. From machine learning field supervised classification and regression algorithms were adopted for brain encoding and decoding models (see review in [7]). The encoding model maps stimulus representation to the voxel activity in selected region of interest (ROI) in the brain. Usually, stimulus features in encoding/decoding studies consist of categorical constructs or individual representations of many stimuli. The encoding model is built upon learning the differences between corresponding brain responses to the different categories or stimuli. Trained model, can then predict the voxel activations for a new stimulus. The decoding model has an opposite aim - to predict the stimulus from the voxel activities.

We employed data-driven approach based on independent component analysis (ICA) decomposition of fMRI and correlating temporal courses of the obtained independent components with stimulus features. It should be noted that the method is different from supervised enco-ding/decoding methods mentioned above; it does not need construction of stimulus categories, or to have preliminary assumptions on responses to define regions of interest. Benefits of our approach in the analysis of data obtained from naturalistic experiment have been addressed in [8].

In spite of the availability of the growing body of research on fMRI responses to natural stimuli, most of the studies to our knowledge are focused on visual, virtual reality settings [1, 3, 4, 9, 10] or speech [11]. However, in one recent study brain responses during passive music listening environment were explored in [2]. Authors employed integrated analysis approach involving computational extraction and perceptual validation of stimulus features, and then finding corresponding activations in brain by correlating stimulus features with voxel time courses. The dimensionality of the initially extracted 25 acoustic descriptors was reduced using PCA to obtain compact stimulus representation expressed by six high-level features. As a linear method, PCA is blind to nonlinear inter-relationships between variables, should such relationships exist. This issue was addressed here. To assess possible nonlinear relationships among initial acoustic descriptors, we employed kernel PCA (KPCA) [12] to generate a new set of high-level features. Kernel PCA has been very extensively used for feature selection and dimension reduction in the field of machine learning and has shown its superiority over PCA [13-15]. Therefore, it is worth examining whether KPCA can assist to find better stimulus features from the extracted acoustic descriptors for analyzing fMRI data during real-world experiences.

Usually, one objective in fMRI studies is to find naturalistic stimulus-related brain activations that are consistently present across different participants' responses. In our paradigm, the objective translates in finding similar ICA components (spatial maps) among subjects such that the time courses of these components are significantly correlated with the time courses of stimulus features. This was the main evaluation criterion for comparing KPCA and PCA performances in search of better stimulus representation.

2. DATA DESCRIPTION

The dataset analyzed in this study consists of fMRI scans of eleven healthy musicians (mean age: 23.2; SD: 3.7; 5 females), who listened to a 512 second-long piece of modern tango Adios Nonino by Astor Piazzolla. The fMRI measurements were made in 3T scanner at sampling frequency of 0.5 Hz. Obtained fMRI scans went through conventional preprocessing routine. Detailed description of preprocessing steps can be found in [2].

The preprocessed fMRI data were first band-pass filtered with FFT-based digital filter. Benefits of applying such filter to fMRI data is discussed in [8]. The pass-band was set between 0.008 and 0.1 Hz [8]. Lower limit of the pass-band was in accordance to the filter applied during the preprocessing, while higher limit was set to match the frequency range where most of the power of acoustic features was contained. Overall, 231 fMRI scans corresponding to stimulus between 21 to 480 seconds were used for analysis.

Next, PCA and model order selection method SORTE [16] was applied to further remove noise and estimate number of sources. Selected PCA components of each participant were decomposed using independent component analysis (ICA). FastICA [17] was employed as part of the ICASSO [18] software package, which addresses the stability of ICA decomposition. For each subject 94 independent components (ICs) were obtained. From the temporal courses of all the ICs, those significantly correlated ($p < 0.01$) with musical features were selected for further analysis. Significance thresholds for correlations were set for each feature via Monte Carlo simulation [2].

From the set of selected components we rejected those with normalized kurtosis less than 5 to avoid artifacts. As a result, for each stimulus feature a set of significantly correlated spatial maps from each subject were obtained. Finally, six sets of spatial maps corresponding to each feature were clustered separately to find common activations among different subjects. We employed diffusion map first to reduce dimensions and then clustered data using simple spectral clustering. Detailed description of this method is provided in [19]. Two clusters are usually produced where similar activation maps (common map) from different subjects formed one dense cluster, whereas dissimilar maps formed sparse cluster. The features were considered as interesting if the common map in associated dense cluster included contribution from more than five (half of all) participants.

Described analysis scheme was employed twice - with PCA and KPCA descriptors, and therefore two large sets of BOLD responses corresponding to each feature set was obtained.

3. STIMULUS FEATURE EXTRACTION AND SELECTION

3.1 Acoustic feature preprocessing

The feature extraction procedure follows already well-established window-based extraction scheme employed in music information retrieval [20, 21]. Overall, 25 features representing timbral, tonal, and rhythmic information were extracted from the stimulus. The features were extracted from the overlapping windows of two different lengths. The shorter window length of 25 ms with 50% overlap was selected for so called low-level features capturing timbral characteristics of the sound. These features usually are of high temporal resolution and reflect fast changes in music. The longer 3s windows with 67% overlap were employed for features that depict higher level concepts in music, such as tonality and rhythm. Hereafter we will refer these two subsets as short-term and long-term features based on the window length employed for their extraction. For the features and their descriptions, the reader is referred to [2].

The features were centered and normalized with respect to their standard deviation, after which long term features were up-sampled to match the sampling rate of short-term features. Next, all features were convolved with double gamma HRF (hemodynamic response function) to consider the hemodynamic lag. Following the convolution, 21 to 480 seconds were extracted from feature time courses to synchronize with fMRI scans. The final step of the preprocessing was the high-pass filtering with cutoff frequency at 0.008 Hz, in accordance with the low cutoff of band-pass filter applied on fMRI voxel series.

3.2 PCA-based musical features

PCA is a widely used method to reduce dimensionality [22]. It is an orthogonal transformation of the centered matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, where d is the number of dimensions and n is the number of samples. This is achieved by solving the following eigenvalue problem:

$$\mathbf{CU} = \boldsymbol{\lambda}\mathbf{U} \tag{1}$$

where $\mathbf{C}$ is a covariance matrix:

$$\mathbf{C} = \frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X} \tag{2}$$

Eigenvectors $\mathbf{U} \in \mathbb{R}^{d\times d}$ of (1) represent the directions to largest variances sorted in decreasing order and eigenvalues $\boldsymbol{\lambda}$ are variances across eigen-directions. The common heuristic to reduce dimensions is to select the first l eigenvectors explaining most of the variance (usually 95%) in the data. Finally, data are projected onto the principal components to get its representation in the principal component space:

$$\mathbf{Y} = \mathbf{U}^{\mathrm{T}}\mathbf{X} \tag{3}$$

This scheme was used for reducing dimensions of 25 preprocessed features. Initially nine principal components were selected explaining 95% of variance in the data. The principal component axes were rotated using varimax rotation [2]. Perceptual labels of principal components were applied based on loadings from raw features. The perceptual labels were validated through the experiment where 21 musicians rated the excerpts of the stimuli in which the labels were exhibited in varying degrees. Finally, a set of six features including Activity, Fullness, Brightness, Timbral Complexity, Key Clarity and Pulse Clarity were selected for further analysis. First four of the six features characterize polyphonic timbre of music. Key Clarity represents tonal clarity, and Pulse Clarity is an estimate of clarity of perceived pulse [2].

3.3 Kernel PCA features

Kernel PCA is a nonlinear extension of PCA for nonlinear data distributions where mapping into linear subspace is not useful [12, 23]. To introduce kernel PCA, let us consider the data matrix consisting of n column vectors with d dimensions: $\mathbf{X} \in \mathbb{R}^{d\times n}$, The basic way to do nonlinear extension of PCA is to introduce nonlinear mapping to a (generally) higher dimensional feature space $\mathcal{F}$:

$$\mathbf{X} \rightarrow \boldsymbol{\Phi}(\mathbf{X}) \tag{4}$$

Then calculate covariance using inner product $\boldsymbol{\Phi}(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{X})$ in $\mathcal{F}$, and apply linear PCA as described above. Usually, this will quickly blow up computational complexity with increasing dimensionality of the data. It is possible to avoid mapping (4) by introducing kernel function: $\mathbf{K} = \boldsymbol{\Phi}(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{X})$, $\mathbf{K} \in \mathbb{R}^{n\times n}$, which replaces the inner product in feature space. It can be shown that eigenvectors of covariance matrix in $\mathcal{F}$ can be represented as linear combinations of data vectors: $\mathbf{V} = \sum_{i=1}^{n} a_i \boldsymbol{\Phi}(\mathbf{x}_i)$. Coefficients a_i can be found to solving the following eigenvalue problem:

$$\widetilde{\mathbf{K}}\mathbf{a} = \tilde{\lambda}\mathbf{a} \tag{6}$$

where $\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{1_N}\mathbf{K} - \mathbf{K}\mathbf{1_N} + \mathbf{1_N}\mathbf{K}\mathbf{1_N}$ is a Gram matrix that is used for centering the kernel matrix, $\tilde{\lambda}_k, \boldsymbol{a}^k,\ k = 1, \dots n$ represent k-th eigenvectors and eigenvalues. The projections of points in the feature space $\boldsymbol{\Phi}(\mathbf{X})$ onto the eigenvectors are given by:

$$\mathbf{Y} = \mathbf{V}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{X}) = \sum_{i=1}^{n} \boldsymbol{a}_i\, \mathbf{K}(\mathbf{x}_i\,, \mathbf{x}) \tag{7}$$

A polynomial kernel of third degree was selected in this study:

$$\mathbf{K} = (a\mathbf{X}^{\mathrm{T}}\mathbf{X} + b)^{\mathbf{3}}$$

where $\mathbf{K} \in \mathbb{R}^{n\times n}$ is the kernel matrix and $\mathbf{X} \in \mathbb{R}^{n\times d}$ matrix of features. For simplicity we set the slope parameter a to 1. To select b, we tested the method on several sample values from wide range. Thus, the final form of our polynomial kernel was: $\mathbf{K} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + 1)^{\mathbf{3}}$. We selected the first 14 eigenvectors explaining 95% of variance in the data to reduce dimensions. Hence, 14 new features were obtained.

4. MUSICAL FEATURES AND FMRI DATA ANALYSIS

Overall, 14 kernel PC scores were generated from the initial set of 25 features. We explored similarities between kernel and linear PCA features by finding Pearson correlations between their temporal courses. For the simplicity, we will refer to kernel PC scores as 'new features' and linear PC scores as 'old features' hereafter. Several new features showed moderate to moderately high correlations with old features, while some features were very weakly correlated with linear PC scores (Fig. 1). Therefore, polynomial kernel was able to find new stimulus features that are moderately or not at all correlated with the old ones.

Another interesting fact was that few combinations of the old features were submerged into first few KPCA features (e.g. Fullness, Brightness, and Activity are represented with different weights in the first KPCA feature). It can be explained by existence of inter-correlations between mentioned PCA features, introduced by varimax rotation applied on principal components (see section 3.2). For example, absolute value of the correlation between Fullness and Activity is 0.92. As described in section 2, the fMRI data were analyzed for two cases, involving six old and 14 new feature sets as stimulus sequences. The features for which we failed to find significantly correlated (associated) ICs from more than half of the participants were eliminated from further analysis. After the elimination, four old (Fullness, Brightness, Timbral Complexity, and Activity) and three new features (features 2, 3, and 12) were left for further analysis. Next, we were interested in finding common spatial activation maps among ICs associated with each of the selected features from each set. To this end, we applied diffusion maps and spectral clustering [19].

For the old features, two common spatial maps were revealed by spectral clustering of the associated components for Brightness and Activity. Both common maps showed large bilaterally activated areas predominantly within auditory cortices. For Brightness, the common map was obtained from ten subjects and for Activity - from seven subjects. For the remaining two features common maps were not observed [8].

The common map was also found for one new feature KPCA #3. The common map showed the same activity patterns as found by PCA-based musical features, but was observed in eight subjects' ICs. Furthermore, the set of ICs showing common activation maps were subset of the ones selected by PCA features. In other words, there is an intersection between the sets of ICs corresponding to each of the feature set, while the two musical features by PCA and kernel PCA are not very similar. The temporal courses of KPCA #3, Brightness, and Activity are depicted in Fig.2. The common map consisting of averaged eight components from eight subjects is shown in Fig. 3.

We also tested Gaussian kernel for KPCA. For kernel parameters outside certain range, for which we could select reasonable amount of eigenvectors, generated features were highly correlated with old features in somewhat similar pattern as in Fig. 1. However, we did not find the common map for those features. Due to the space limitation, we do not report results of Gaussian kernel in this paper.

5. DISCUSSION

We aimed to exploit possible nonlinear relationships among initial set of descriptors by employing kernel PCA with third degree polynomial kernel. The set of generated KPCA features were employed in our individual ICA-based framework to analyse real fMRI dataset from free listening experiment. We found similar brain responses as with previously used PCA stimulus sequences and the same fMRI data. It should be noted that the analysis framework presented in this study was tested and shown to be producing results in agreement with previous findings from the same dataset, obtained from other established models [8].

Two interesting points can be highlighted from the results: First, one of the three selected KPCA features, namely KPCA #3, was highly correlated with the temporal courses of spatial maps from majority of subjects. Considering only moderate-level correlations between this new feature and Brightness (Fig.1), both showing significant correlation with the same brain responses is an interesting finding. It indicates PCA features as a representation of the auditory stimulus are not the unique solution, and the mapping between initial descriptors and stimulus representation can be nonlinear. Second, two KPCA features exhibited contributions from several PCA features. Such aggregation

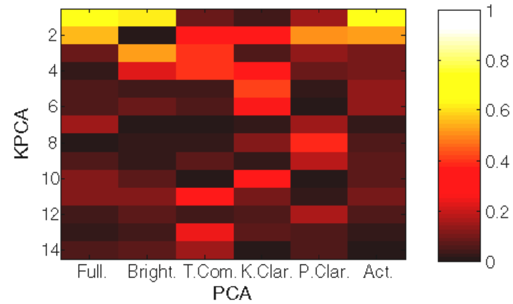

Fig. 1. Correlation coefficients (absolute value) between KPCA and PCA features.

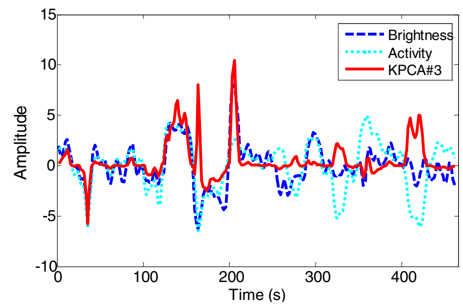

Fig. 2. Temporal courses of Brightness, Activity and KPCA#3.

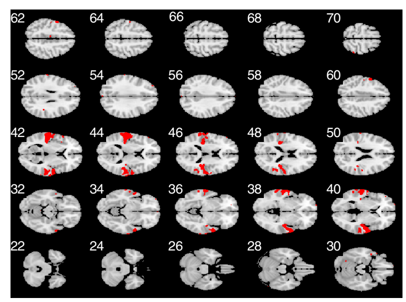

Fig. 3. Common spatial map based on the new feature, KPCA#3.

seems to be useful as it enables compact representation and, considering the existing inter-correlations between the old features, reduces redundancies.

From the perspective of finding stimulus-related consistent brain responses, both sets of features produced comparable results. Both PCA and KPCA found the same common activations.

To summarize, despite the fact that KPCA did not clearly outperform PCA in our exploratory study, finding new stimulus representation that correlates well with brain responses is a positive result. Moreover, not finding common map for most of the features might point to some limitations of our analysis method. Indeed, issues with our method and group ICA-based methods for analysing fMRI in naturalistic settings are discussed in [8]. This motivates us to explore kernel-based or other non-linear methods for finding stimulus representation further (e.g. see [24]). To overcome possible limitations introduced by model selection and to validate our results, we plan to compare generated features within different analysis method.

Acknowledgement: This work was financially supported by TEKES (Finland) under grant 40334/10 "Machine Learning for Future Music and Learning Technologies", and Jyväskylä Doctoral Program in Computing and Mathematical Sciences. A. K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

6. REFERENCES

[1] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science,* vol. 303, pp. 1634-1640, Mar 12, 2004.

[2] V. Alluri, P. Toiviainen, I. P. Jaaskelainen, E. Glerean, M. Sams and E. Brattico, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *Neuroimage,* vol. 59, pp. 3677-3689, Feb 15, 2012.

[3] H. J. Spiers and E. A. Maguire, "Decoding human brain activity during real-world experiences," *Trends Cogn. Sci. (Regul. Ed.),* vol. 11, pp. 356-365, 2007.

[4] K. N. Carvalho, G. D. Pearlson, R. S. Astur and V. D. Calhoun, "Simulated Driving and Brain Imaging," *CNS Spectr,* vol. 11, pp. 52-62, 2006.

[5] J. P. Kauppi, I. P. Jääskeläinen, M. Sams and J. Tohka, "Intersubject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency," *Frontiers in Neuroinformatics,* vol. 4, 2010.

[6] J. Ylipaavalniemi, E. Savia, S. Malinen, R. Hari, R. Vigário and S. Kaski, "Dependencies between stimuli and spatially independent fMRI sources: Towards brain correlates of natural stimuli," *Neuroimage,* vol. 48, pp. 176-185, 2009.

[7] T. Naselaris, K. N. Kay, S. Nishimoto and J. L. Gallant, "Encoding and decoding in fMRI," *Neuroimage,* vol. 56, pp. 400-410, 2011.

[8] F. Cong, T. Puoliväli, V. Alluri, T. Sipola, I. Burunat, P. Toiviainen, A. K. Nandi, E. Brattico and T. Ristaniemi, "Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis," *J. Neurosci. Methods,* vol. 223, pp. 74-84, 2014.

[9] A. Bartels and S. Zeki, "Brain dynamics during natural viewing conditions—a new guide for mapping connectivity in vivo," *Neuroimage,* vol. 24, pp. 339-349, 2005.

[10] K. N. Kay, T. Naselaris, R. J. Prenger and J. L. Gallant, "Identifying natural images from human brain activity," *Nature,* vol. 452, pp. 352-355, 2008.

[11] E. Formisano, F. De Martino, M. Bonte and R. Goebel, ""Who" is saying "what"? Brain-based decoding of human voice and speech," *Science,* vol. 322, pp. 970-973, Nov 7, 2008.

[12] K. Muller, S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf, "An introduction to kernel-based learning algorithms," *Neural Networks, IEEE Transactions On,* vol. 12, pp. 181-201, 2001.

[13] B. Scholkopf, S. Mika, C. J. Burges, P. Knirsch, K. Muller, G. Ratsch and A. J. Smola, "Input space versus feature space in kernel-based methods," *Neural Networks, IEEE Transactions On,* vol. 10, pp. 1000-1017, 1999.

[14] T. Mu, A. K. Nandi and R. M. Rangayyan, "Classification of breast masses via nonlinear transformation of features based on a kernel matrix," *Med. Biol. Eng. Comput.,* vol. 45, pp. 769-780, 2007.

[15] T. Mu, A. K. Nandi and R. M. Rangayyan, "Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers," *J. Digital Imaging,* vol. 21, pp. 153-169, 2008.

[16] Zhaoshui He, A. Cichocki, Shengli Xie and Kyuwan Choi, "Detecting the Number of Clusters in n-Way Probabilistic Clustering," *Pattern Analysis and Machine Intelligence, IEEE Transactions On,* vol. 32, pp. 2006-2021, 2010.

[17] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.,* vol. 9, pp. 1483-1492, 1997.

[18] J. Himberg, A. Hyvarinen and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage,* vol. 22, pp. 1214-1222, Jul, 2004.

[19] T. Sipola, F. Cong, T. Ristaniemi, V. Alluri, P. Toiviainen, E. Brattico and A. K. Nandi, "Diffusion map for clustering fMRI spatial maps extracted by independent component analysis," in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP-2013),* Southampton, UK, 22-25 Sep, 2013.

[20] N. Scaringella, G. Zoia and D. Mlynek, "Automatic genre classification of music content: a survey," *Signal Processing Magazine, IEEE,* vol. 23, pp. 133-141, 2006.

[21] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE Transactions On,* vol. 10, pp. 293-302, 2002.

[22] I. T. Jolliffe, *Principal Component Analysis* . New York: Springer New York, 2002.

[23] B. Scholkopf, A. Smola and K. Müller, "Kernel principal component analysis," in *Advances in Kernel Methods-Support Vector Learning,* 1999,

[24] H. Guo, L. B. Jack and A. K. Nandi, "Feature generation using genetic programming with application to fault classification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions On,* vol. 35, pp. 89-99, 2005.