JYVÄSKYLÄ STUDIES IN COMPUTING
268

Valeri Tsatsishvili

Data-driven Analysis for fMRI during Naturalistic Music Listening

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN COMPUTING 268

Valeri Tsatsishvili

Data-driven Analysis for fMRI during Naturalistic Music Listening

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteellä, salissa YlistöKem4
marraskuun 24. päivänä 2017 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in Ylistönrinne, auditorium YlistöKem4, on November 24, 2017 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2017

Data-driven Analysis for fMRI during Naturalistic Music Listening

JYVÄSKYLÄ STUDIES IN COMPUTING 268

Valeri Tsatsishvili

Data-driven Analysis for fMRI during Naturalistic Music Listening

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2017

Editors
Timo Männikkö
Faculty of Information Technology, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-7240-0

URN:ISBN:978-951-39-7240-0
ISBN 978-951-39-7240-0 (PDF)

ISBN 978-951-39-7239-4 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2017

ABSTRACT

Tsatsishvili, Valeri
Data-driven Analysis for fMRI during Naturalistic Music Listening
Jyväskylä: University of Jyväskylä, 2017, 51 p. (+included articles)
(Jyväskylä Studies in Computing
ISSN 1456-5390; 268)
ISBN 978-951-39-7239-4 (nid.)
ISBN 978-951-39-7240-0 (PDF)
Finnish summary
Diss.

Interest towards higher ecological validity in functional magnetic resonance imaging (fMRI) experiments has been steadily growing since the turn of millennium. The trend is reflected in increasing amount of naturalistic experiments, where participants are exposed to the real-world complex stimulus and/or cognitive tasks such as watching movie, playing video games, or listening to music. Multifaceted stimuli forming parallel streams of input information, combined with reduced control over experimental variables introduces number of methodological challenges associated with isolating brain responses to individual events.

This exploratory work demonstrated some of those methodological challenges by applying widely used data-driven methods to real fMRI data elicited from continuous music listening experiment. Under the general goal of finding functional networks of brain regions involved in music processing, this work contributed to improvement of the methodology from two perspectives. One is to produce a set of representative features for stimulus audio that can capture different aspects of music, such as timbre and tonality. Another is to improve reliability and quality of separation of the observed brain activations into independent spatial patterns. Improved separation in turn enables better differentiation of stimulus-related activations from the ones originating from unrelated physiological, cognitive, or technical processes.

More specifically, part of the research explored an application of a nonlinear method for generating perceptually relevant stimulus features representing high-level concepts in music. Another part addressed dimensionality reduction and model order estimation problem before subjecting fMRI data to source separation and offered few methodological developments in this regard.

Keywords: fMRI, naturalistic experiment, ICA, CCA, PCA, kernel PCA dimension reduction

Author's address Valeri Tsatsishvili
Faculty of Information Technology
University of Jyväskylä
valeri.tsatsishvili@jyu.fi

Supervisors Professor Tapani Ristaniemi
Faculty of Information Technology
University of Jyväskylä
Finland

Professor Fengyu Cong
Department of Biomedical Engineering
Dalian University of Technology
China

Reviewers Professor Tarmo Lipping
Tampere University of Technology
Finland

Professor Dezhong Yao
University of Electronic Science and Technology
China

Opponent Professor Karen Egiazarian
Tampere University of Technology
Finland

ACKNOWLEDGEMENTS

I would like to acknowledge support of my supervisors Fengyu Cong and Tapani Ristaniemi. Fengyu Cong helped at the initial stage to formulate the research questions. I have learnt what is needed to be a good academic in the modern world. Thanks to Tapani Ristaniemi for his trust and support in funding my studies, especially first year.

One 'thank you' goes to the graduate school COMAS for funding the major part of my work. While inconsistent funding throughout my PhD has not really helped with motivation, it definitely sharpened my funding application writing skills and taught me to keep focus under a lot of pressure. Thanks also to the other sources financially supporting the remining part of my studies.

I am truly grateful to my friend and colleague Tuomas Puoliväli. Thank you Tuomas for countless hours of discussions about work, science, and whatnot. Directly or indirectly you contributed in addressing some of the challenges I was facing in my project.

I wish to thank all my co-authors, especially Iballa Burunat and Petri Toiviainen. Ibi, it was great fun working with you and have all those inspiring discussions. Petri, interacting with you throughout my PhD (and beyond) contributed much to my understanding of how to be a good scientist.

Thank you, Nino and Matti, for being always happy to provide an accommodation and even host the dinners for me and my guests.

Finally, warm thanks to my relatives and family. Those who made me who I am, and who have been along for the ride of this long and winding road.

27.10.2017
Valeri Tsatsishvili

LIST OF ABBREVIATIONS

fMRI	Functional magnetic resonance imaging
BOLD	Blood oxygen level dependent
PET	Positron emission tomography
HRF	Hemodynamic response function
GLM	General linear model
ROI	Region of interest
ICA	Independent component analysis
GICA	Group independent component analysis
MOS	Model order Selection
PCA	Principal component analysis
KPCA	Kernel principal component analysis
CCA	Canonical correlation analysis
MCCA	Multiset canonical correlation analysis
SNR	Signal to noise ration

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

PI Tsatsishvili, V., Cong, F., Puoliväli, T., Alluri, V., Toiviainen, P., Nandi, A. K., Brattico, E., & Ristaniemi, T. (2013). Dimension reduction for individual ICA to decompose FMRI during real-world experiences: Principal component analysis vs. canonical correlation analysis. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium. 137.

PII Tsatsishvili, V., Cong, F., Ristaniemi, T., Toiviainen, P., Alluri, V., Brattico, E., & Nandi, A. (2014). Generation of stimulus features for analysis of FMRI during natural auditory experiences. Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European, 2490-2494.

PIII Tsatsishvili, V., Cong, F., Toiviainen, P., & Ristaniemi, T. (2015). Combining PCA and multiset CCA for dimension reduction when group ICA is applied to decompose naturalistic fMRI data. Neural Networks (IJCNN), 2015 International Joint Conference On, 1-6.

PIV Burunat, I., Tsatsishvili, V., Brattico, E., & Toiviainen, P. (2017). Coupling of action-perception brain networks during musical pulse processing: Evidence from region of interest based independent component analysis. Frontiers in Human Neuroscience, 11, 230.

PV Tsatsishvili, V., Burunat, I., Cong, F., Alluri, V., Toiviainen, P., & Ristaniemi, T. (2017). On Application of Kernel PCA for Generating Stimulus Features for fMRI during Continuous Music Listening. Submitted to: Journal of Neuroscience Methods.

1 INTRODUCTION

"The human brain has 100 billion neurons, each neuron connected to 10 thousand other neurons. Sitting on your shoulders is the most complicated object in the known universe" - Michio Kaku

Mankind's fascination towards human mind can be found throughout all its documented history. According to timeline of neuroscience provided in (Kandel & Squire, 2000), first steps in neuroscience dates back to 2nd century AD when brain was identified as an 'organ of mind'. The field, however, did not excel as well as some other areas of science over the centuries. Fast forward almost two millennia to 19th century, and first strong evidence demonstrating localization of different mental processes in different areas of brain emerge from the lesion studies. Around the same period, first precursors leading to the development of functional brain imaging appeared (Raichle, 2009). Invention of modern functional brain imaging techniques including positron emission tomography (PET) and magnetic resonance imaging (MRI) in the 20th century marked the new era in modern cognitive neuroscience as these methods opened the window in living and functioning human brain.

Since its inception in early 1990s, functional magnetic resonance imaging (fMRI) quickly gained popularity due to superior temporal (1-3s) and spatial (1-3mm) resolution, safety (no harmful exposure of ionizing radiation), and relatively lower costs of experiment than PET. Soon after the first fMRI experiments on human subjects, as Hasson & Honey (Hasson & Honey, 2012) puts it, 'age of exploration' started to map different cognitive functions to cortical and subcortical regions of the brain. This traditional view on localized cognitive processing has been gradually updated by more modern 'distributed processing' view, suggesting that a network of brain regions is involved in most of the cognitive processes, and given anatomical region can have more than one function. As will be discussed later, traditional experimental design and analysis methodology applied in fMRI have limitations that in some cases hinder exploration of distributed processing. Naturalistic experiments are more recently emerged alternative to controlled experiments, aimed towards increasing an ecological validity. In

the naturalistic experiments, brain responses are acquired while participants are performing complex cognitive or behavioral task they would do in real life. Such experiments can provide all needed leverages to observe and understand how different facets of complex sensory input are integrated and processed in the brain. This work explores an application of data-driven analysis methods to fMRI data from naturalistic music listening experiment in search of music-related brain activations, and focuses on certain aspects in analysis workflow.

1.1 Structure of the Work

The rest of the thesis is organized as follows. The remainder of this chapter describes origins of the fMRI signal, common experimental designs in fMRI and main methodological approaches. Chapter 2 overviews the methods employed in the included publications. Chapter 3 provides summaries of the included articles, and chapter 4 discusses the main contributions and limitations of this work.

1.2 From Neurons to BOLD to fMRI

Magnetic resonance imaging utilizes variations in magnetic properties of different brain tissues, such as gray matter and white matter, to generate highly detailed anatomical images of the brain. More specifically, large electromagnet in the MRI scanner produces very intense magnetic field that forces nuclei of all hydrogen atoms to align their spins to the same direction. An electromagnetic pulse excites the nuclei, causing their spins to temporarily tilt away from the magnetic field. The spins will eventually return to equilibrium and during this 'relaxation' process, the hydrogen atoms emit electromagnetic wave. Emitted waves are detected by the receiver in the scanner and relaxation time is measured. Since concentration of hydrogen varies in different structures, contrasting between different tissues is possible based on the relaxation times. While very useful for providing high resolution anatomical images, MRI does not provide information on dynamics of brain activity.

Functional MRI relies on similar principles as MRI, but exploits the difference between magnetic properties of oxygenated versus deoxygenated blood, and repeated electromagnetic pulses to detect changes in blood oxygenation across different brain regions over time. More detailed explanation of physiological mechanisms and operating principles of MRI and fMRI can be found in (Huettel, Song, & McCarthy, 2004). What will suffice to know for the purpose of this thesis is that fMRI measures blood-oxygen-level-dependent (BOLD) signal.

How blood oxygenation relates to neuronal activity? Oxygen is one of the main ingredients involved in the cell metabolism, hence, a proper function of our every organ depends on its undisrupted supply. The supplier is cardiovascular system and the transport is blood. When the group of neurons start firing, there

is a local surge of blood providing oxygen and other vital ingredients. Based on the changes in the oxygen consumption in each voxel (a volumetric unit of the brain image, a three-dimensional equivalent of a pixel), activation state of it can be inferred.

This relationship between neural activity and BOLD is known as neurovascular coupling. The inner workings of process are still not fully understood, yet, progress has been made over the past decade (Mukamel et al., 2005; Nir et al., 2007). The mathematical function characterizing dynamics of neurovascular coupling is called hemodynamic response function (HRF). While HRF can considerably differ between persons, in a generic form it is approximated by two combined gamma functions. (Friston, Jezzard, & Turner, 1994). Hemodynamic response is significantly slower than neuronal response. Accordingly, temporal resolution of fMRI relying to this indirect measure is lower than electrophysiological methods such as Electroencephalography (EEG) directly measuring neuronal activity. It is generally agreed that the hemodynamic response is linear and time invariant (Boynton, Engel, Glover, & Heeger, 1996; Dale & Buckner, 1997). In other words, superposition, scaling and time shifting the neural response will be linearly related to corresponding changes in BOLD signal. This enables convolving stimulus time courses to HRF in order to account for the lag and smoothing effect of BOLD to neuronal activity.

1.3 Experimental Design in fMRI

Experimental design in fMRI studies is mainly bound to the temporal structure of stimulus presentation or the behavioral task performed in the MRI scanner. Block design and event-related design are most common types. In the block design experiments, stimulus trials are grouped by experiment condition and are presented as stimulus 'blocks' lasting 30-60 seconds. Entire experiment is structured as alternating blocks of different conditions. The simplest example would be a block of experimental condition (Task) followed a block of control condition (Rest). Then such experiment can have 'Task-Rest-Task-Rest…' structure, and the temporal course of stimulation can be characterized by boxcar function with ones representing task condition and zeros - rest. Stimulus trials are spaced less than duration of HRF in time, leading HR to individual trials within the block to overlap. As a result, one long plateau of response in each block is produced. Such structure allows testing hypothesis about contrasting response during one block with respect to another (D'esposito, Zarahn, & Aguirre, 1999).

Stimulus trials are not grouped in the event-related design. Instead, individual stimulus trials or 'events' can be presented in an arbitrary order. This enables examining signal changes associated with individual events rather than blocks of trials. An intuitive explanation of application areas for each design is provided in (Poldrack, Mumford, & Nichols, 2011). Authors draw an interesting analogy between a trial in experiment and a granule of sand. A pile of sand is easy to detect from the distance, whereas spread around it is difficult to locate.

Similarly, in block design experiment, trials are grouped as the grains in a pile of sand and thus response to the stimulus block is easy to detect. However, it is difficult to model responses to individual trials. On the other hand, as closer look at spread out sand provides better information on the shape of individual granules, event-related design allows modelling responses to individual trials. Naturally then, choice of the experiment design depends on the goal of the experiment. Both block and event-related designs allow repeatedly presenting stimulus trials and averaging responses across trials. This is an important advantage, because BOLD signal is much weaker compared to the artifacts and noises originating from the imaging hardware as well as from the human body. Various physiological processes, including muscle activities related to heartbeat and respiration, blood pulses in veins, or simply slight head movements are strong contaminators of fMRI data. Averaging across trials eliminates much of the random sources and increases signal-to-noise ratio (SNR). Mixed design experiments have also been conducted where block and event-related paradigms are combined within one experiment.

Another type of fMRI experiments examines intrinsic brain activity when person is not engaged in any goal-oriented task. In these so called resting state experiments, brain images of a person resting in the scanner (eyes closed or simply fixating to a blank screen) are obtained. Brain activity in resting state, also called default mode network has been subject of growing interest since turn of the millennium. An extensive discussion about its importance and applications can be found in (Raichle & Snyder, 2007).

Finally, there is an evident increase of interest towards of naturalistic experiments in neuroimaging. Naturalistic experiments aim to increase ecological validity at the expense of less control over independent variable. In fMRI, this usually means that stimulus timing or behavioral task are not controlled and complex real-world input (such as motion picture, or music) are not simplified to isolate variables of interest. Complex stimulation makes it very difficult to separate stimulus-related responses from the ones originating from other sources due to and the fact that trial-based averaging is frequently not possible, and SNR is low. Nevertheless, there is an ongoing discussion and emerging evidence that traditional controlled experiments with simplified stimulation might not always reveal complete information on how multiple streams of information – as experienced in real world- are processed and integrated in the brain. Hasson & Honey (Hasson & Honey, 2012) point out that in traditional block, or event-related experiments individual events are assumed to be processed in isolation. Authors argue that processing of complex continuous phenomena is not necessarily a superposition of separate event processing. This is relevant to music as well. In order to isolate the attribute of interest, music stimulus is usually simplified (e.g. melody can be composed by pure sine tones rather than by musical instrument to eliminate timbre as a variable). While understanding neural processing mechanisms of separate musical attributes is important, in real life they are rarely processed in isolation. Indeed, in line with (Hasson & Honey, 2012), growing evidence suggests that the functional network involved in music processing is not a

simple integration over networks associated to isolated attributes (Abrams et al., 2013; Alluri et al., 2012). Therefore, studying neural correlates of musical features in naturalistic music listening context can extend existing knowledge on how brain processes music, which will have implications in range of fields outside neuroscience, including therapy and education.

To summarize, early stages of neuroimaging research was focused on localization of brain functions with strictly controlled paradigms to maintain control over variables of interest. In light of accumulated theoretical and empirical foundations from those studies, there is a growing interest towards how networks of brain regions integrate information in more natural environment. It should be noted that the term 'functional network' in this work refers to the set of active regions in response to stimulus, and is not related to functional connectivity.

1.4 Methodology: A Brief Overview

Before analysis, fMRI images are typically subjected to preprocessing routine. Complete preprocessing pipeline might considerably vary according to specific needs of the collected data, but typical routines include: timing correction for separate brain slices, compensating for mismatch of voxels in different images due to the head movement, spatial normalization to account for size and shape differences between brains, and spatial smoothing to increase SNR. More detailed description of preprocessing can be found in any fMRI textbook, e.g. (Poldrack et al., 2011).

Methodology for fMRI analysis has been expanding and improving to increase signal-to-noise ratio, provide more reliable statistical inference, and to address specific needs of new experimental designs. Existing methods have been categorized in multiple ways in the literature: based on statistical approach they use, application areas, etc. Exhaustive review would require a separate dissertation. Here, I briefly overview more established as well as relatively new approaches in functional neuroimaging community.

The most common methodological approach has been statistical parametric mapping, where activation maps are built by testing the hypothesis about temporal dynamics of each voxel separately. The simplest method is correlation analysis between individual voxel time courses in hypothesized ROI and stimulus time course (typically boxcar function representing task and rest conditions), convolved with HRF (Bandettini, Jesmanowicz, Wong, & Hyde, 1993). Subsequently, statistically significant correlations are selected, and a binary spatial map is produced where active voxels (significant correlations) will be visible.

More advanced method is general linear model (GLM). GLM is essentially multiple regression in which measured voxel time series are represented as linear combination of predictor variables that are hypothesized to be contributing factors, and some residual error. Advantage of this method over simple correlation described above is that here multiple variables contributing to the observed activations can be considered in the model. If the measured data matrix is denoted

by $\mathrm{Y} \in \mathrm{R}^{t \times v}$ with t time points and v voxels, and $\mathrm{X} \in \mathrm{R}^{t \times n}$ is the matrix containing n predictor time courses, then the measured responses can be modeled by:

$$\mathrm{Y} = \mathrm{BX} + e$$

where B are weights and indicate contribution of each predictor to a given voxel, and e is some residual error. The weights of the predictor variables are optimized to model each voxel response accurately (i.e. minimize e) in least square sense. The obtained weights are then tested for statistical significance, typically with standard statistical parametric tests, such as t-test. In statistical parametric mapping and related methods, each voxel is treated as an independent measure and statistical testing is repeated across all voxels. Such methods are understandably referred as 'massive univariate analysis'. Shortcoming of massive univariate approach, however, is that it overlooks existing interaction between voxels and raises multiple comparison issues. Furthermore, prior hypothesis on contributing predictor variables and their temporal dynamics is important for the outcome of the model (this is where another label 'hypothesis-driven' stems from). Experiments, where temporal structure of the independent variable is known, are main application area for GLM and other hypothesis-driven methods.

In naturalistic and resting state experiments, as well as in many clinical studies, it is usually difficult to build a regressor time courses due to the complex nature of the input, or complete absence of it. Data-driven methods are more suitable alternative for such cases. These are typically multivariate exploratory methods, that do not require any prior knowledge of temporal models for stimulus or noise. Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000; McKeown & Sejnowski, 1998) is arguably the most commonly applied data-driven method in functional neuroimaging. Typical ICA approach decomposes fMRI responses in statistically independent spatial activation patterns and corresponding temporal courses with no prior information about the data or stimulus, apart from a few statistical assumptions. It has been evidenced in numerous applications that ICA is very powerful method for uncovering brain activation patterns from noisy fMRI data, frequently hidden from the massive univariate analysis (e.g. PIV). A good review of exploratory techniques is provided in (McIntosh & Mišić, 2013).

Inter-subject correlation (ISC) is more recently developed data-driven method (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Kauppi, Jaaskelainen, Sams, & Tohka, 2010). The method essentially measures synchrony by correlating corresponding voxel time series across different subjects. Consequently, it reveals regions with high inter-subject similarity in terms of temporal dynamics. Underlying assumption is that the responses should be highly consistent between participants exposed to the same stimulation or performing the same task, whereas the responses unrelated to the experimental condition will most likely show high variation. Therefore, activations of high inter-subject consistency can be regarded as reliable task-related responses.

A common drawback of most of the standard data-driven methods is that they look for patterns in brain responses independently from the experiment conditions or underlying cognitive processes. As a result, the obtained activation patterns are not as explicitly linked to experimental variables as in GLM, which makes them difficult to interpret. Perhaps slightly biased but still interesting discussion on limitations of both hypothesis-driven and data-driven exploratory methods can be found in (O'Toole et al., 2007) .

During the last decade, application of machine learning classifiers for brain decoding gained popularity. Classifier is an algorithm that predicts class membership for an input data point based on the discrimination function learned from the pre-labeled training data. In fMRI settings, a classifier links voxel activity patterns to the different categories of the experimental variables, such as pictures of faces versus objects (Cox & Savoy, 2003) or spoken words and speaker (Formisano, De Martino, Bonte, & Goebel, 2008). In other words, sensory input or cognitive states of the brain can be inferred or 'decoded' from the elicited neural responses. For continuous variables, regression instead of classification can be used. For example, Toiviainen et al. (Toiviainen, Alluri, Brattico, Wallentin, & Vuust, 2014) applied principal component regression to predict six stimulus features from the brain responses. The approach that direction of inference is reversed is encoding. Above described GLM can be viewed as simplified encoding model that predicts one voxel time series from the stimulus time course. Principles, application, and limitations of the encoding and decoding models are discussed in (Haynes, 2015; Naselaris, Kay, Nishimoto, & Gallant, 2011).

Connectivity studies look at the synchrony of the regions within the brain. Connectivity studies consider brain as a complex network of interconnected nodes. Nodes can be any neural elements of interest such as cell, voxel, or anatomical region, and can be connected biologically or in terms of dynamics of interaction. Sporns (Sporns, 2010) defines structural, functional, and effective connectivity. Structural connectivity refers to the anatomical or structural connection between nodes. Functional connectivity is defined as synchronous covariation between remote regions. Effective connectivity implies stronger directed causal relationships. Correlation-based measure of connectivity has been used in resting state studies describing intrinsic functional connectivity (e.g. (Fox et al., 2005)). Correlation does not provide information on how information flows through the network, or whether the connection between nodes is direct or indirect (e.g. if two nodes are linked by the third node). Granger causality is common measure for estimating effective connectivity based on time lags between the two nodes. In the essence, the node X of a network is considered to 'drive' the node Y, if the past values of X can predict the future values of Y, better than the past values of Y itself. A good introduction to the field can be found in (Rubinov & Sporns, 2010) informative review of the methodology and associated challenges can be found in (Smith et al., 2011; Sporns, 2014).

1.5 Aims

Importance of studying neural correlates of complex real-world input through naturalistic experiments has been discussed in Section 1.3. This work attempts to apply and improve existing exploratory methods to the fMRI images elicited from naturalistic music listening experiment. Under the general goal of linking brain responses to timbral, tonal and rhythmic aspects of music information, data-driven methodology is explored from two perspectives: one part focuses on stimulus representation, by means of generating features from music stimulus that capture the mentioned perceptual dimensions of music. Another part searches for alternative ways of dimension reduction before subjecting data to independent component analysis (ICA) in order to improve extraction of stimulus-related brain activation patterns.

The dataset employed in all publications except for PIV originated from the experiment where 11 participants with formal music training listened to 8.5-minute-long recording of modern Argentinian tango by Astor Piazzolla in the fMRI scanner without any specific task other than to listen attentively. Detailed description of the demographics of the sample as well as initial preprocessing of the elicited fMRI images can be found in (Alluri et al., 2012). This dataset represents complex stimulation, where multiple interconnected perceptual dimensions of music form the concurrent continuous streams. Therefore, it is the perfect platform for exploring data-driven methods and methodological challenges associated to analyzing naturalistic experiments.

Figure 1 depicts the overall setup of the studies PI-PIV. Entire analysis workflow can be divided into two parts. The first part deals with the feature generation, i.e. problem of finding the representative set of features capturing perceptual dimensions of stimulus (left side of Figure 1). The problem is addressed in PII and PV. Both studies employed nonlinear method to generate sets of high-level features. Neural correlates of these new features were compared to the neural correlates of the stimulus features designed by Alluri and colleagues in (Alluri et al., 2012). By employing the two stimulus representations in our ICA-based analysis framework (will be described in Section 2.4.5), study PII compared stimulus-related consistent independent components. General aim was to test whether nonlinear features provide an advantage in finding music-related networks of activation. Study PV also explored neural correlates of nonlinear stimulus features, under the methodological design proposed in (Alluri et al., 2012). Application of the same methods on the same data also enabled comparison of the results. In addition, perceptual correlates of the features were examined.

The second part is related to finding stimulus-related activation patterns from the observed fMRI images (right side of Figure 1). The analysis framework in studies PI-PIV was based on ICA. In this framework, a set of statistically independent activation patterns were obtained by ICA decomposition, of which a subset of the independent components (IC) were selected that were temporally related to the stimulus features and consistent across subjects. When SNR is small,

as is the case for complex naturalistic fMRI data, ICA is likely to have reliability and stability issues. This problem is typically addressed by dimension reduction of the input data before prior source separation. However, 'correct' or intrinsic dimensionality (also referred as model order) of the input data is usually unknown. When the model order is under- or overestimated, ICA model assumptions are violated and consequently, subdivision or merging of several sources (i.e. brain activation patterns) might occur. These issues were addressed from different perspectives in PI, PIII and PV. In PI and PIII standard two-set canonical correlation analysis (CCA) and the multiset extension were employed for reducing dimensionality before ICA. In addition, PIII addressed model order estimation problem and offered a heuristic solution. Study PIV offers an innovative methodological development that eliminates the need of model order estimation. Furthermore, hypothesis-driven ROI selection is applied to improve ICA decomposition.

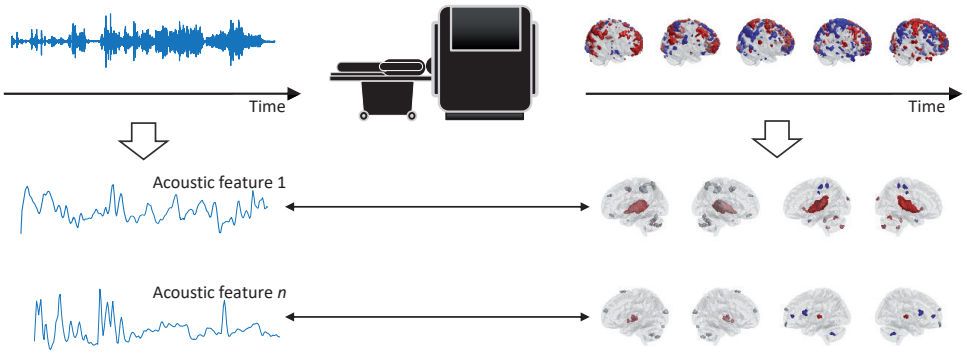

FIGURE 1 Overall setup of the study

2 METHODOLOGY BACKGROUND

This chapter overviews the methods applied in publications. After the general and concise introduction of methods, Section 2.4.5. will detail ICA analysis framework applied throughout this work.

2.1 Principal Component Analysis

Principal component analysis (PCA) is widely applied tool for exploratory data analysis. It finds transformation of the input data into a new basis where projection to each successive axis will have the next largest variance. It enables reducing number of variables in the data, yet preserving most of the variance. It is commonly assumed that signal of interest has largest variance, therefore, PCA reduction can also be seen as separating signal and noise subspaces. PCA is achieved by diagonalizing covariance matrix, while variances on diagonal elements will be maximized. In practice this method is usually implemented by eigen decomposition of the covariance matrix of the centered data. For the centered data matrix $\mathrm{X} \in \mathrm{R}^{v \times t}$ (v - voxels, t - time points) with x_i as column vectors, eigenvectors and eigenvalues of the covariance matrix $\mathrm{C} = \mathrm{E}\{\mathrm{X}\mathrm{X}^{\mathrm{T}}\}$ are estimated by solving the following eigenvalue problem:

$$\mathrm{C}\mathrm{V} = \Lambda\mathrm{V} \qquad \text{2.1-1}$$

where $\mathrm{V} \in \mathrm{R}^{t \times t}$ is an orthonormal matrix containing eigenvectors as columns, i.e. the new basis, and $\Lambda \in \mathrm{R}^{t \times t}$ is the diagonal matrix of eigenvalues, i.e. variances corresponding to each eigenvector. Principal components are obtained by projecting the input data to the new basis:

$$\mathrm{P} = \mathrm{V}^{\mathrm{T}}\mathrm{X} \qquad \text{2.1-2}$$

Dimension reduction using PCA can be achieved by selecting a subset of $d < t$ eigenvectors that cumulatively explain predefined amount of variance in the data. Projecting input data X to the reduced space $\mathrm{V} \in \mathrm{R}^{t \times d}$ in the equation 2.1-2, will

produce d principal components. Due to its relative simplicity, PCA is very popular tool for data exploration, reduction, and compression.

2.2 Kernel Principal Component Analysis

Relationship between variables is not always linear and the linear methods are blind to nonlinear patterns in the data. One way to use linear method for learning nonlinear patterns is by mapping the data to a higher dimensional space, where the pattern becomes linear. The straightforward way to use PCA for exploiting nonlinear relationships is to introduce a nonlinear function $\phi\ x$ that will map all elements of input matrix from input space χ to the new space $\mathcal{F}$, calculate covariance matrix in the new space and solve the eigenvalue problem as in linear PCA. However, nonlinear mapping to the higher dimensional space and estimation of covariance matrix is often computationally expensive or even infeasible. An elegant solution is to use 'kernel trick'. In essence, kernel trick is a way to directly estimate of dot product in the $\mathcal{F}$ space by kernel function $K(x,x) = E\{\phi\ x\ \phi\ x\ ^T\} = \frac{1}{m}\sum_{i=1}^{m} \phi\ x_i\ \phi\ x_i\ ^T$, without explicitly calculating the nonlinear mapping $x \to \phi\ x$. For the methods that depend only on dot products, such as PCA, calculating kernel function for input data directly provides an estimate of the covariance matrix (or Gram matrix produced by outer product here precisely, but they have same eigenvalues in our case). With kernel function, eigenvalue equation for linear PCA in eq. (2.1-1) can be rewritten:

$$Cv = \lambda v \Rightarrow \frac{1}{m}\sum_{i=1}^{m} \phi(x_i)\phi(x_i)^T v = \lambda v \qquad 2.2\text{-}1$$

It can be showed that all solutions v lie in the span of $\phi\ x$:

$$v = \sum_{i=1}^{m} a_i \phi\ x_i \qquad 2.2\text{-}2$$

Substituting (2.2-2) into (2.2-1) and a few operations we get:

$$Ka = \lambda a$$

where $K(x,x) = \frac{1}{m}\sum_{i=1}^{m} \phi\ x_i\ \phi\ x_i\ ^T$, λ and a denote eigenvalues and eigenvectors, respectively. Principal components can be obtained by projecting sample to K, i.e. to the eigenvectors of covariance matrix in $\mathcal{F}$:

$$\langle v, \phi\ x\ \rangle = \sum_{i=1}^{m} a_i\, K(x_i, x) \qquad 2.2\text{-}3$$

More detailed derivation can be found in (Muller, Mika, Ratsch, Tsuda, & Scholkopf, 2001). Gaussian kernel was applied in PV, calculated by:

$$k_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\varepsilon^2}\right) \qquad 2.2\text{-}4$$

where k_{ij} are elements of the kernel matrix, x_i, x_j are column vectors of X, and ε is the parameter, estimated by:

$$\varepsilon = \underset{i}{\text{median}} \min_{i \neq j} \left\| x_i - x_j \right\| \qquad 2.2\text{-}5$$

Polynomial kernel of third degree, applied in PII, is given by:

$$k_{ij} = \left(x_i x_j^T + b \right)^3$$

For simplicity, the parameter b was set to 1. Dimension reduction is achieved similarly as in linear PCA, by selecting a subset of eigenvectors based on the explained variance.

2.3 Canonical Correlation Analysis

Canonical correlation analysis (CCA) has been developed by Hotelling (Hotelling, 1936). While PCA reduces data by finding the subspace of maximum variation, CCA finds the subspace of shared covariation between two datasets. In other words, for each of the two multidimensional vectors x and y, it finds two bases with w_x and w_y such that correlations between canonical components $p_x = w_x^T x;\ p_y = w_y^T y$ are maximized. This can be achieved by solving the following eigenvalue equations (Borga, 2001):

$$\begin{cases} \mathrm{C}_{xx}^{-1}\mathrm{C}_{xy}\mathrm{C}_{yy}^{-1}\mathrm{C}_{yx} w_x = \rho^2 w_x \\ \mathrm{C}_{yy}^{-1}\mathrm{C}_{yx}\mathrm{C}_{xx}^{-1}\mathrm{C}_{xy} w_y = \rho^2 w_y \end{cases} \qquad 2.3\text{-}1$$

where $\mathrm{C}_{xx} = E\{xx^T\}$ $\mathrm{C}_{xy} = E\{xy^T\}$ are within-set and between-set covariance matrices, respectively. Eigenvalues ρ^2 are squared canonical correlation and eigenvectors are normalized basis vectors. If the input data are whitened beforehand, 2.3-1 reduces to singular value decomposition of the covariance matrix:

$$\mathrm{C}_{xy} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}} \qquad 2.3\text{-}2$$

where U and V are orthogonal matrices containing basis vectors, and Σ is a rectangular diagonal matrix that contains canonical correlations on the main diagonal. Canonical components are rank ordered by canonical correlations, i.e. by correlations between corresponding k-th set of canonical variables p_x^k, p_y^k. Dimension reduction can be achieved by setting the threshold on canonical correlations and retaining only canonical components correlated above the threshold.

Multiset CCA (MCCA) extends CCA to multiple datasets. MCCA optimizes an objective function that maximizes overall correlation between multiple canonical components. Kettenring (Kettenring, 1971) proposed five measures to calculate overall correlations. Here, sum of correlations is applied to measure overall correlation between k-th set of canonical components:

$$R^k = \sum_{m,n=1}^{M} |\rho_{m,n}^k| \qquad 2.3\text{-}3$$

where M is the number of datasets. Several approaches have been proposed to solve MCCA problem (Karhunen, Hao, & Ylipaavalniemi, 2012a; Vía, Santamaría, & Pérez, 2007). Here an implementation offered in (Li, Adali, Wang, & Calhoun, 2009) was employed, which follows the algorithm proposed in the original publication (Kettenring, 1971). In this algorithm, canonical components are estimated recursively such that each of K iterations produce set of components maximizing R^k function:

$$\{w_1^k, w_2^k, \dots, w_M^k\} = \underset{w}{argmax}\, R^k \qquad 2.3\text{-}4$$

After each step, the obtained components are removed from the data and estimation procedure is repeated. Similar to two set CCA, estimated canonical components are subjected to the following constraints:

1. canonical components Z_m are uncorrelated within each dataset and have zero mean and unit variance:

$$E\{Z_m^T Z_m\} = I, m = 1,2, \dots, M \qquad 2.3\text{-}5$$

2. canonical components between datasets have nonzero correlations only on corresponding indices, i.e. $E\{Z_n^T Z_m\}, m \neq n$ is the diagonal matrix with canonical correlations over the main diagonal.
3. canonical correlations between corresponding datasets are sorted in non-decreasing order:

$$|r_{m,n}^1 \geq |r_{m,n}^2, \dots, \geq |r_{m,n}^K| \text{ for any } m, n \in \{1,2, \dots, M\}$$

As a result of this optimization, a set of basis vectors and canonical components are obtained for each dataset (fMRI data from one subject). Dimensions are reduced by retaining canonical components with average canonical correlation above predefined threshold.

2.4 Independent Component Analysis

2.4.1 ICA Model

Initially developed for solving cocktail party problem, ICA established itself as a multi-purpose tool for wide range of problems including feature extraction and source separation of signals in various domains. The idea behind ICA is very nicely illustrated by the cocktail party problem - a situation where people are speaking simultaneously, and several microphones record the signal (Hyvärinen & Oja, 2000). Each microphone picks up voices of all speakers with different intensity, depending on their relative position. The aim is to recover each speaker's

voice from the recorded mixtures with no prior information but a few assumptions about mixing or sources. In a matrix form this process is expressed by:

$$\mathrm{X} = \mathrm{AS} \qquad \text{2.4-1}$$

where X is the matrix containing the observed signals, A contains the mixing coefficients or weights, and S is the matrix of sources. ICA estimates demixing matrix $\mathrm{W} \approx \mathrm{A}^{-}$ such that independent component estimates $\mathrm{Y} = \mathrm{WX}$ are as close to true sources as possible. In the ideal case:

$$\mathrm{W} = \mathrm{A}^{-} \text{ i.e. } \mathrm{Y} = \mathrm{WX} = \mathrm{WAS} = \mathrm{S} \qquad \text{2.4-2}$$

Not having any information about two members makes Equation 2.4-1 an ill-defined problem. However, estimation of A^{-} and S is possible with few assumptions on sources. These assumptions are:

a) Sources are non-gaussian (except possibly one) and mutually independent
b) Sources are mixed linearly; the mixing matrix is square and invertible.

With these assumptions, estimation of the mixing matrix can be achieved using central limit theorem, according to which the sum of independent random variables are distributed more closely to gaussian than the individual variables. In the Equation 2.4-1, X is a linear combination of non-gaussian sources, i.e. X will be maximally non-gaussian when it is equal to one of the sources s_i. As shown in the Equation 2.4-2, Y can also be represented as the linear combination of the sources and by the same logic, Y will be maximally non-gaussian when it is equal to one of the sources. Therefore, ICA estimation can be defined as an optimization problem, where weights in W are selected such that non-gaussianity of $\mathrm{Y} = \mathrm{WX}$ is maximized. There are more than one ways to measure gaussianity and therefore several cost functions can be used for the mentioned optimization. More extensive presentation of different ICA algorithms and cost functions is provided in (Hyvärinen & Oja, 2000).

Here we applied FastICA algorithm (Hyvarinen, 1999) with an approximation of negentropy as the cost function, due to its efficiency and speed. Negentropy as a measure of non-gaussianity stems from the link between gaussianity and entropy in the information theory. Namely, it is proven that among all variables with equal variance, gaussian variable will have the largest entropy (Hyvärinen & Oja, 2000). Negentropy is the function that considers entropy of any variable in relation to the entropy of a gaussian variable:

$$\mathrm{J}\ y\ = \mathrm{H}(y_{\mathrm{gaussian}}) - \mathrm{H}(y) \qquad \text{2.4-3}$$

where $\mathrm{H}(y)$ is an entropy of the variable y, $\mathrm{H}(y_{\mathrm{gaussian}})$ is the entropy of a gaussian variable with the same variance as y. In practice, estimation of $\mathrm{J}\ y$ from the formulation in 2.4-3 is difficult due to the need of probability density function estimate. An approximation that has been used throughout this work is given by:

$$\mathrm{J}\ y\ \cong\ E\{G\ y\ \} - E\{G\ v\ \}^{\ 2} \qquad \text{2.4-4}$$

$G(y) = \tanh(a_1 y)$ is a nonlinear function with constant term $a_1, 1 \leq a_1 \leq 2$, and v is the standardized gaussian variable.

One consequence of simultaneously solving mixing coefficients and sources in ICA is that there exists more than one solution satisfying Equation 2.4-1. For example, having an independent component scaled by some scalar will not change outcome of the equation as long as corresponding weight in the mixing matrix is divided by the same scalar. Similarly, the sign and the order of the estimated independent component are arbitrary.

2.4.2 Reliability of the Decomposition

Optimization of the above-mentioned objective function is done by starting with random initial condition and there is no guarantee that it will converge to the global optimum. In other words, multiple runs of ICA can converge to different solutions, hence producing different estimates of the demixing matrix and IC-s. Therefore, reliability of ICA decomposition needs to be assessed. Few approaches have been proposed to assess reliability, such as RAICAR (Yang, LaConte, Weng, & Hu, 2008), ARABICA (Ylipaavalniemi & Soppela, 2009) and ICASSO (Himberg, Hyvarinen, & Esposito, 2004). The latter has been selected to assess reliability of decomposition here. The main principle behind ICASSO reliability assessment is that reliable estimates of the sources will appear frequently in the multiple runs of ICA. To estimate reliability of IC-s, ICA is run many times and estimates of ICs produced from different rounds are clustered. Finally, the software offers few parameters and visualization tools for quantifying and exploring the clustering results.

In order to better demonstrate ICASSO operation, consider the preprocessed and reduced fMRI dataset $\mathrm{X} \in \mathrm{R}^{v \times d}$ to be analyzed, where d represents number of sources. After setting the parameters of FastICA algorithm, ICA is run 100 times, each round producing d source estimates. All $100d$ ICs are subsequently clustered using simple hierarchical agglomerative clustering with average linkage strategy. Distance between the estimates is calculated by $\sigma_{ij} = 1 - |\rho_{ij}|$, where $|\rho_{ij}|$ represents an absolute value of correlation. The most reliable components are expected to appear repeatedly in the majority of rounds and be highly similar to each other, which means they will be clustered together in a compact and well separated cluster. ICASSO offers a measure for quantifying compactness and separation of a given cluster. Following the author's definition (Himberg et al., 2004) cluster quality index I_q for m-th cluster is calculated by:

$$\mathrm{I}_q(C_m) = \frac{1}{|C_m|^2} \sum\nolimits_{i,j\ \in C_m} \sigma_{ij} - \frac{1}{|C_m||C_{-m}|} \sum\nolimits_{i \in C_m} \sum\nolimits_{j \in C_{-m}} \sigma_{ij} \qquad \text{2.4-5}$$

where C_m contains source estimates within m-th cluster, whereas C_{-m} denotes the estimates outside of it. $|C_m|$ and $|C_{-m}|$ denote number of elements within

and outside the m-th cluster respectively. Ranking the clusters by I_q index provides an indirect measure of reliability. Finally, the best IC estimate from the cluster, or cluster 'centrotype', is an element in the cluster c_i with the highest overall similarity with others:

$$c_i = \max_{i \in C_m, i \neq j} \{ \sum\nolimits_{j \in C_m} \rho_{ij} \} \qquad 2.4\text{-}6$$

Combined with visualization tools, cluster quality is useful measure for exploring the reliability of IC estimates. ICASSO also features an index that assists in estimating optimal number of clusters. However, it has not been employed here since number of clusters equal to number of sources has been the best solution for our data.

It should be noted that reliability of ICA result does not necessarily mean validity. Even the most reliable components can be related to artifact rather than the task of interest. Therefore, ICASSO does not provide means for finding interesting task-related components.

2.4.3 Model Order Selection (MOS)

In the simple ICA model (Equation 2.4-1) $X \in R^{t \times v}$ represents the observed fMRI images, organized as t time points in rows and image voxels v as columns. $A \in R^{t \times t}$ is the square mixing matrix, with columns representing time courses of the corresponding sources, and $S \in R^{t \times v}$ contains independent activation patterns as rows. This model assumes determined mixture, where the number of fMRI images is equal to the number of sources. In practice, however, few hundred images are acquired in fMRI experiments and with such number of sources in the model, ICA algorithm typically does not converge, and the estimated sources are not reliable. Therefore, to conform with the determined mixing assumption, dimensionality of the input data is typically reduced using PCA prior subjecting to ICA decomposition. The problem is that 'true' number of sources (also referred as intrinsic dimensionality and model order) of the data is not known.

Number of solutions for estimating intrinsic dimensionality has been suggested in the literature, including model order estimation techniques from information theory, such as Akaike information criterion (AIK), Minimum distance length (MDL) and Bayesian information criterion (BIC). The former two have been very popular in the field and reportedly produced good results (Calhoun, Adali, Pearlson, & Pekar, 2001). However, they do not always work. Indeed, Cordes & Nandy (Cordes & Nandy, 2006) demonstrated in simulations as well as in real resting state fMRI data that none of the mentioned methods estimated model order correctly. Authors argue that existence of strong autocorrelations in noise and low SNR renders the methods relying on eigenspectrum for model order estimation (including AIK and MDL) unreliable. Further evidence was provided by Cong and colleagues (Cong, Nandi, He, Cichocki, & Ristaniemi, 2012), who showed in their simulations that both AIK and MDL tend to underestimate

number of sources when the SNR is low. Finally, unreasonable model order estimations by AIK and MDL were observed in experiments on real fMRI data in PIII and also in (Cong et al., 2014).

In PI and PII, model order was estimated by the method proposed in (He, Cichocki, Xie, & Choi, 2010). This method, referred as SORTE, relies on the second order statistics of the eigenvalues to find a large gap in eigenspectrum. Formally, SORTE is defined as:

$$\text{SORTE } p = \begin{cases} \dfrac{var\ \{\nabla\lambda\}_{i=p+1}^{m-1}}{var\ \{\nabla\lambda\}_{i=p}^{m-1}}, & var|\{\nabla\lambda\}_{i=p}^{m-1}| \neq 0 \\ +\infty\ , & var\ \{\nabla\lambda\}_{i=p}^{m-1} = 0 \end{cases} \quad \text{2.4-7}$$

where $var\ \{\nabla\lambda\}_{i=p}^{m-1} = \frac{1}{m-p}\sum_{i=p}^{m-1}(\nabla\lambda_i - \frac{1}{m-p}\sum_{i=p}^{m-1}\nabla\lambda_i)$; and $\nabla\lambda_i = \lambda_i - \lambda_{i+1}, i = 1, \ldots, (m-1)$. Number of sources are determined by finding minimum of $\text{SORTE}(p)$:

$$d = \underset{p=1,\ldots,m-1}{\text{argmin}}\ \text{SORTE}(p) \quad \text{2.4-8}$$

SORTE provided reasonable estimate under single subject ICA approach only if the fMRI data were filtered by band-pass filter with 8mHz and 100mHz cutoffs. For group ICA approaches in PIII and PIV, however, none of the tested MOS methods worked.

In summary, none of the model order estimation techniques are universally suited to all settings. Real data rarely comply with all assumptions of the estimators, and no ground truth is available for the model order. Some heuristics have been proposed for robust estimation and repeatability (Onton & Makeig, 2006; Särelä & Vigário, 2003), but these estimates are higher limits of reasonable model order that can also be detected if reliability testing of ICA results are conducted. Studies PIII and PIV offer two different approaches to address model order selection problem.

2.4.4 Overview of ICA Approaches in Neuroimaging

ICA has become one of the standard data-driven method for fMRI since its first applications in 1990-s (McKeown et al., 1998). There are several workflows for decomposing multisubject fMRI data and drawing group inferences, summarized in Figure 2. At the highest level of the flowchart is the most general subdivision: spatial ICA and temporal ICA. Even though in practice difference between the two is mere orientation of input matrix, assumptions on mixing model are different. In spatial ICA, each of t scans in the observed data matrix $\mathrm{X} \in \mathrm{R}^{t \times v}$ is assumed to be a linear mixture of statistically independent activation patterns (sources), i.e. t is the number of dimensions in the data. In temporal ICA, each voxel time course in $\mathrm{X} \in \mathrm{R}^{v \times t}$ is assumed to be a mixture of v independent time courses. The latter approach is not widely used in fMRI because very large spatial dimension of the whole brain image - $v > 200\ 000$ makes it computationally expensive without substantial data reduction. EEG and MEG are more natural domain for temporal ICA, since it is applied in the sensor space and the number of

electrodes or squid sensors are three orders of magnitude smaller compared to fMRI voxels.

Spatial ICA is further divided by how data from multiple subjects are integrated (Figure 2). Each subject's data are decomposed separately in single subject ICA, and source estimates are combined for group analysis afterwards. However, such unconstrained ICA solution, combined with associated permutation and sign ambiguity (see Section 2.4.1), makes matching of IC estimates across subjects challenging. Multiple ways for making group inferences for single subject ICA has been proposed in literature (Calhoun et al., 2001; Esposito et al., 2005). Our approach for drawing group inferences will be detailed in summaries of PI and PII, where single subject ICA was employed.

Group ICA (GICA) approach is a popular alternative where instead of individual decomposition, data from all subjects are concatenated before subjecting to ICA. As depicted in Figure 2, data can be concatenated across temporal or spatial dimensions. Temporal concatenation strategy assumes common activation patterns across all subjects, but allows spatial maps to have subject specific time courses. Conversely, spatial concatenation assumes common temporal courses and allows unique spatial maps. When necessary, contributions of individual subjects into common activation maps or temporal courses can be reconstructed. Both single subject and group ICA have advantages and limitations that can be tailored for specific purposes. For instance, unconstrained decomposition in single subject ICA can reveal sources that are present only in a subset of subjects, whereas such sources will be challenging to detect for group ICA approaches, especially with temporal concatenation strategy.

More generally, single subject ICA is not affected by high inter-subject variability across temporal or spatial dimensions, at the expense of challenges associated with integration of ICs across subjects. Group ICA, on the other hand, dramatically simplifies group inferences. Moreover, possibility to reconstruct subject specific parts of a common map dramatically simplifies comparisons between subjects or different groups. Calhoun and colleagues in (Calhoun, Liu, & Adalı, 2009) reason that assumption of common spatial maps in temporal concatenation approach is better suitable to fMRI because temporal variation in signal is much larger than spatial variation. This is in line with the simulation study by Schmithorst & Holland (Schmithorst & Holland, 2004). Authors simulated a scenario where some of the sources are present only in a subset of subjects, and found that temporal concatenation approach outperformed spatial concatenation. Indeed, the former allows subject-specific temporal courses (i.e. weights of sources in the mixture), which means absence of a source in a given subject can be easily handled by having corresponding weights in the mixing matrix close to zero. The list of approaches is not exhaustive, but these are arguably the most commonly applied ones in modern neuroimaging. For more extensive review see (Calhoun et al., 2009).

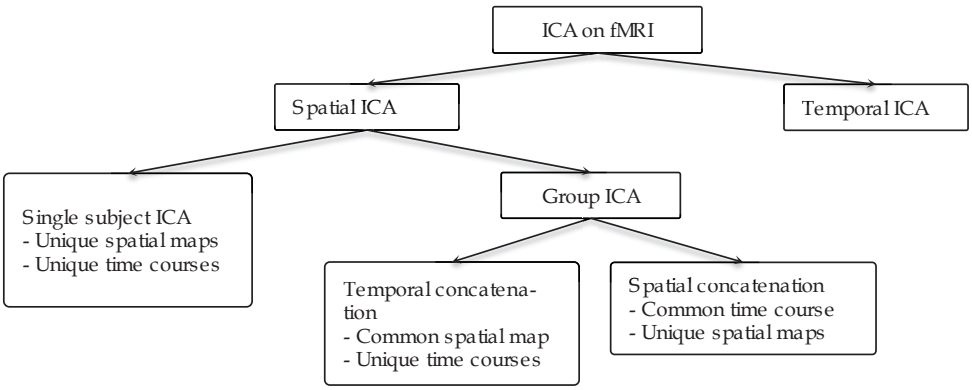

FIGURE 2 ICA approaches commonly employed in neuroimaging

2.4.5 ICA-based Methodology Applied in this Work

This section overviews single subject and group ICA analysis workflows employed in PI, PII, and PIII. Individual ICA framework is depicted in Figure 3. Each participant data was first reduced by PCA, followed by ICA decomposition within ICASSO framework to produce a set of ICs and corresponding time courses. In order to be qualified as task-related consistent activation map, the produced ICs are subjected to the two constraints:

1. Correlation between temporal course of the given IC and at least one stimulus feature should be statistically significant.
2. Spatial map should be observed in more than 5 (half) of the participants' responses.

Statistical significance of the correlations was estimated by permutation tests where phase scrambled features were correlated with randomly selected IC time course. The process was repeated 100 000 times and from the distribution of the computed correlations, significance threshold at the level $p<0.01$ was selected.

After selecting the subset of ICs satisfying the first constraint, group-level consistency analysis was conducted. To this end, similar stimulus-related activation maps from different participants were grouped together and the ones present in over half of the subjects are retained. In PI grouping of similar spatial maps was achieved simply by visual examination. This laborious process of component grouping was automated in PII by applying diffusion map to cluster ICs (Sipola et al., 2013). Finally, the components satisfying both criteria are reported as stimulus-related common spatial maps.

General workflow of conventional GICA, as proposed in (Calhoun et al., 2001), is depicted in Figure 4. The method consists of two stage dimensionality reduction: first, each subject data is reduced separately by PCA. Next, the reduced data from all subjects are concatenated and reduced again by second stage PCA. The output of the second stage reduction defines number of sources in ICA model and thus has significant effect to decomposition results. After the source

separation, a set of ICs and mixing matrix is estimated. The set of spatial maps is common to all subjects, whereas aggregate mixing matrix combines all subject-specific temporal courses for each component. The process of finding consistent stimulus-related activations is similar to single subject ICA approach: for each group IC, 11 subject-specific temporal courses are correlated to the stimulus features. The component is retained if more than half of its subject-specific time courses are significantly correlated with the given stimulus feature.

Figure 5 illustrates schematic view of the MCCA extension proposed in PIII. Here, additional dimensionality reduction is introduced between the two PCA stages of the conventional technique. While MCCA combines data from all subjects for computing correlated subspace, canonical components are produced for each dataset. The third stage reduction consists of concatenation and PCA reduction followed by ICA, as described above. More details on how this addition is implemented and how it effects the reconstruction of images and time courses is provided in PIII. Once the set of common ICs and aggregate time courses are estimated, rest of the analysis is identical to standard GICA.

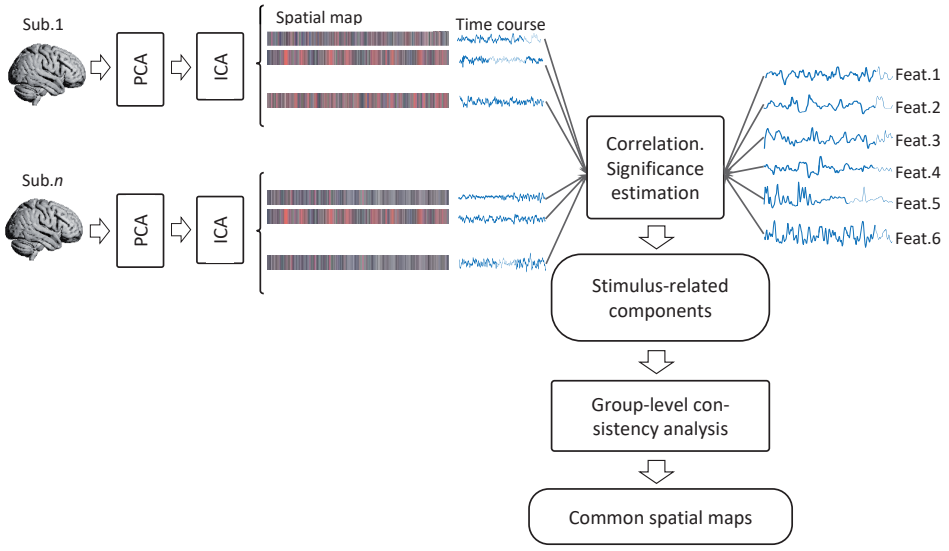

FIGURE 3 Single subject ICA framework

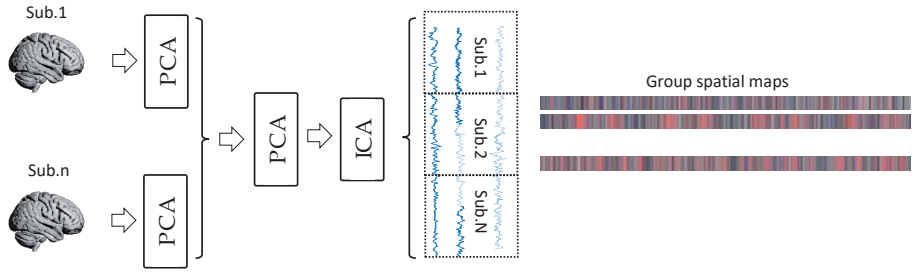

FIGURE 4 Group ICA with temporal concatenation

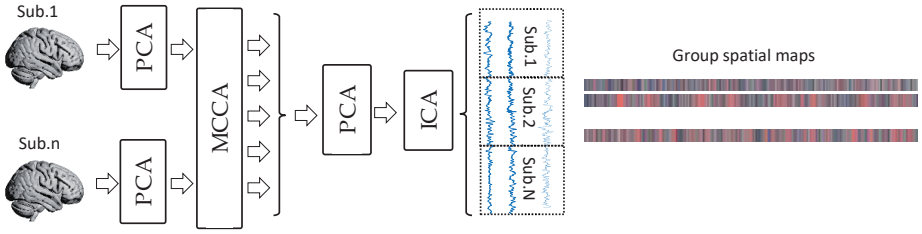

FIGURE 5 Group ICA with MCCA extension

3 STUDY SUMMARIES

3.1 Contribution of the Author

The author had an active role in all stages of research presented in this thesis. He was the main contributor in data analysis, interpreting results, and writing the manuscripts in PI, PII, PIII, and PV. Together with Fengyu Cong the author was an active contributor to the methodological design of studies PI and PII, and PIII. In PIV, the author contributed to the methodological design of the study, performed dimension reduction, conducted preliminary ICA analyses, and revised the manuscript. Besides the author, Iballa Burunat and Petri Toiviainen also contributed to the methodological design of PV.

3.2 Study PI

Aims: In this study CCA was applied for reducing data before subjecting to ICA decomposition. In contrast to more commonly applied PCA, which selects the subspace of maximal variance, CCA selects common (correlated) subspace between two datasets. It is widely accepted in neuroimaging community that task-related signal in fMRI is much weaker than noise originating from physiological processes or other sources. Therefore, retaining most of the variation might not necessarily mean retaining the signal of interest, especially for low SNR. On the other hand, common subspace between brain responses of participants exposed to the same stimulation is more likely to be linked to the stimulus itself. Furthermore, benefits of CCA has been shown in simulations and real fMRI data (Karhunen, Hao, & Ylipaavalniemi, 2012b). We tested CCA and PCA reduction before ICA and compared the results.

Methods: Methodological design is shown in Figure 6: fMRI data were reduced by both PCA and CCA, following which ICA was applied. One additional step before dimension reduction was bandpass filtering of fMRI voxel time series. The

filter cutoffs were set at 0.008Hz-0.05Hz, corresponding to the frequency band where most of the energy of stimulus features were located. Benefits of the filtering step is discussed in (Cong et al., 2011). Three model orders were tested for each method: 20 and 30 were predefined by us and 46 was estimated by SORTE (He et al., 2010). The predefined model orders were added to qualitatively examine optimality of SORTE estimate. All six datasets were subsequently subjected to individual ICA framework (see Section 2.4.5.) to find stimulus-related consistent spatial maps.

Performance of each method was assessed by examining reliability of ICA results and produced spatial maps. Primary interest was to find out whether CCA helps finding more stimulus-related activations (i.e. the sources whose temporal course significantly correlated with at least one acoustic feature).

Results: Reliability of ICA, measured by cluster quality index (see Section 2.4.2), was not significantly different between the two methods at any model order. Difference was also subtle in terms of the stimulus-related common activation maps produced by the two methods under model order of 46. Common map was observed in seven and nine subjects' responses for CCA and PCA respectively. Interestingly, the produced stimulus-driven activation maps by PCA+ICA declined for model orders below the SORTE estimate. The trend, however, was not prominent for CCA+ICA model. A possible explanation is that over-reduction forces PCA to retain only few components accounting for the largest portions variance. These components are likely to be related to the artifacts rather than stimulus due to the fact that stimulus-related signal energy is lower. Consequently, signals of interest can be discarded even before source separation. CCA, on the other hand, finds the subspace of shared variance between a pair of datasets, which is more likely to be related to the experimental manipulation. Thus, even at suboptimal model orders, canonical components are likely to contain stimulus-related sources, while noisy components will be discarded.

To summarize, while CCA reduction did not provide a large gain in ICA performance according to our criteria, observed results suggest that CCA-based reduction might lead to more stable results for suboptimal model orders. Furthermore, there were several limitations that could blur the contrast between the outcomes. First, comparison criteria were overly general for observing large differences under very narrow range of tested model orders, given the dimensionality of the input data (231 dimensions). Such restricted range was mainly consequence of the bandpass filtering step, which smoothed voxel time series and significantly reduced the rank of the data matrices. In the following study, we widened the pass band of the filter to 0.008 – 0.1Hz in order to avoid excessive smoothing and the associated collinearity problems.

Another issue that can affect the outcome is that CCA operates on two datasets. Therefore, dimensions of 11 datasets was reduced in randomly selected pairs in this study. Common subspace can vary across different pairs, which certainly affects the subsequent source separation step. In PIII, multiset extension of CCA was employed in order to extract common subspace from all datasets.

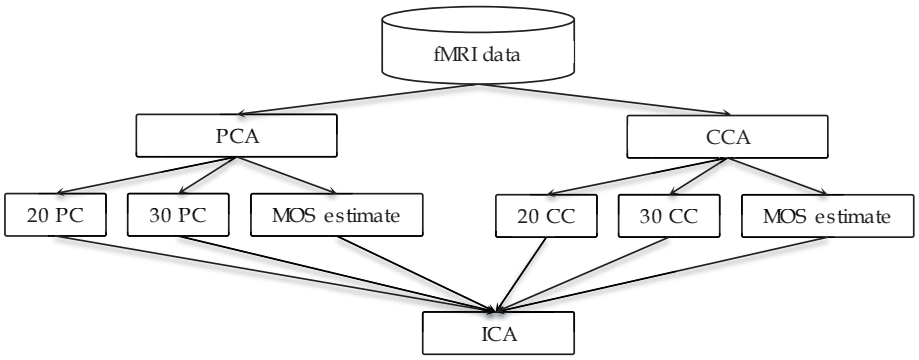

FIGURE 6 Flowchart of the study design

3.3 Study PII

Aims: While study PI explored the role of the dimension reduction in improvement of ICA decomposition, PII focused on stimulus representation. As mentioned above, (Alluri et al., 2012) extracted a set of 25 acoustic descriptors commonly used in music information retrieval field and transformed those into nine principal components, retaining over 95% variance in the data. By examining contributions of the initial descriptors to each principal component, Alluri and colleagues labeled them by the musical dimension they represent. The applied labels were subsequently validated in the perceptual experiment. Finally, six high-level features were retained. Four features including Fulness, Brightness, Timbral Complexity, and Activity, represent polyphonic timbre of music; Key Clarity represents clarity of the key, and Pulse Clarity represents clarity of the main beat in music.

Aim of this study was to explore an alternative set of stimulus features, generated by applying KPCA to the same acoustic descriptors that were employed in (Alluri et al., 2012). As a nonlinear method, KPCA can exploit nonlinear patterns in the data and generate novel stimulus features, which can subsequently help finding interesting brain activations.

Methods: Figure 7 depicts the process of kernel PCA feature generation. The initially extracted raw descriptors went through the same preprocessing routine as in (Alluri et al., 2012). Namely, the sampling rates of the descriptors were matched and convolved with HRF in order to account for the lag produced by hemodynamic response. The features were band-pass filtered with cut offs at 0.008 and 0.1Hz. After the preprocessing, KPCA was applied with third degree polynomial kernel, producing 14 nonlinear features. The generated features and

their neural correlates were compared with the six PCA features designed in (Alluri et al., 2012). Single subject ICA approach (see Section 2.4.5.) was employed for finding consistent activation maps related to each set of features.

Results: Results revealed 53 ICs from all 11 subjects were significantly correlated with KPCA features. In the Table 1 (top row) number of the components linked to each stimulus feature is shown. As can be seen from the table, more than one components from each subject can be associated with each stimulus feature. For example, K_2 is correlated to time courses of 11 components from seven subjects. However, due to large variation in activation patterns across participants, majority of the spatial maps did not satisfy our group-level consistency criterion (being present in responses of at least six subjects) and were discarded from the subsequent analyses. Figure 8 demonstrates spatial variability of the stimulus-related responses across subjects by showing averaged (left half) versus combined thresholded maps (right half), associated to each of the features. Averaged spatial maps associated to KPCA tend to consist of smaller clusters with relatively strong presence in frontal areas, whereas in averaged PCA maps activations are more concentrated in the auditory areas. Visual examination of averaged and combined spatial maps shows that PCA maps tend to be more consistent across subjects than KPCA counterparts. The reason behind this phenomenon is difficult to pinpoint.

Overall, brain correlates of the two stimulus representations feature similar and unique areas. Namely, presence of large clusters of activation in the frontal areas of averaged K_2 and K_3 maps is absent in PCA maps. On the other hand, we see significantly larger auditory areas in all PCA maps. In addition, activations in the parietal and occipital areas in Timbral Complexity map only partly overlaps with averaged K_2 and K_3 responses. Nevertheless, group-level consistency criterion was not fulfilled by majority of the features. Common activation patterns were found only for K_3 from KPCA features (see Figure 9, left half), and for Brightness and Activity from PCA (Figure 9, right half). Both spatial maps show activations predominantly in bilateral auditory cortices. Additional activation clusters in left inferior frontal gyrus and left precentral gyrus are visible in responses to K_3 feature.

KPCA features showed low to moderate correlation to PCA features (see Figure 10). Feature K_3 is moderately correlated to Brightness (r=0.52), and Timbral complexity feature (r=0.39), while not correlated to Activity. Temporal course of K_3tends to capture significant changes in temporal courses of Brightness and Timbral complexity, but otherwise exhibits less dynamics than the two.

To conclude, nonlinear mapping between initial descriptors produced new interesting stimulus features. However, to interpret neural correlates of K_3, finding its perceptual correlates is needed. Such analysis was out of scope of this study. As in the case of linear PCA features, common activation pattern was not observed for majority of the features. This might be attributed to excessively strict constraints for group-level consistency for source estimates. Study PV offers

more extensive analysis of both perceptual and neural correlates of nonlinear features. However, the results are not comparable since the generated features as well as the employed methodology was different.

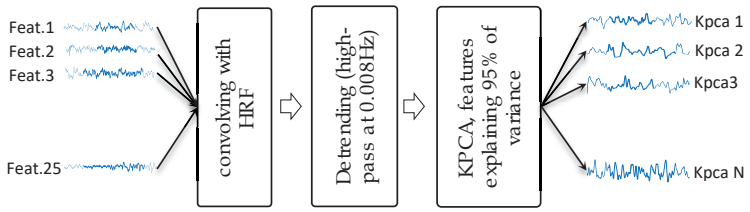

FIGURE 7 KPCA feature generation process

TABLE 1 Source estimates related to the selected stimulus features

	K_2	K_3	K_{12}	Ful.	Bri.	T.C.	Act
stimulus-related components (components/subjects)	11/7	28/10	14/8	9/7	17/11	6/6	15/8
Common map (subjects)	-	9	-	-	10	-	7

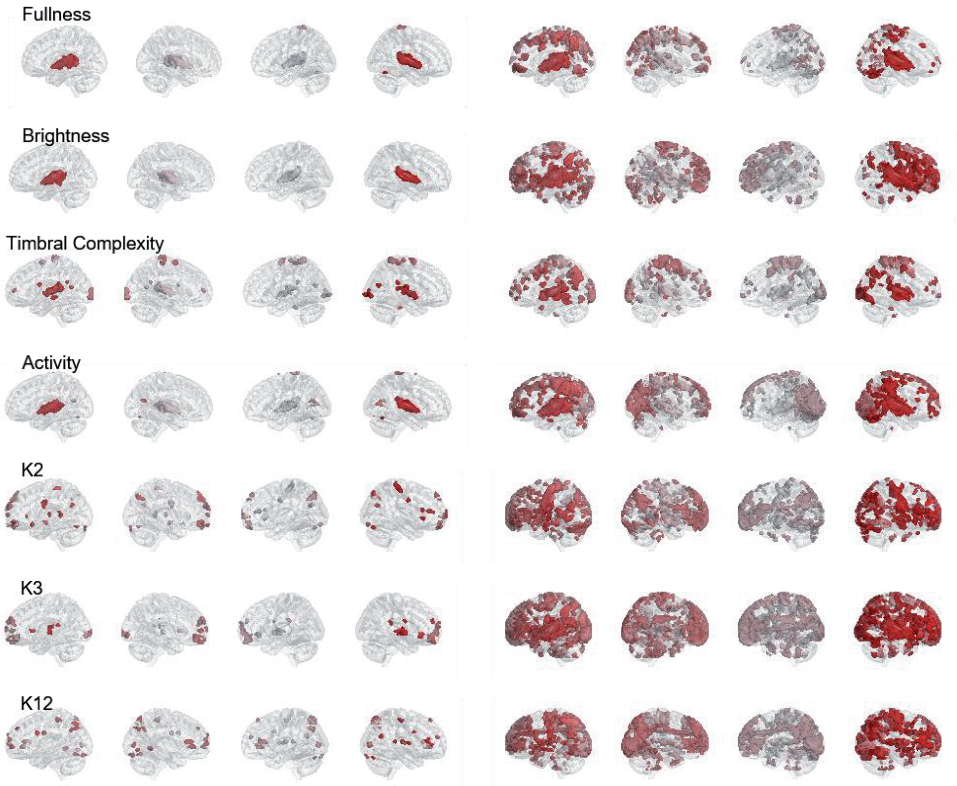

FIGURE 8 Averaged and combined functional networks

FIGURE 9 Common spatial maps for K_3 and Brightness

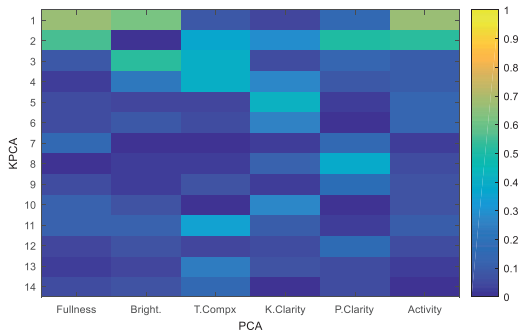

FIGURE 10 Correlation between KPCA and PCA features

3.4 Study PIII

Aims: This study presents an extension of group ICA by adding multiset CCA (MCCA) to the conventional two-stage PCA reduction. MCCA addressed the limitations of standard two set CCA discussed in Section 3.2. In addition, single subject ICA framework in PI was replaced by group ICA in order to simplify group-level analysis.

Methods: General design of study PIII was similar to PI, conventional GICA and our extension were tested on fMRI data from music listening experiment and the results were compared. However, criteria for comparison according were extended. In addition to ICA decomposition stability, here quality of the produced independent components was assessed. Decomposition stability was measured by three variables monitoring convergence of ICA and reliability of IC-s. Quality of the separated sources was characterized by another three variables: number of stimulus-related components produced, number of consistent maps across subjects, and the size of the stimulus-driven activation maps.

One additional problem we faced was that MOS methods failed to provide reasonable estimates of dimensionality for the second stage dimension reduction. For example, for the concatenated data consisting of 80*11=880 dimensions, SORTE estimated model order of 878. Other MOS methods were also tested but without success (Cong et al., 2014). Such high estimate is unreasonable if we consider the rate of decline of ICA convergence as a function of model order, pro-

vided in Figure 11 (left). Clearly, regardless of the output of the first stage reduction (denoted as 'indPCA' in the figure), ICA virtually does not convergence for model orders above 100. This is in line with the heuristic for estimating upper limit of model order, proposed in (Onton & Makeig, 2006), which states that for number of voxels (samples) v needed for estimating n parameters (sources) is $v = k \times n^2$, where k is an arbitrary constant, suggested to be larger than 25. With 228453 voxels in our data, $k = 25$ corresponds to model order of 95, at which ICA convergence rate is low (see Figure 11, left).

In order to estimate optimal model order, experiments on conventional GICA was conducted by consecutively fixing output of the first stage reduction, while varying the model order (i.e. the output of the stage two). Outputs from all combinations were evaluated by the six variables mentioned above. Consequently, dimensionality for stage 1 and stage 2 PCA reductions in standard GICA were fixed to 80 and 40 respectively.

In extended GICA there are three stages of reduction. Output dimensions of the two PCA stages were matched to those in standard GICA, to make results comparable. Next, different output of MCCA stage were tested and the one providing the best result was selected. Finally, after determining outputs of dimension reduction, results of the two methods were compared.

Results: One outcome of this study was the proposed heuristics for evaluating ICA results. As mentioned above, ICASSO indices are useful to quantify reliability and stability of ICA decomposition. Here, those indices were combined with the three additional variables that provide quality measure for the estimated sources. Although initially developed for comparing results of two methods, the proposed variables can also serve as useful heuristic for estimating number of sources when MOS methods fail.

Our extension of GICA outperformed conventional GICA according to our criteria. While the improvement is in the results was not significant, it was systematic throughout extensive testing under multiple parametrizations. For example, it can be seen from Figure 11 (right) that convergence of GICA with MCCA extension (blue line) was better than conventional GICA at all model orders up to 90.

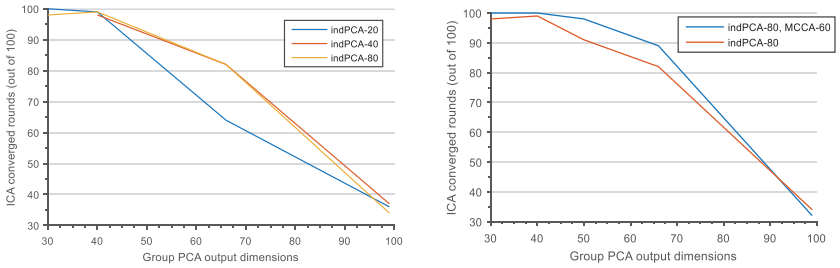

FIGURE 11 ICA convergence as a function of model order

3.5 Study PIV

Aims: Study PIV examined functional networks involved in the processing of the perceived beat in music. The dataset analyzed originated from the experiment, where brain responses to 18 musicians and 18 non-musician controls were acquired while continuously listening to three pieces of music. Pulse Clarity feature, modelling perceived clarity of the main beat, was extracted from stimulus audio. Brain correlates of the feature were computed by ICA. In addition, responses of musicians and control group were compared in order to examine effect of music training on the brain circuitry involved in rhythm processing. What is more relevant to this thesis, PIV also offered an interesting methodological development that addresses challenges associated with MOS.

Methods: The methodological approach was semi data-driven: first, hypothesized regions of interest (ROI) involved in processing of musical rhythm were selected based on the existing knowledge. Next, data-driven GICA with temporal concatenation was applied to the selected ROI (comprising roughly 10% of voxels in the entire brain) to identify stimulus-related activation maps. As in PIII, GICA was tested under several model orders, but instead of selecting the 'correct' model order, here ICs were selected from different model orders. Specifically, one IC per model order was selected that best correlated with the stimulus feature.

Results: Results revealed auditory and motor areas activated in response of Pulse Clarity feature. Interestingly, non-musicians' neural responses were better predicted by Pulse Clarity feature than musicians' responses. It can be explained by the fact that computational model of pulse clarity is built solely on acoustic features of music. While non-musicians may also rely on the same acoustic features of music to estimate main beat, better understanding of metric structure allows musicians to have more advanced models, which use additional high-level features to identify and follow the beat.

To the methodological end, we observed that estimated ICs under different model orders were different: at low model orders, stimulus-related components contained larger networks of activation, while at higher model orders the components maps were split into smaller sub-networks. This finding is in line with results of previous studies (Abou - Elseoud et al., 2010; Särelä & Vigario, 2003). Such hierarchical organization of networks would not be possible to observe if one 'correct' model order was selected for ICA.

As mentioned above, GLM was also applied to the data. Results were in concordance with the ones produced by GICA. However, only part of the significant activations was uncovered and only in non-musicians' responses, adding to the existing evidence that data-driven methods frequently outperform traditional massive univariate approaches in finding neural correlates of complex naturalistic stimulus.

3.6 Study PV

Aims: In this study, perceptual and neural correlates of KPCA features were explored and the results were compared to (Alluri et al., 2012). For simplicity (Alluri et al., 2012) will be referred as original study. In spite of similarities in main research goals, PV is different from PII in three aspects. First, gaussian kernel was selected for kernel PCA, while PII featured polynomial kernel. Therefore, the generated stimulus features are different. Second, methodological design is different: instead of single subject ICA, here the method proposed in the original study was employed to find neural correlates of the generated features. Third, PV investigated perceptual relevance of the nonlinear features.

Methods: Apart from the choice of the kernel, feature generation process was similar to the one presented in PII (see Figure 7). First part of the work focused on finding perceptual correlates of the KPCA features.

Perceptual validation procedure in the original study is illustrated schematically in Figure 12. Average feature values over 6 second frames with 1 second step size were first calculated, producing several hundred frames per feature. The obtained frames were rank ordered according to average feature values, from which 30 frames were equidistantly sampled under the constraint that the selected frames captured entire dynamic range of the feature. Audio excerpts corresponding the sampled 30 frames were presented to 21 musicians, who rated perceived levels of the musical feature on nine-point scale.

Data from the described experiment were employed here. Average KPCA values from all $\mathrm{W}_{si}, i = 1,2,\ldots,30$ frames were estimated and were subsequently correlated to corresponding R_i ratings (see Figure 12). Significance of correlations was estimated by Monte-Carlo sampling. A given feature K_i was associated to the rating R_i if: 1. Correlation between the two was significant, 2. The selected 30 frames W_s captured at least 80% of the entire range of K_i.

Neural correlates of the two sets of features were first compared at global level by finding an overlap between combined activation maps. Next, similarity between the individual activation maps were assessed by correlations.

Results: Half of the KPCA features showed significant correlations with the perceptual ratings. Among the ratings Rhythmic Complexity and Event Synchronicity, which PCA features failed to predict. Interestingly, some parts of frontal and visual areas were present in the activation maps of the new features, which were not observed in the original study. More detailed analysis of behavioral and neural correlates of the KPCA features are presented in the article.

To summarize, main contribution of PV was to demonstrate that KPCA has a potential to capture high-level musical percepts. For increased reliability of the results, the limitations of the study stemming from methodological issues as well as from insufficient perceptual rating data will need to be addressed in the future

investigations. These limitations are discussed in the article and solutions to are proposed.

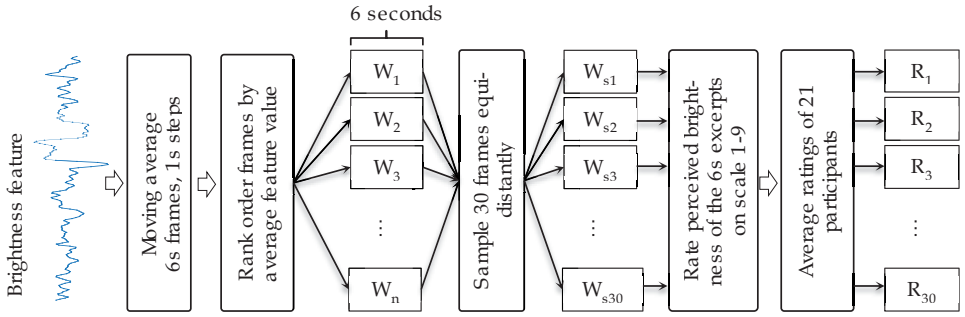

FIGURE 12 Perceptual validation procedure in (Alluri et al., 2012)

4 DISCUSSION

This exploratory work contributed to the development of data-driven methodology for naturalistic fMRI experiment. Complex continuous stimulation in natural music listening settings makes finding functional networks involved in processing of various aspects of music challenging. This thesis approaches the challenge from two perspectives. First is to improve source separation quality by reducing complexity and noisiness of fMRI data prior subjecting ICA decomposition. Second is to explore whether generation of a new features as the way of forming alternative representation of stimulus can unveil additional brain areas involved in music processing. Following is the summary of the main contributions and limitations of this work:

This work demonstrated that model order estimation for ICA on theoretical grounds is not always possible for noisy fMRI data. It has been shown that model order has significant influence on the decomposition results. Specifically, underestimation of model order can cause merging of multiple functional network into one IC or omit some of the weaker activations, whereas overestimation produces unpredictable subdivisions of networks, and the model suffers from overfitting (cf. PIV, also (Abou - Elseoud et al., 2010; Beckmann & Smith, 2004)). Based on existing evidence, Burunat (Burunat, 2017) argues that in neuroimaging notion of correct model order is inadequate. Instead, the choice of the model order can be guided by the specific aims and research questions of the study. I share this view and add that in exploratory work when required level of 'magnification' for functional networks is not known in advance, examining source estimates under several model orders is the most appropriate approach. Certainly, analyzing all components can become prohibitively expensive, especially when data dimensionality is large. One way to address this is proposed in PIV, where one (or more) IC from each model order were selected based on criteria that suits the aim of the study.

An extension of conventional GICA was proposed that integrates MCCA in the two-stage dimension reduction process. The method produced promising results

in our tests. However, comparison of results from our extension and the conventional approach proved challenging due to the scarcity of the ICs satisfying group-level consistency criteria.

Multiple studies, including PI - PIII, and also (Cong et al., 2014; Puoliväli et al., 2013), where ICA-based method was applied to naturalistic music listening data, common activation maps have been usually observed for only one or two of the six stimulus features, even though the number of stimulus-related source estimates varied considerably. There are several possible contributing factors to this issue. Sub-optimal model order selection can be one reason. However, MOS can only be a part of the issue since in most cases at least one IC has been found for each stimulus feature, but those usually did not satisfy group-level consistency criterion (to be present in responses of over 5 subjects) and discarded.

Another likely contributor might be the group-level consistency constraint itself (See section 2.4.5.) that rejected some of the stimulus-related components. This constraint assumes virtually identical functional networks processing higher-level concepts in music information. Combined with inevitable stationarity assumption for source estimates in ICA context, the assumption of homogeneity across responses of different brains are not likely to hold for all features, especially for Key Clarity and Pulse Clarity. The stimulus piece was fairly dynamic and complex musically, and depending on the background (such as genre preferences, main instrument) musicians can have different models and expectations for musical structure, tonality, rhythm, and other high-level concepts. This can affect the cognitive processes involved in analysis of different sections of music. Therefore, there can be significant variability in functional networks and more importantly, significant changes in those across time. In support of this argument, (Alluri et al., 2012) also found that neural correlates of Key Clarity featured substantial inter-subject variability. Furthermore, responses to Pulse Clarity and Key Clarity did not replicate well in the follow up study (Burunat et al., 2016). Perhaps relaxing group-level consistency constraint in our ICA framework, which requires similar activation patterns to be present across majority of subjects, could prevent rejection of potentially interesting task-related components.

Two studies included in this work demonstrated that nonlinear methods can also generate perceptually relevant interesting musical features. In my opinion, generation of high-level features based on statistically motivated criteria (e.g. preserving most of the variance as was the case for linear PCA features) is valid but not optimal solution for finding relevant neural representations. For instance, in (Alluri et al., 2012) part of the activations with high inter-subject consistency were not predicted by any of the stimulus features. Authors attributed this to the incomplete representation of stimulus and concluded that more acoustic components are needed to explain all stimulus-driven activations. Defining and producing an exhaustive list of stimulus features and their perceptual validation would be very challenging if not impossible. A better solution would be supervised fea-

ture generation, e.g. by involving fMRI responses in production of stimulus features and then find perceptual correlates. My approach to this issue is in its early development phase. Conceptually the idea is to extract wide range of low-level descriptors from the stimulus audio (as has been done throughout this thesis), but instead of finding linear combination that retains most of the variance, find the combinations that maximize correlation to brain responses. One example would be applying CCA to a set of acoustic descriptors and spatially reduced representation of brain responses. CCA will find linear combination of acoustic descriptors and linear combination of fMRI features such that correlations between them are maximized. By selecting a subset of the most significant contributors, stimulus features and relevant spatial maps can be obtained. There are, however, multiple methodological considerations still to be resolved. For example, due to high dimensionality and sparsity of fMRI image space, voxel time series cannot be directly used in the model. Thus, spatial reduction is necessary, e.g. by using time courses of ICs instead of voxels, or by applying spatial PCA to ROIs. Nevertheless, this can be an interesting topic for the future research.

SUMMARY

Under the general goal of finding functional networks involved in music processing, this work contributed to improvement of data-driven methodology for naturalistic fMRI studies. By analyzing fMRI data collected from continuous music listening experiment, certain challenges in methodology were demonstrated and addressed. From the broader point of view, source separation using ICA is well suited tool for exploratory analysis of fMRI. However, discarding redundancies and increasing SNR is crucial to be in accordance with the model and assumptions of this source separation method. Moreover, absence of any prior information about true number of sources renders the results dependable to the heuristics for selecting model order.

Part of this work tested an application of CCA for data reduction and showed comparable results to more widely applied reduction method. In line with other scholars, I advocate the benefits of testing multiple model orders for finding stimulus-driven functional networks in complex naturalistic experiments, even in the cases where model order selection method seemingly works.

Another part explored stimulus feature generation process. In search of the novel features capturing high-level musical percepts, nonlinear kernel PCA was tested as opposed to liner PCA in the previous studies. Results turned out promising, even though finding perceptual correlates to nonlinear features proved more challenging than for linear counterparts.

Presented findings are relevant not only to fMRI elicited from music listening experiments. Rather, these can be generalized also for naturalistic studies in general, where continuous multifaceted stimulation requires exploratory approaches such as ICA to be involved.

YHTEENVETO (FINNISH SUMMARY)

Datalähtöisiä menetelmiä jatkuvan musiikin kuuntelun aikana kerättyihin fMRI-signaaleihin

Kiinnostus todellisuutta vastaaviin toiminnallisiin magneettikuvantamiskokeisiin on kasvanut tasaisesti vuosituhannen vaihteen jälkeen. Tämä trendi on heijastunut kasvavaan määrään todellisuutta jäljitteleviä kokeita, joissa koehenkilöt prosessoivat monimutkaisia ärsykkeitä tai suorittavat vaikeita kognitiivisia tehtäviä kuten elokuvien katsomista, pelien pelaamista tai musiikin kuuntelua. Tällaiset kokeet kuitenkin johtavat lukuisiin metodisiin haasteisiin koeasetelmaan vaikuttavien muuttujien hallinnassa.

Tässä eksploratiivisessa työssä osoitettiin joitakin näistä haasteista soveltamalla usein käytettyjä datalähtöisiä menetelmiä jatkuvan musiikin kuuntelun aikana kerättyihin fMRI-signaaleihin. Työn yleisenä tavoitteena oli löytää musiikin prosessointiin osallistuvia toiminnallisia aivoverkkoja, jonka osalta edistettiin nykyistä metodologiaa kahdesta eri näkökulmasta. Ensimmäisenä ärsykkeinä käytetyistä audiosignaaleista pyrittiin eristämään piirteitä, jotka kuvaavat musiikin erilaisia ominaisuuksia, kuten sointiväriä tai tonaalisuutta. Toisena tavoitteena oli edistää aivosignaaleiden purkamista luotettaviin ja laadukkaisiin riippumattomiin alueellisiin aktivaatiokarttoihin. Tämä erittely mahdollisti musiikkiärsykkeestä ja muista fysiologisista, kognitiivisista ja teknisistä prosesseista lähtöisin olevien aktivaatioiden aiempaa paremman jaottelun.

Yksityiskohtaisemmalla tasolla tutkimuksessa kartoitettiin epälineaarisen menetelmän avulla musiikkiärsykkeen korkeatasoisia piirteitä. Toisessa osassa keskityttiin fMRI-signaalien dimension pienennyksen ja mallikompleksisuuden arviointiin ja saavutettiin metodisia parannuksia.

BIBLIOGRAPHY

Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., & Kiviniemi, V. (2010). The effect of model order selection in group PICA. *Human Brain Mapping, 31*(8), 1207-1216.

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience, 37*(9), 1458-1469.

Alluri, V., Toiviainen, P., Jaaskelainen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage, 59*(4), 3677-3689. doi:10.1016/j.neuroimage.2011.11.019

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine, 30*(2), 161-173.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging, 23*(2), 137-152.

Borga, M. (2001). Canonical correlation: A tutorial. *On Line Tutorial Http://People.Imt.Liu.Se/Magnus/Cca, 4*, 5.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 16*(13), 4207-4221.

Burunat Pe□rez, I. (2017). Brain integrative function driven by musical training during real-world music listening. *Jyväskylä Studies in Humanities 302.,*

Burunat, I., Toiviainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., & Brattico, E. (2016). The reliability of continuous brain responses during naturalistic listening to music. *NeuroImage, 124*, 224-231.

Calhoun, V., Adali, T., McGinty, V., Pekar, J., Watson, T., & Pearlson, G. (2001). fMRI activation in a visual-perception task: Network of areas detected using the general linear model and independent components analysis. *NeuroImage, 14*(5), 1080-1088.

Calhoun, V., Adali, T., Pearlson, G., & Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping, 14*(3), 140-151.

Calhoun, V. D., Liu, J., & Adalı, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage, 45*(1), S163-S172.

Cong, F., Nandi, A. K., He, Z., Cichocki, A., & Ristaniemi, T. (2012). Fast and effective model order selection method to determine the number of sources in a linear transformation model. *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European,* 1870-1874.

Cong, F., Puoliväli, T., Alluri, V., Sipola, T., Burunat, I., Toiviainen, P., . . . Ristaniemi, T. (2014). Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis. *Journal of Neuroscience Methods, 223*, 74-84.

Cong, F., Leppanen, P. H., Astikainen, P., Hamalainen, J., Hietanen, J. K., & Ristaniemi, T. (2011). Dimension reduction: Additional benefit of an optimal filter for independent component analysis to extract event-related potentials. *Journal of Neuroscience Methods, 201*(1), 269-280. doi:10.1016/j.jneumeth.2011.07.015

Cordes, D., & Nandy, R. R. (2006). Estimation of the intrinsic dimensionality of fMRI data. *NeuroImage, 29*(1), 145-154.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI)"brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage, 19*(2), 261-270.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*(5), 329-340.

D'esposito, M., Zarahn, E., & Aguirre, G. K. (1999). Event-related functional MRI: Implications for cognitive psychology. *Psychological Bulletin, 125*(1), 155.

Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., . . . Di Salle, F. (2005). Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage, 25*(1), 193-205.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? brain-based decoding of human voice and speech. *Science (New York, N.Y.), 322*(5903), 970-973. doi:10.1126/science.1164318; 10.1126/science.1164318

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*(27), 9673-9678. doi:0504136102 [pii]

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping, 1*(2), 153-171.

Hasson, U., & Honey, C. J. (2012). Future trends in neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage, 62*(2), 1272-1278.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science (New York, N.Y.), 303*(5664), 1634-1640. doi:10.1126/science.1089506 [doi]

Haynes, J. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron, 87*(2), 257-270.

He, Z., Cichocki, A., Xie, S., & Choi, K. (2010). Detecting the number of clusters in n-way probabilistic clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(11), 2006-2021.

Himberg, J., Hyvarinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage, 22*(3), 1214-1222. doi:10.1016/j.neuroimage.2004.03.027

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika,* , 321-377.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* Sinauer Associates Sunderland.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626-634.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4), 411-430.

Kandel, E. R., & Squire, L. R. (2000). Neuroscience: Breaking down scientific barriers to the study of brain and mind. *Science (New York, N.Y.), 290*(5494), 1113-1120.

Karhunen, J., Hao, T., & Ylipaavalniemi, J. (2012a). A generalized canonical correlation analysis based method for blind source separation from related

data sets. *Neural Networks (IJCNN), the 2012 International Joint Conference On,* 1-9.

Karhunen, J., Hao, T., & Ylipaavalniemi, J. (2012b). A canonical correlation analysis based method for improving BSS of two related data sets. *International Conference on Latent Variable Analysis and Signal Separation,* 91-98.

Kauppi, J. P., Jaaskelainen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics, 4*, 5. doi:10.3389/fninf.2010.00005 [doi]

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 58*(3), 433-451.

Li, Y., Adali, T., Wang, W., & Calhoun, V. D. (2009). Joint blind source separation by multiset canonical correlation analysis. *Signal Processing, IEEE Transactions On, 57*(10), 3918-3929.

McIntosh, A. R., & Mišić, B. (2013). Multivariate statistical analyses for neuroimaging data. *Annual Review of Psychology, 64*, 499-525.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., & Sejnowski, T. J. (1998). *Analysis of fMRI data by blind separation into independent spatial components*. *Analysis of fMRI Data by Blind Separation into Independent Spatial Components, 6*(3), 160.

McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping, 6*(5-6), 368-372.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science (New York, N.Y.), 309*(5736), 951-954. doi:309/5736/951 [pii]

Muller, K., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. *Neural Networks, IEEE Transactions On, 12*(2), 181-201.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage, 56*(2), 400-410.

Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., & Malach, R. (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD

fMRI is related to interneuronal correlations. *Current Biology, 17*(15), 1275-1285.

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Progress in Brain Research, 159*, 99-120.

O'Toole, A. J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience, 19*(11), 1735-1752.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis* Cambridge University Press.

Puoliväli, T., Cong, F., Alluri, V., Lin, Q., Toiviainen, P., Nandi, A. K., . . . Ristaniemi, T. (2013). Semi-blind independent component analysis of functional MRI elicited by continuous listening to music. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference On,* 1310-1314.

Raichle, M. E. (2009). A brief history of human brain mapping. *Trends in Neurosciences, 32*(2), 118-126.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage, 37*(4), 1083-1090.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*(3), 1059-1069.

Särelä, J., & Vigário, R. (2003). Overlearning in marginal distribution-based ICA: Analysis and solutions. *Journal of Machine Learning Research, 4*(Dec), 1447-1469.

Schmithorst, V. J., & Holland, S. K. (2004). Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *Journal of Magnetic Resonance Imaging, 19*(3), 365-368.

Sipola, T., Cong, F., Ristaniemi, T., Alluri, V., Toiviainen, P., Brattico, E., & Nandi, A. K. (2013). Diffusion map for clustering fMRI spatial maps extracted by independent component analysis. *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop On,* 1-6.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., . . . Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage, 54*(2), 875-891.

Sporns, O. (2010). *Networks of the brain* MIT press.

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience, 17*(5), 652-660.

Toiviainen, P., Alluri, V., Brattico, E., Wallentin, M., & Vuust, P. (2014). Capturing the musical brain with lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage, 88*, 170-180.

Vía, J., Santamaría, I., & Pérez, J. (2007). A learning algorithm for adaptive canonical correlation analysis of several data sets. *Neural Networks, 20*(1), 139-152.

Yang, Z., LaConte, S., Weng, X., & Hu, X. (2008). Ranking and averaging independent component analysis by reproducibility (RAICAR). *Human Brain Mapping, 29*(6), 711-725.

Ylipaavalniemi, J., & Soppela, J. (2009). Arabica: Robust ICA in a pipeline. *International Conference on Independent Component Analysis and Signal Separation,* 379-386.

ORIGINAL PAPERS

PI

DIMENSION REDUCTION FOR INDIVIDUAL ICA TO DECOMPOSE FMRI DURING REAL-WORLD EXPERIENCES: PRINCIPAL COMPONENT ANALYSIS vs CANONICAL COMPONENT ANALYSIS

by

Valeri Tsatsishvili, Fengyu Cong, Tuomas Puoliväli, Vinoo Alluri, Petri Toiviainen, Asoke Nandi, Elvira Brattico & Tapani Ristaniemi, 2013.

21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) 2013, pp.137-142

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Dimension Reduction for Individual ICA to Decompose FMRI during Real-World Experiences: Principal Component Analysis vs. Canonical Correlation Analysis

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

1. -University of Jyväskylä, Department of Mathematical Information Technology, PO Box 35, FI-40014, Finland
2. - Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Department of Music, PO Box 35, FI-40014,Finland
3. -Brunel University, Department of Electronic and Computer Engineering, Uxbridge, Middlesex UB8 3PH, UK
4. -University of Helsinki, Institute of Behavioral Sciences, Cognitive Brain Research Unit, PL 9, 466, Finland

Abstract. Group independent component analysis (ICA) with special assumptions is often used for analyzing functional magnetic resonance imaging (fMRI) data. Before ICA, dimension reduction is applied to separate signal and noise subspaces. For analyzing noisy fMRI data of individual participants in free-listening to naturalistic and long music, we applied individual ICA and therefore avoided the assumptions of Group ICA. We also compared principal component analysis (PCA) and canonical correlation analysis (CCA) for dimension reduction of such fMRI data. We found interesting brain activity associated with music across majority of participants, and found that PCA and CCA were comparable for dimension reduction.

1. Introduction

Study of brain activations elicited by natural continuous auditory and visual stimuli is relatively new and a promising domain in the field of fMRI research[1-3].Generated brain responses by such stimuli are of much more complex nature than in commonly utilized controlled design (block or event-related) experiments. This yields to adopting more data-driven approaches rather than holding on more traditional methods following to the hypothesis-driven models[4]. Group ICA has been used for analyzing fMRI during real-world experiences [4,5]. Assumptions for Group ICA require at least the number of sources and their order to be invariant for different subjects [6]. However, it is unknown whether these assumptions are met in real life. Therefore, in this study we apply individual ICA to each participant's fMRI dataset elicited by naturalistic, continuous and long piece of music.

* This work was financially supported by TEKES (Finland) under grant40334/10 "Machine Learning for Future Music and Learning Technologies". A.K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Before subjecting fMRI to ICA decomposition, dimension reduction using PCA with model order selection is a common pre-processing routine that helps in identifying and separating signal and noise subspaces.

Although fairly old method, CCA [7] has only been recently employed for pre-processing [8] or post-processing [5] fMRI data. It finds correlated and uncorrelated subspaces from two datasets using second order statistics [8,9]. In an experiment to collect brain data, it is often expected to find the common information across different participants belonging to the same group. Therefore, CCA theoretically matches this goal, and its strength in the dimension reduction for ICA has been shown through the analysis of simulated and real fMRI data obtained during the controlled design experiment[8]. However, it is unknown whether CCA can also work well as the pre-processing step for ICA to decompose very noisy fMRI data elicited during real-world experiences. Present study compares performances of CCA, implemented according to [8], and more widely used PCA for dimension reduction for ICA.

2. Method

2.1. Data description

Dataset here consists of continuous fMRI scans (time resolution was 2 seconds) obtained from eleven healthy musicians (mean age: 23.2 ± 3.7 SD; 5 females) while they listened to the tango 'Adios Nonino' by Astor Piazzolla with duration of 8 minutes and 32 seconds. Six high-level musical features including Fullness, Brightness, Timbral Complexity, Key Clarity, Pulse Clarity, and Activity were extracted from the stimulus. Detailed information about the fMRI data can be found in [1].

2.2. Dimension Reduction

2.2.1 PCA

If we denote the matrix of observed centered (zero mean) signals by $\boldsymbol{X} \in \mathbb{R}^{n \times l}$, $l \gg n$, then the goal of PCA is to find orthogonal transform diagonalizing the covariance matrix of $\boldsymbol{X}$,$\boldsymbol{C_{xx}} = \frac{\mathbf{1}}{\boldsymbol{n}} \boldsymbol{X X^T}$. This is achieved e.g. by eigenvalue decomposition:

$$\boldsymbol{C_{xx} V = D V},$$

where $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ is a matrix whose each column contains eigenvectors and $\boldsymbol{D}$ is a diagonal matrix of eigenvalues ranked decreasingly.

2.2.2 CCA

While PCA analyses one dataset at a time, CCA analyses two datasets to measure linear relationships between them. It finds two bases $\boldsymbol{W_1}$ and $\boldsymbol{W_2} \in \mathbb{R}^{n \times n}$ for two centered data matrices $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2} \in \mathbb{R}^{n \times l}$, such that correlations between the projections $\boldsymbol{Z_1} = \boldsymbol{Y_1}^T \boldsymbol{W_1}$ and $\boldsymbol{Z_2} = \boldsymbol{Y_2}^T \boldsymbol{W_2}$ are mutually maximized. According to [8] CCA can be calculated by singular value decomposition of cross-covariance matrix of two whitened and normalized datasets:

$$\boldsymbol{C_{y_1 y_2}} = \mathbf{U \Sigma Y}^T,$$

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

where $\boldsymbol{C_{y_1y_2}} = \frac{1}{n}\boldsymbol{Y_1 Y_2}^{\boldsymbol{T}}$ is the cross-covariance matrix, $\mathbf{U}$ and $\mathbf{Y}$ are two orthogonal bases, one for each input dataset, and $\mathbf{\Sigma}$ contains singular values representing the canonical correlations.

2.2.3 Dimension reduction

In the both methods described above, dimension reduction is achieved in similar fashion. For PCA first k eigenvectors are selected from $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ basis, such that it becomes $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n\times k}$, and then input dataset $\boldsymbol{X}$ is projected onto it: $\boldsymbol{\mathfrak{X}} = \boldsymbol{\mathcal{V}^T X}$. The procedure is similar for CCA where $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$ are projected onto $\mathbf{U}$ and $\mathbf{Y}$.

Neither of the presented methods estimates target dimensionality of the input data automatically. In fact, evaluating the number of sources (i.e. target dimensions) is one of the challenges in fMRI analysis, which is frequently solved by empirical approaches [10]. Nevertheless, several methods for estimation of number of sources from the data have been proposed [11,12]. We employed model order selection method Gap proposed in [13] and previously employed for EEG data due to its computing efficiency [14]. With different numbers of sources experimented, the strength of Gap was examined for dimension reduction.

2.3. ICA decomposition

In this study we decompose each participant's fMRI dataset separately using spatial ICA, as opposed to the group-level approach where the data is concatenated first. The model of spatial ICA is $\boldsymbol{x} = \boldsymbol{\mathcal{A}s}$, where $\boldsymbol{x} \in \mathbb{R}^{n\times l}$ is a matrix of fMRI scans (n denotes time points and l - voxels), $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n\times k}$is the mixing matrix containing respective time courses of the sources in $\boldsymbol{s}$, and $\boldsymbol{s} \in \mathbb{R}^{k\times l}$ is the source matrix containing spatial activation patterns. If we denote dataset after dimension reduction by $\mathbf{x}$, then the above model will become determined by $\mathbf{x} = \boldsymbol{\mathcal{V}^T \mathcal{A} s} = \boldsymbol{As}$, where $\boldsymbol{\mathcal{V}^T} \in \mathbb{R}^{k\times n}$ is dimension reduction matrix obtained from the dimension reduction method, and $\boldsymbol{A} = \boldsymbol{\mathcal{V}^T \mathcal{A}}$, $\boldsymbol{A} \in \mathbb{R}^{k\times k}$ becomes the mixing matrix of the determined ICA model. The goal is to learn unmixing matrix $\mathbf{W}$ such that: $\mathbf{y} = \mathbf{Wx}$. After the decomposition, original time courses of extracted sources are reconstructed by projecting extracted sources back to the scan field [15] via $\mathbf{U} = \mathbf{VW^{-1}}$.

As a stochastic algorithm ICA is not intrinsically stable and therefore, it can provide different results if run several times. A software package Icasso [16] analyzes the stability and robustness of ICA decomposition. The idea of Icasso is to run ICA repeatedly N times (N=100 in this study), each time with randomly initialized unmixing matrix and to cluster extracted independent components into the predefined number of clusters. In this study, FastICA algorithm with the nonlinear function *tanh* was selected as the separation algorithm. For the clustering, the agglomerative hierarchical clustering with average-linkage criterion was used. The number of clusters was the same to the number of components extracted by ICA. For characterizing decomposition stability, cluster quality index $\boldsymbol{I_q}$ was calculated, which is a parameter estimating compactness of each cluster and degree of separation from others [16]. It is calculated by:

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

$$I_q = \frac{1}{|C_m|^2} \sum_{i,j \in C_m} \sigma_{ij} - \frac{1}{|C_m||C_{-m}|} \sum_{i \in C_m} \sum_{j \in C_{-m}} \sigma_{ij}$$

where C_m denotes the set of estimated independent components in the cluster m, $|C_m|$ is the size of the cluster, C_{-m} is the set of indices outside the cluster m, and σ_{ij} is an absolute value of mutual correlations between estimated independent components. It is a good measure for estimating stability of the extracted component as well as detecting possible overfitting. Therefore, I_q is a suitable parameter for performance comparison of employed dimension reduction algorithms.

2.4. Individual-level data processing

Obtained fMRI images went through the pre-processing procedure described in [1]. Next, temporal course of each voxel in the dataset was filtered using digital filter based on Fourier transform. The cut-off frequencies of the band-pass filter were set to 0.008Hz and 0.05Hz, determined by power spectrum of stimulus feature time series.

Dimension of the filtered data was reduced using two different methods. First, PCA and Gap were employed, where Gap estimated 46 sources. Next, CCA was performed on six pairs of subjects. We implemented CCA according to the algorithm proposed in [8]. However, for dimension reduction authors in [8] rejected CCA components with correlations below an arbitrary threshold of 0.5. Here we employed Gap method again that determined different number of sources for different pairs of datasets, varying between 43 and 45. To test if Gap performance was optimal we also experimented with different numbers of sources (k=20 and k=30).

Resulted six datasets (three for each dimension reduction method) were separately decomposed using Icasso [16].

2.5. Group-level data analysis

Obtained time courses of independent components were correlated with time courses of stimulus features. Significance thresholds of the correlations were set using Monte-Carlo simulation presented in [1] and only significant correlations at the significance level $p<0.01$ were considered for further analysis. Finally, spatial maps with significant correlations were visually examined to find common stimulus-related brain activations. We considered common activation map only if it was shared between more than five (half of all) participants.

3. Results

For compactness of representation we denote CCA and PCA-based ICA results as PCA+ICA and CCA+ICA. Experiments showed that ICA decomposition stability is affected little by employed dimension reduction method. In the Fig.1 quality indexes for CCA and PCA are provided. Indeed, for all numbers of components the difference between two methods for mean ICA decomposition stability is subtle.

Visual examination of activation maps significantly correlated with one or more musical features ($p<0.01$) revealed one common map showing activation in Auditory cortex, shared between more than five participants. Table 1 summarizes the observed

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

common map for PCA and CCA. In overall, the spatial map was detected in activations of nine subjects for PCA+ICA and seven subjects for CCA+ICA. Due to the space limitation, spatial maps are not shown.

Manually reducing dimensionality to 20 and 30 resulted in less stable ICA decomposition for CCA as well as PCA. However, desired common map was still observed for both methods: for PCA+ICA among seven and six participants respectively. For CCA+ICA the common map was found in 7 participants' activations regardless of the number of sources.

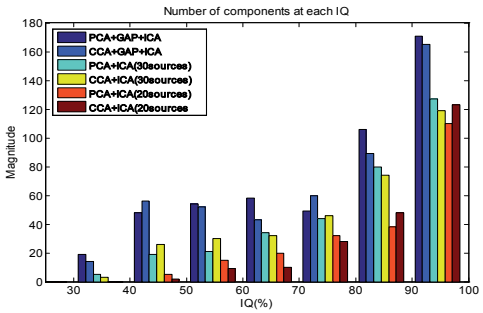

Fig 1: Cluster quality indexes for CCA and PCA for GAP,30, and 20 components.

Feature	Subject numbers	
	CCA+GAP+ICA	PCA+GAP+ICA
Fulness	3,4,5,10	3,7,10
Brightness	3,4,5,7,9,11	1,3,4,5,7,8,9,10,11
Timb. Complexity	0	8,11
Key Clarity	0	3
Pulse Clarity	0	0
Activity	3,4,5,9,10,11	3,4,7,10

Table 1: Summary of common spatial map. Numbers represent numbering of subjects and zero denotes absence of participant for which observed map was significantly correlated with acoustic features.

4. Conclusions

In order to study fMRI during real-world experiences, we proposed an individual-level data processing and group-level analysis approach mainly based on ICA and correlation. Meanwhile, two different methods for dimensionality reduction were tested for ICA in processing such challenging data.

We found similar spatial maps with corresponding temporal courses significantly correlated with musical features among individual participants. For dimension reduction in processing fMRI during real-world experiences, we found both PCA and CCA performed reasonably well.

In addition, we repeated the process with two different numbers of sources to check whether the employed model order selection was optimal in estimating number of target dimensions. We found that the number of sources suggested by model order selection was optimal for ICA decomposition stability for both methods. Interestingly, in production of stimulus-related spatial maps CCA was less sensitive to lower dimensions than PCA in our experiment.

It should be noted that CCA was implemented according to [8], which does not precisely follow the conventional CCA definition [7]. In the future we will investigate

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

the conventional CCA and Partial least squares [17] for dimension reduction of fMRI during real-world experiences.

References

[1] V. Alluri, P. Toiviainen, I.P. Jaaskelainen, E. Glerean, M. Sams, E. Brattico, Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm, Neuroimage. 59 (2012) 3677-3689.
[2] J.P. Kauppi, I.P. Jääskeläinen, M. Sams, J. Tohka, Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency, Frontiers in neuroinformatics. 4 (2010) DOI: 10.3389/fninf.2010.00005.
[3] K.N. Carvalho, G.D. Pearlson, R.S. Astur, V.D. Calhoun, Simulated Driving and Brain Imaging, CNS Spectr. 11 (2006) 52-62.
[4] S. Malinen, Y. Hlushchuk, R. Hari, Towards natural stimulation in fMRI—Issues of data analysis, Neuroimage. 35 (2007) 131-139.
[5] J. Ylipaavalniemi, E. Savia, S. Malinen, R. Hari, R. Vigário, S. Kaski, Dependencies between stimuli and spatially independent fMRI sources: Towards brain correlates of natural stimuli, Neuroimage. 48 (2009) 176-185.
[6] F. Cong, Z. He, J. Hämäläinen, P.H.T. Leppänen, H. Lyytinen, A. Cichocki, T. Ristaniemi, Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection, J. Neurosci. Methods. 212 (2013) 165-172.
[7] H. Hotelling, Relations between two sets of variates, Biometrika. 28 (1936) 321-377.
[8] J. Karhunen, T. Hao, J. Ylipaavalniemi, A canonical correlation analysis based method for improving BSS of two related data sets, Latent Variable Analysis and Signal Separation. Lecture Notes in Computer Science. 7191 (2012) 91-98.
[9] M. Koskinen, J. Viinikanoja, M. Kurimo, A. Klami, S. Kaski, R. Hari, Identifying fragments of natural speech from the listener's MEG signals, Hum. Brain Mapp. DOI: 10.1002/hbm.22004 (2012).
[10] V. van de Ven, C. Bledowski, D. Prvulovic, R. Goebel, E. Formisano, F. Di Salle, D.E.J. Linden, F. Esposito, Visual target modulation of functional connectivity networks revealed by self-organizing group ICA, Hum. Brain Mapp. 29 (2007) 1450-1461.
[11] D. Cordes, R.R. Nandy, Estimation of the intrinsic dimensionality of fMRI data, Neuroimage. 29 (2006) 145-154.
[12] Y.O. Li, T. Adalı, V.D. Calhoun, Estimating the number of independent components for functional magnetic resonance imaging data, Hum. Brain Mapp. 28 (2007) 1251-1266.
[13] Zhaoshui He, A. Cichocki, Shengli Xie, Kyuwan Choi, Detecting the Number of Clusters in n-Way Probabilistic Clustering, Pattern Analysis and Machine Intelligence, IEEE Transactions on. 32 (2010) 2006-2021.
[14] F. Cong, Z. He, J. Hamalainen, A. Cichocki, T. Ristaniemi, Determining the number of sources in high-density EEG recordings of event-related potentials by model order selection, (IEEE MLSP2011) 1-6.
[15] F. Cong, I. Kalyakin, Z. Chang, T. Ristaniemi, Analysis on subtracting projection of extracted independent components from EEG recordings, Biomed. Tech. (Berl). 56 (2011) 223-234.
[16] J. Himberg, A. Hyvarinen, F. Esposito, Validating the independent components of neuroimaging time series via clustering and visualization, Neuroimage. 22 (2004) 1214-1222.
[17] H. Wold, Soft modelling : The Basic Design and Some Extensions, Systems Under Indirect Observation, Part II. (1982) 36-37.

PII

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

by

Valeri Tsatsishvili, Fengyu Cong, Tapani Ristaniemi, Petri Toiviainen, Vinoo Alluri, Elvira Brattico, & Asoke Nandi, 2014

22nd European Signal Processing Conference (EUSIPCO), Proceedings of the, 2014, 2490-2494

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 reali-

ty 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 syn-

chronize 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:

$$\mathbf{CU} = \boldsymbol{\lambda}\mathbf{U} \quad (1)$$

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

$$\mathbf{C} = \frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X} \quad (2)$$

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:

$$\mathbf{Y} = \mathbf{U}^{\mathrm{T}}\mathbf{X} \quad (3)$$

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 $\mathcal{F}$:

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

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:

$$\widetilde{\mathbf{K}}\mathbf{a} = \tilde{\lambda}\mathbf{a} \quad (6)$$

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, $\tilde{\lambda}_k, \boldsymbol{a}^k,\ k = 1, \ldots n$ represent k-th eigenvectors and eigenvalues. The projections of points in the feature space $\boldsymbol{\Phi}(\mathbf{X})$ onto the eigenvectors are given by:

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

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 activati-

on 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

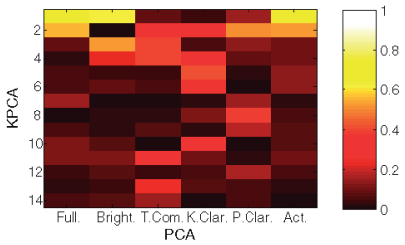

Fig. 1. Correlation coefficients (absolute value) between KPCA and PCA features.

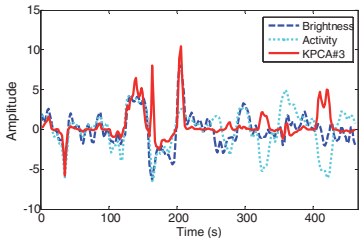

Fig. 2. Temporal courses of Brightness, Activity and KPCA#3.

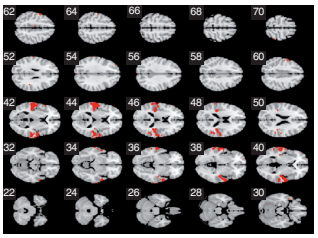

Fig. 3. Common spatial map based on the new feature, KPCA#3.

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, "Inter-subject 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.

PIII

COMBINING PCA AND MULTISET CCA FOR DIMENSION REDUCTION WHEN GROUP ICA IS APPLIED TO DECOMPOSE NATURALISTIC FMRI DATA

by

Valeri Tsatsishvili, Fengyu Cong, Petri Toiviainen, & Tapani Ristaniemi, 2015

Neural Networks (IJCNN), International Joint Conference On, 2015, 1-6

Combining PCA and Multiset CCA for Dimension Reduction when Group ICA is Applied to Decompose Naturalistic fMRI Data

Valeri Tsatsishvili[1], Fengyu Cong[1,2], Petri Toiviainen[3], Tapani Ristaniemi[1]
[1]Department of Mathematical Information Technology
University of Jyvaskyla
P.O. Box 35, FI-40014, University of Jyvaskyla, Finland
valeri.v.tsatsishvili@jyu.fi

[2]Department of Biomedical Engineering
Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology, China

[3]Department of Music
University of Jyvaskyla
P.O. Box 35, FI-40014, University of Jyvaskyla, Finland

***Abstract*—An extension of group independent component analysis (GICA) is introduced, where multi-set canonical correlation analysis (MCCA) is combined with principal component analysis (PCA) for three-stage dimension reduction. The method is applied on naturalistic functional MRI (fMRI) images acquired during task-free continuous music listening experiment, and the results are compared with the outcome of the conventional GICA. The extended GICA resulted slightly faster ICA convergence and, more interestingly, extracted more stimulus-related components than its conventional counterpart. Therefore, we think the extension is beneficial enhancement for GICA, especially when applied to challenging fMRI data.**

***Keywords*—Group ICA; temporal concatenation; naturalistic fMRI; dimension reduction; Multiset CCA; PCA**

I. Introduction

Naturalistic neuroimaging experiments are increasingly utilized by neuroscience community to study cognitive brain function [1-6]. In such experiments real-world experiences are reproduced in laboratory conditions, enabling more ecologically valid stimulation than in common controlled experimental paradigms. From the methodological perspective, analysis of naturalistic functional magnetic resonance imaging (fMRI) data involves new challenges associated with less control over stimulus and response timing, and lack of predefined models for such complex brain responses. That makes data-driven analysis approaches preferable over traditional hypothesis-driven ones due to their flexibility and few assumptions on the nature of data and noise. Independent component analysis (ICA) is arguably the most widely applied data-driven method to fMRI data. The ICA model assumes the observed fMRI data to be a linear mixture of several independent and non-Gaussian sources, and extracts them from observations. Various extensions of ICA exist nowadays.

Group ICA (GICA) is commonly applied ICA variant on fMRI due to its advantage in drawing group inferences from multi-subject datasets. Within Group ICA, several approaches with different assumptions and data grouping strategies have been developed [1]. In this study, GICA with temporal concatenation was selected [7], since it was shown to outperform other strategies through the simulation study in [8]. Conventional GICA method [7] consists of three main stages: first, the dimensionality of data from each subject is reduced using principal component analysis (PCA). Second, the reduced datasets are concatenated across time dimension and PCA is applied again to further reduce dimensions. Third, the thus reduced dataset is subjected to ICA decomposition. As a result, one mixing matrix is estimated containing partitions corresponding to each subject, and one set of independent components (i.e. spatial maps) - that are common to all subjects. Spatial maps for each subject can also be reconstructed for exploring between-subject differences within each aggregate component. Although the described model has been successfully applied on fMRI before (see e.g. [1] for review), it does not always perform well, as demonstrated on our naturalistic fMRI data in the present study as well as in [9].

In this work we extended the described conventional GICA model by integrating multiset canonical correlation

analysis (MCCA) as an additional pre-processing step before subjecting data to ICA decomposition. Multiset CCA is the multiset generalization of standard two-set CCA introduced in [10], which finds correlated subspace from different but related datasets using second-order statistics [11]. In fMRI experiments finding common brain activation patterns among different subjects is important. Therefore, finding correlated subspace in the data before applying ICA is well justified for our purposes. The extension of blind source separation (BSS) by first finding correlated subspace was proposed in [12] and was reported to be significantly better than ICA alone in separating simulated mixtures. It should be noted, however, that the model we introduce here is not identical with the one applied in [12]. The main difference is in the ICA approaches employed. We selected group ICA with temporal concatenation strategy, whereas in [12] ICA was applied on each dataset separately. Furthermore, MCCA implementation in the present study is from [13], which in turn follows the original publication [11]. On the other hand, authors in [12] offered their own implementation of MCCA (or generalized CCA, as referred in the article). Nevertheless, the main principle of finding the correlated subspace in the data prior subjecting to ICA decomposition is the same. Interestingly, even though Karhunen et al. also included experiments on real fMRI data in [12], only two-set CCA was applied as source separation method rather than as a preprocessing of ICA. In this study, the developed GICA extension was applied on challenging fMRI data acquired during task-free continuous music listening experiment and the results were compared with outcome of conventional GICA.

A crucial question for any ICA-based approach is how many sources should be extracted from the observed data. Estimation of number of sources has tremendous influence on the success of the decomposition. The authors in [7] recommended commonly applied heuristics (such as selecting number of components explaining most of the variance in the data) as well as information-theoretic criteria for estimation of number of sources from the aggregate dataset. However, in the recent study Cong et al. [14] showed through simulations that popular information-theoretic methods, such as Akaike's information criterion [15] and minimum distance length [16] fail to accurately estimate number of sources when the SNR is low. In fact, the suboptimal performance of estimators was later demonstrated in practice, where model order selection methods failed to estimate reasonable number of sources for GICA with temporal concatenation on the same naturalistic fMRI data as here [9]. To address this issue, we took a simple empirical approach for estimating the number of sources. Specifically, we examined GICA results for different number of sources to find the most appropriate number. Two general criteria were used for evaluating and comparing the results from different trials: first, the stability and reliability of ICA decomposition, and second, quality of the produced independent components. We devised six parameters, which provide concise description of ICA decomposition results according to the two criteria. The parameters, detailed in the section II.*D*, are simple and straightforward to compute from ICA decomposition results. The proposed evaluation heuristics can be adapted for testing and comparison of different ICA parametrizations in any application.

To summarize, we developed an extension of conventional GICA, where correlated subspace from all subject data is found prior concatenating and separating sources. The model was applied to challenging naturalistic fMRI data, and the results were compared to conventional GICA. In addition, we proposed a set of parameters that enables evaluation and comparison of different parametrizations in terms of the two criteria.

II. Method

A. 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. The fMRI measurements were made in 3T scanner at sampling frequency of 0.5 Hz. Obtained fMRI scans went through common preprocessing routine, which is described in [6]. Overall, 231 fMRI scans corresponding to stimulus between 21 to 480 seconds were used for analysis. Six high-level features were obtained from stimulus audio, capturing timbral rhythmic and tonal information in music. The feature set, consisting of fullness, brightness, timbral complexity, pulse clarity, key clarity, and activity, were generated as a linear combination of 25 long-term and short-term features via PCA. The multi-stage process of their extraction and perceptual validation is described in [6].

B. Group ICA extension

In this section, extension of GICA is introduced. For description of the conventional GICA algorithm with two-stage PCA reduction refer to [7]. The entire model can be divided into three-stage dimensionality reduction, followed by source separation by ICA. First, PCA separates signal and noise subspaces in each dataset. Next, MCCA is applied to select correlated subspace across all subjects' responses. Particularly, it extracts an orthogonal set of canonical components from each subject data such that the canonical components are correlated across different datasets only on corresponding indices. Subsequently, the correlated subspace is extracted by selecting the canonical components that are correlated above the predefined threshold. The third stage dimensionality reduction is applied on concatenated data using PCA. The main reason is that number of signals in the concatenated data are commonly assumed to be larger than number of sources. In other words, we have the overdetermined data model, where the number of sources in the mixture is less than number of signals (or samples, statistically speaking). Therefore, prior subjecting to source separation, dimension reduction needs to be applied in order to produce determined mixtures with equalized sources and samples. Finally, BSS is achieved by ICA, extracting statistically independent sources in the concatenated data. The approach consists of the following steps:

Dimensionality reduction:

1. Reduce dimensionality of each dataset using PCA
2. Apply MCCA to select correlated subspace from each dataset
3. Concatenate reduced data from all subjects

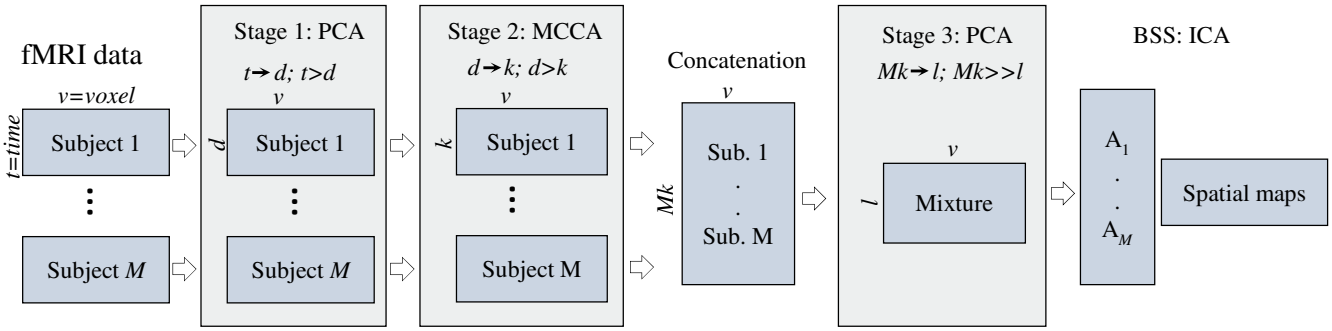

Fig. 1. Schematic view of the extended GICA model. Dimensions of data matrices after each stage are shown.

4. Reduce dimensionality of concatenated data using PCA

Blind source separation:

5. Apply ICA and get sources and unmixing matrix
6. Reconstruct time courses and spatial maps

Fig.1 depicts more detailed view of the entire processing chain. In the figure dimensionalities of data matrices after each step are shows for more clarity.

More formal description of the model is as follows. Let us denote dataset from each subject by $\mathbf{X}_m \in \mathbb{R}^{t\times v}$, where $m = [1, 2, .. M]$ denotes number of subjects, t is number of fMRI scans and v is number of voxels. First, dimensions of each dataset is reduced using PCA. If we denote dimension reduction matrix (i.e. selected d eigenvectors) as $\mathbf{V}_m \in \mathbb{R}^{t\times d}$, then the reduced data for each subject (i.e. projection of data vectors to a new space) can be expressed as:

$$\mathbf{Z}_m = \mathbf{V}_m^{-1}\mathbf{X}_m \tag{1}$$

Note that pseudoinverse will be assumed in the cases when a matrix to be inverted is not square. Now $\mathbf{Z}_m$ becomes an input for MCCA. Solving MCCA problem can be considered as finding orthogonal matrix $\mathbf{B}_m$ and canonical components are obtained by projecting the input data to $\mathbf{B}_m$:

$$\mathbf{U}_m = \mathbf{B}_m^{-1}\mathbf{Z}_m = \mathbf{B}_m^{-1}\mathbf{V}_m^{-1}\mathbf{X}_m = \mathbf{D}_m\mathbf{X}_m \tag{2}$$

where, $\mathbf{D}_m = \mathbf{B}_m^{-1}\mathbf{V}_m^{-1}$ and $\mathbf{U}_m$ contains components, such that the canonical correlations, or correlation between corresponding components among different subjects are in decreasing order. The dimensions are reduced from d to k by selecting canonical components showing k highest mean correlations. Subsequently, datasets from all subjects are concatenated. The concatenated dataset $\widetilde{\boldsymbol{U}} \in \mathbb{R}^{Mk\times v}$ consists of $M * k$ samples and is an overdetermined mixture. Then, dimensionality of the data is reduced using PCA to make the mixture determined:

$$\widetilde{\boldsymbol{U}} = \boldsymbol{\mathcal{V}}^{-1}\boldsymbol{U} \tag{3}$$

where, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{Mk\times l}$ is a dimension reduction matrix. Finally, $\widetilde{\boldsymbol{U}}$ is subjected to ICA decomposition to find sources of activations as independent components (IC) and mixing matrix, or temporal courses of IC-s. The model ICA follows is $\widetilde{\boldsymbol{U}} = \mathbf{AS}$. In practice, ICA algorithm estimates the unmixing matrix $\boldsymbol{\mathcal{W}} = \mathbf{A}^{-1} \in \mathbb{R}^{l\times l}$ and IC-s such that:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{W}}\widetilde{\boldsymbol{U}} \tag{4}$$

According to ICA model applied here, aggregate unmixing matrix containing partitions unique to each subject, while the independent components are common. Reconstruction of temporal courses in the reduced data space is done by simply inverting the estimated unmixing matrix. However, we are interested in time courses in the original scan space. It can be achieved by inserting (3) in (4), partitioning aggregate data for each subject, and then inserting (2):

$$\boldsymbol{\mathcal{Y}}_m = \boldsymbol{\mathcal{W}}_m\boldsymbol{\mathcal{V}}_m^{-1}\mathbf{D}_m\mathbf{X}_m = \mathbf{W}_m\mathbf{X}_m \tag{5}$$

Here, $\boldsymbol{\mathcal{Y}}_m$ denotes reconstruction of subject-specific spatial maps. Then, temporal courses can be estimated by:

$$\mathbf{T}_m = \mathbf{W}_m^{-1} = \mathbf{D}_m^{-1}\boldsymbol{\mathcal{V}}_m\boldsymbol{\mathcal{W}}_m \tag{6}$$

For ICA calculation FastICA algorithm was used [17].

C. Analysis of extracted components by ICA

Within the temporal concatenation approach, each extracted common activation pattern may feature different temporal dynamics among subjects. The aim is to find the common maps temporally correlated with the stimulus. To this end, we first correlate subject-specific temporal courses of each IC with the time courses of the six stimulus features. The significance thresholds for correlations are estimated using Monte-Carlo simulation presented in [6] and the level of significance selected throughout this study is $p<0.01$. If majority of the subject-specific temporal courses of a given IC are significantly correlated with the temporal course of any of the stimulus features, then the IC is considered to be stimulus-related and is selected for further analysis. At this point we reconstruct subject-specific spatial maps of the selected IC-s in order to observe individual differences in activation patterns. Ideally, we would have one or more similar activation patterns from all the subjects per acoustic feature. However, in practice similar spatial patterns from different subjects are not synchronized in time and show heterogeneous correlations with acoustic features, which makes them difficult to interpret.

TABLE 1. Evaluation parameters for MCCA+GICA vs conventional GICA (last entry in the table). The output dimensionalities of the first and the third stages (PCA) are fixed to 80 and 40 respectively, while the output of the second stage (i.e. MCCA) is varied from 10 to 60

MCCA dimensions	Converged runs	Convergence rate (mean/SD)	Cluster quality index (Mean/SD)	Selected independent components	Common maps (feature/sub)	Active voxels in common map (left/right hemisphere)
MCCA-10	66	82/11	0.64/0.21	37	-	-
MCCA-20	72	74/13	0.78/0.19	27	-	-
MCCA-40	96	64/17	0.9/0.09	24	-	-
MCCA-60	100	59/14	0.89/0.1	31	1 (Bright./6 sub)	2317/2283
No MCCA	99	65/14	0.9/0.11	29	-	-

D. Evaluation of different parametrisations

Two general criteria were used for evaluation and comparison of the results from different trials. The first criterion is stability and reliability of ICA decomposition. Stochastic nature of ICA-based decomposition renders direct comparison of results unreliable. To address the issue the software package ICASSO [18] was utilized. This tool has been designed for analysis of the stability and robustness of ICA decomposition. Essentially ICASSO runs ICA repeatedly N times (N=100 in this study), each time with randomly initialized unmixing matrix, clusters the extracted independent components, and provides multiple parameters for observing and visualizing the clustering and separation results.

One interesting parameter in our list is ICASSO cluster quality index:

$$I_q = \frac{1}{|C_m|^2}\sum_{i,j\in C_m}\sigma_{ij} - \frac{1}{|C_m||C_{-m}|}\sum_{i\in C_m}\sum_{j\in C_{-m}}\sigma_{ij}$$

where C_m denotes the set of estimated independent components in the cluster m, $|C_m|$ is the size of the cluster, C_{-m} is the set of indices outside the cluster m, and σ_{ij} is an absolute value of mutual correlations between estimated independent components extracted in different runs. Cluster quality index characterizes compactness and separation of clusters and is a good measure for estimating stability of the extracted component as well as detecting possible overfitting.

We propose three parameters for assessing the stability and reliability of ICA decomposition: 1. number of converged runs (out of N=100). If ICA decomposition does not converge, the estimation of mixing matrix is not reliable. 2. mean/standard deviation of convergence rate across N runs. Convergence rate is the number of steps required for convergence. 3. Cluster quality index.

Since a stable ICA decomposition does not guarantee good quality of produced independent components, it is necessary to quantify the quality of the output in order to make comparison between different ICA decompositions possible. Last three of the six parameters assess the quality of produced components and consist of: 4. Number of the selected components, 5. Number of the common maps, and 6. Number of the active voxels in a common map(s).

III. RESULTS

For simplicity we denote conventional group ICA as simply GICA, and the extension introduced in this study - as MCCA+GICA. Furthermore, to differentiate between the two stage PCA reductions, the term individual PCA will be applied to the reduction of each subject's data, while group PCA will refer to dimension reduction of concatenated data.

We first investigated different parametrizations of GICA in order to select appropriate number of sources. To this end, the number of dimensions was consecutively fixed for one of the two PCA stages while varying the other, and examined the GICA output. For individual PCA we wanted to retain most of the variance data and reduced the dimension from 231 to 80, explaining about 88% of variance in average over all subjects. As the tests involving different numbers of dimensions showed, individual PCA did not have as major influence on ICA decomposition as group PCA. It is expected because individual PCA does not define number of sources to be extracted by ICA. However, it increases SNR by separating signal and noise subspaces, and therefore, intuitively will have more influence on quality of independent components. For group PCA, conversely, significant influence on ICA decomposition was observed. Intuitively, it is probably related to the fact that it directly determines the number of sources in the mixture and suboptimal estimation of sources has devastating effect on ICA decomposition. Indeed, in our experiments ICA decomposition always failed above certain number of sources, regardless of the settings of individual PCA, or MCCA. For GICA, none of the tested number of dimensions for each PCA stage led to finding common spatial maps. Hence, it was difficult to evaluate quality of the produced independent components except counting the number of selected ones. Nevertheless, the first three evaluation parameters were helpful for dramatically narrowing down the range of all possible number of sources to a few. As a result, we reduced data to 80 and 40 dimensions after individual and group PCA respectively. Next, the developed MCCA+GICA extension was applied and different sizes of correlated subspace were tested. Throughout the tests, individual and concatenated PCA reduction outputs were fixed to 80 and 40 respectively, since this parametrization produced the best results and we wanted any other parameters except MCCA to be similar among GICA and MCCA+GICA. Average canonical correlations corresponding to the selected

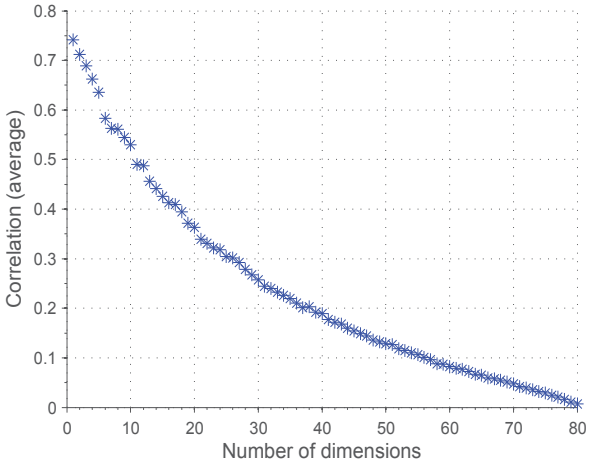

Fig. 2. Average correlations between canonical components

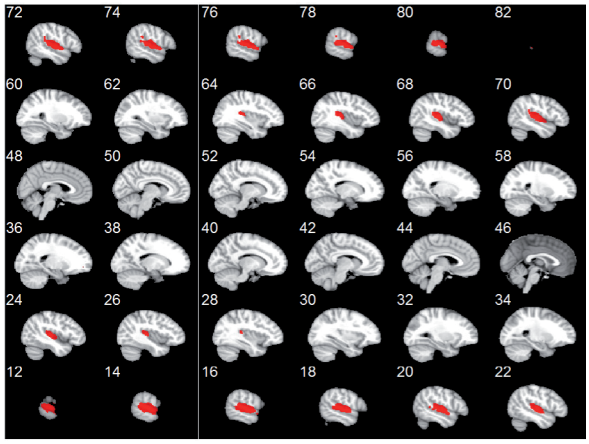

Fig. 3. Common spatial map, averaged across six contributing subjects

number of canonical components are shown in Fig.2. As depicted in the figure, the examined correlated subspace sizes (see Table 1) correspond to a wide range of thresholds for canonical correlations varying from fairly strict (above 0.5) to very liberal (above 0.1). In Table 1, six parameters summarizing MCCA+GICA output for different sizes of correlated subspace are shown. For comparison, GICA results for the same PCA outputs are also provided (labelled as 'No MCCA'). In terms of ICA stability, it is evident that increasing the size of correlated subspace improves ICA convergence (see Table 1). However, the rate of improvement quickly decreases, and for 40 and 60 components (MCCA-40 and MCCA-60 in the table) first three evaluation parameters are very similar. And yet, the parameters related to quality show more contrasting picture. For MCCA-40, none of the stimulus features were selected and common maps were not extracted, whereas for MCCA-60 two acoustic features including Brightness and Activity were selected, and common map related to Brightness was found with six contributing subjects. The common map in Fig.3 shows large clusters of bilateral auditory cortex activations, which is in line with previous findings [6, 9].

To summarize, the empirically selected parametrization of GICA after multiple tests was the following: individual PCA reduced dimensions from 231 to 80 and group PCA reduced aggregate data from 880 to 40. For MCCA+GICA: individual PCA reduced data from 231 to 80, MCCA reduced each dataset from 80 to 60, and group PCA – from 660 to 40. Based on our evaluation parameters (see Table 1), it can be seen that MCCA+GICA shows improvement over conventional GICA both in ICA convergence and quality of produced components.

IV. Discussion

In this study we introduced MCCA-based extension of GICA, applied it on challenging fMRI data and compared the obtained results with the outcome from the conventional GICA. Interestingly, the latter failed to extract stimulus-related common spatial maps for any of the tested parametrizations.

The reason behind improved results by MCCA+GICA is that finding correlated subspace reduces complexity of sources and therefore, has positive effect on inter-subject variability in terms of the temporal courses. As a result, for a given component, temporal courses from more than half of the subjects showed significant correlations with acoustic features. Reduced complexity of sources also explains faster convergence of ICA decomposition observed for the extended GICA model. In favour of this argument, Karhunen et al. [12] indicated that CCA alone already provides separation of sources at some degree using second-order statistics. Subsequently, in the next stage, partly separated sources form the new mixture of less complex sources in turn is decomposed further by ICA using higher-order statistics. Hence, benefits of introduced extension is expected and justified.

In addition, we proposed means for quantifying output of ICA-based method, in order to: *a*) estimate reasonable number of sources for ICA-based methods, should an automatic model order selection method fail. *b*) compare results from multiple different ICA parametrizations. A somewhat intuitive explanation of why evaluation parameters are useful for estimating number of sources stems again from the fact that suboptimal estimation of number of sources has strong influence on ICA decomposition. Over-estimated number of sources leads to complete failure of ICA decomposition, which will be easily reflected in first three evaluation parameters. Under-estimated number of sources, on the other hand, is primarily exhibited in reduced quality of produced independent components, and will probably be detected by last three parameters. It should be noted that the proposed set of evaluation parameters is still under the development. For example, parameters such as convergence rate and converged rounds can be improved by normalizing with respect to the number of sources.

Temporal concatenation approach assumes common spatial maps and individual time courses among subjects. The reconstructed subject-specific spatial activation patterns were very similar as expected, but within the standard model only part of the corresponding temporal courses were significantly correlated with the stimulus features. Large variability in temporal dynamics among subjects' responses elicited from a complex stimulation is presumably the major contributing factor to the failure of the conventional temporal

concatenation approach. In light of this, and with the supporting evidence in [9], perhaps spatial concatenation approach with the constraint of common mixing matrix for all subjects would be more appropriate for the specific cases such as the one considered here, although the opposite has been suggested previously [8]. We are planning to test the extension of spatial concatenation strategy with MCCA.

V. Acknowledgment

This work was financially supported by TEKES (Finland) under grant 40334/10, Department of Mathematical Information Technology at the University of Jyvaskyla, Academy of Finland (project 272250), the Fundamental Research Funds for the Central Universities [DUT14RC(3)037], XingHai Scholar in Dalian University of Technology in China, and National Natural Science Foundation of China (Grant No. 81471742). The first and the second author contributed equally to this work. The authors would also like to thank to Elvira Brattico and Asoke Nandi for their contribution.

VI. References

[1] V. D. Calhoun, J. Liu and T. Adalı, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *Neuroimage,* vol. 45, pp. S163-S172, 2009.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] V. Calhoun, T. Adali, G. Pearlson and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum. Brain Mapp.,* vol. 14, pp. 140-151, 2001.

[8] V. J. Schmithorst and S. K. Holland, "Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data," *Journal of Magnetic Resonance Imaging,* vol. 19, pp. 365-368, 2004.

[9] 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.

[10] H. Hotelling, "Relations between two sets of variates," *Biometrika,* pp. 321-377, 1936.

[11] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika,* vol. 58, pp. 433-451, 1971.

[12] J. Karhunen, T. Hao and J. Ylipaavalniemi, "A generalized canonical correlation analysis based method for blind source separation from related data sets," in *Neural Networks (IJCNN), the 2012 International Joint Conference On,* 2012, pp. 1-9.

[13] Y. Li, T. Adali, W. Wang and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *Signal Processing, IEEE Transactions On,* vol. 57, pp. 3918-3929, 2009.

[14] F. Cong, A. K. Nandi, Z. He, A. Cichocki and T. Ristaniemi, "Fast and effective model order selection method to determine the number of sources in a linear transformation model," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European,* 2012, pp. 1870-1874.

[15] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions On,* vol. 19, pp. 716-723, 1974.

[16] J. Rissanen, "Modeling by shortest data description," *Automatica,* vol. 14, pp. 465-471, 1978.

[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.

PIV

COUPLING OF ACTION-PERCEPTION BRAIN NETWORKS DURING MUSICAL PULSE PROCESSING: EVIDENCE FROM REGION-OF-INTEREST-BASED INDEPENDENT COMPONENT ANALYSIS

by

Iballa Burunat, Valeri Tsatsishvili, Elvira Brattico, & Petri Toiviainen, 2017

Frontiers in Human Neuroscience, 2017, vol.11

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 May 2017
doi: 10.3389/fnhum.2017.00230

Coupling of Action-Perception Brain Networks during Musical Pulse Processing: Evidence from Region-of-Interest-Based Independent Component Analysis

Iballa Burunat[1]*, *Valeri Tsatsishvili*[2], *Elvira Brattico*[3] *and Petri Toiviainen*[1]

[1] Department of Music, Arts and Culture Studies, Finnish Centre for Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland, [2] Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland, [3] Department of Clinical Medicine, Center for Music in the Brain, Aarhus University and The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

Our sense of rhythm relies on orchestrated activity of several cerebral and cerebellar structures. Although functional connectivity studies have advanced our understanding of rhythm perception, this phenomenon has not been sufficiently studied as a function of musical training and beyond the General Linear Model (GLM) approach. Here, we studied pulse clarity processing during naturalistic music listening using a data-driven approach (independent component analysis; ICA). Participants' (18 musicians and 18 controls) functional magnetic resonance imaging (fMRI) responses were acquired while listening to music. A targeted region of interest (ROI) related to pulse clarity processing was defined, comprising auditory, somatomotor, basal ganglia, and cerebellar areas. The ICA decomposition was performed under different model orders, i.e., under a varying number of assumed independent sources, to avoid relying on prior model order assumptions. The components best predicted by a measure of the pulse clarity of the music, extracted computationally from the musical stimulus, were identified. Their corresponding spatial maps uncovered a network of auditory (perception) and motor (action) areas in an excitatory-inhibitory relationship at lower model orders, while mainly constrained to the auditory areas at higher model orders. Results revealed (a) a strengthened functional integration of action-perception networks associated with pulse clarity perception hidden from GLM analyses, and (b) group differences between musicians and non-musicians in pulse clarity processing, suggesting lifelong musical training as an important factor that may influence beat processing.

Keywords: functional magnetic resonance imaging (fMRI), Independent Component Analysis (ICA), rhythm perception, musicians, naturalistic, prediction

OPEN ACCESS

Edited by:
Stephane Perrey,
Université de Montpellier, France

Reviewed by:
Alessandro Tavano,
Max Planck Institute, Germany
Fabiana Mesquita Carvalho,
University of São Paulo, Brazil
Anne Danielsen,
University of Oslo, Norway

***Correspondence:**
Iballa Burunat
iballa.burunat@jyu.fi

Received: *09 October 2016*
Accepted: *21 April 2017*
Published: *09 May 2017*

Citation:
Burunat I, Tsatsishvili V, Brattico E and Toiviainen P (2017) Coupling of Action-Perception Brain Networks during Musical Pulse Processing: Evidence from Region-of-Interest-Based Independent Component Analysis. Front. Hum. Neurosci. 11:230. doi: 10.3389/fnhum.2017.00230

INTRODUCTION

Pulse may be defined as an endogenous periodicity, a series of regularly recurring, precisely equivalent psychological events that arise in response to a musical rhythm (Cooper and Meyer, 1960; Large and Snyder, 2009). Although rhythms in music do not hold one-to-one relationships with auditory features (Kung et al., 2013), humans are able to effortlessly perceive the pulse

in music. This phenomenon keeps challenging cognitive scientists, who pursue understanding of its underlying brain processes (Gabrielsson, 1987; Clarke, 1989; Palmer, 1989; Repp, 1990). This unique ability to perceive pulse allows us to coordinate motor movements to an external auditory stimulus (such as in music-induced foot tapping or dancing). Moving in synchrony with the beat (i.e., the predictive, perceived pulse in music; Patel, 2014) is in fact one of the most intriguing effects of music and a spontaneous behavior which has long puzzled scientists (Zentner and Eerola, 2010; Repp and Su, 2013). Furthermore, rhythm perception is fundamental to the experience of music and thus key for explaining musical behavior (Large and Palmer, 2002; Large and Snyder, 2009).

Beat perception in auditory rhythms is underpinned by interactions between activity in the auditory and motor systems (Zatorre et al., 2007; Grahn, 2009; Kung et al., 2013), which in particular may drive the temporal predictions involved in rhythm perception (Zatorre et al., 2007; Patel and Iversen, 2014). Recent fMRI evidence indicates that listening to salient rhythms in the absence of any overt movement recruits a cortico-subcortical functional network consisting of auditory cortex, premotor cortex (PMC), putamen (PUT), and supplementary motor area (SMA; Grahn and Rowe, 2009). In addition to the SMA and PMC, the cerebellum (CER) has been found to be active while listening to rhythms (Chen J. L. et al., 2008). Moreover, musical training seems to enhance auditory-motor coupling at the cortical level during rhythm processing (Chen et al., 2006; Grahn and Rowe, 2009), which is in line with evidence indicating that musicians show better rhythm synchronization than controls (Chen J. et al., 2008), likely due to a stronger internal representation of the beat or enhanced working memory abilities (Zatorre et al., 2010; Kung et al., 2011).

Connectivity studies have thus provided insights by exploring internal brain dependencies related to rhythm perception as modulated by musical training. They have, however, exclusively examined this phenomenon within the General Linear Model (GLM) approach, which allows studying brain activity as modeled by the researcher. In contrast, data-driven analyses require no explicit model of the temporal course of the brain activations, allowing for a more open, and comprehensive understanding of the brain mechanisms underlying rhythm processing. A well-studied data-driven approach is Independent Component Analysis (ICA), a blind source separation technique for studying networks on which we have no prior information. ICA is intrinsically a multivariate approach, i.e., it considers the relationships between all voxels simultaneously. Thus, it can provide an alternative and complementary approach to voxel-wise analyses. ICA can separate fMRI data into independent components (ICs), each of which represents spatially independent but functionally connected brain networks. What is special and interesting about ICA is that (a) it allows us to study connectivity and find networks without the need to rely on seed-based analysis, (b) it is a completely data-driven approach able to identify brain activity without a-priori assumptions of its dynamics; and (c) it has been applied reliably in naturalistic approaches stimuli (Bartels and Zeki, 2004, 2005; Malinen et al., 2007; Wolf et al., 2010), so complex naturalistic data can be analyzed reliably with consistent results.

In the current study, we aimed to identify the brain networks that respond to clarity of the pulse during music listening. The clarity or salience of the pulse is considered a high-level musical dimension that conveys how easily listeners can perceive the underlying metrical pulsation in a given musical piece (Lartillot et al., 2008). To study this phenomenon, we used a region-of-interest-based ICA (ROI-based ICA) approach. ROI-based ICA improves the separation and anatomical precision of the identified components that represent sources of interest, since (a) the brain volume does not affect the number of obtained components, and (b) informative signals with respect to potentially interesting sources are included in the analysis, thus excluding contributions otherwise used to separate non-interesting processes (e.g., artifacts; Formisano et al., 2004; Sohn et al., 2012; Beissner et al., 2014). To this end, we presented listeners (professional musicians and controls) with three pieces of music in randomized order while their fMRI responses were recorded. A targeted, hypothesis-driven subset of regions related to rhythm processing was included in the analysis, comprising cerebral and cerebellar areas: cortical auditory, motor and somatosensory regions of the cerebrum, cerebellar regions and, subcortically, the basal ganglia. Following a two-stage dimensionality reduction approach, ICA was applied in order to decompose participants' brain responses into spatially ICs. The ICA decomposition was performed under a range of model orders (e.g., dimensionality levels), namely, assuming different numbers of sources. The justification for this approach is that different choices of model order lead to the identification of different networks or subdivisions of networks (Abou-Elseoud et al., 2010; Kalcher et al., 2012). Component selection was based on the highest correlation coefficient between the associated temporal course and a continuous measure of the pulse clarity of the music. Additionally, GLM analyses of the data were performed as a complementary approach for comparison purposes. We expected to observe group differences in pulse clarity processing as a result of musicians' improved models of beat induction, evidenced in previous work using tapping paradigms (Drake et al., 2000; Aschersleben, 2002; Hove et al., 2007; Repp and Doggett, 2007; Krause et al., 2010; Repp, 2010). Because signal sources tend to merge into individual ICs in low models whereas they split into several subcomponents at high model orders (Abou-Elseoud et al., 2010), we additionally hypothesized that large-scale networks underpinning pulse clarity would be observed at low model orders, reflecting a scattered functional network previously reported in studies investigating rhythm processing. Accordingly, subcomponents of the large-scale networks that respond to pulse clarity would be observed at high model orders.

MATERIALS AND METHODS

Participants for the fMRI Experiment

Thirty-six healthy participants with no history of neurological or psychological disorders participated in the fMRI experiment. The participants were screened for inclusion criteria before

admission to the experiment (no ferromagnetic material in their body; no tattoo or recent permanent coloring; no pregnancy or breastfeeding; no chronic pharmacological medication; no claustrophobia). The participant pool was selected to include an equal number of professional musicians ($n = 18$, age $= 28.2 \pm 7.8$, females $= 9$) and non-musicians ($n = 18$, age $= 29.2 \pm 10.7$, females $= 10$, left-handers $= 1$). The criteria for musicianship was having more than 5 years of music training, having finished a music degree in a music academy, reporting themselves as musicians, and working professionally as a performer. As for the type of musicians, there were classical ($n = 12$), jazz ($n = 4$), and pop ($n = 2$) musicians. The instruments played were strings (violin $= 4$; cello $= 2$; double bass $= 1$), piano ($n = 8$), winds (trombone $= 1$; bassoon $= 1$), and mixed ($n = 1$). The musicians' group was homogeneous in terms of the duration of their musical training, onset age of instrument practice, and amount of years of active instrument playing. These details were obtained and crosschecked via questionnaires and HIMAB (Gold et al., 2013; Helsinki Inventory for Music and Affect Behavior). Both groups were comparable with respect to gender, age distribution, cognitive performance, socioeconomic status, and personality and mood questionnaire. The experiment was undertaken with the understanding and written consent of all participants. The study protocol proceeded upon acceptance by the ethics committee of the Coordinating Board of the Helsinki and Uusimaa Hospital District. The present dataset was part of a broad project ("Tunteet") investigating different hypotheses related to auditory processing and its dependence on person-related factors by means of a multidimensional set of paradigms and tests, involving several experimental sessions, brain and behavioral measures as well as questionnaires. The findings related to the various hypotheses investigated appear in separate papers (cf. Alluri et al., 2015, 2017; Burunat et al., 2015, 2016; Kliuchko et al., 2015).

Stimuli

The musical pieces used in the experiment were the following: (a) Stream of Consciousness by Dream Theater; (b) Adios Nonino by Astor Piazzolla; and (c) Rite of Spring (comprising the first three episodes from Part I: Introduction, Augurs of Spring, and Ritual of Abduction) by Igor Stravinsky. These are a progressive rock/metal piece, an Argentinian New Tango, and an iconic twentieth century classical work, respectively, thus covering distinct musical genres and styles. All three selected pieces are instrumental and have a duration of about 8 min. The recording details, musical excerpts used, and Spotify links to the musical stimuli can be found as Supplementary Material.

fMRI Experimental Procedure

Participants' brain responses were acquired while they listened to each of the musical stimuli in a counterbalanced order. For each participant the stimuli loudness was adjusted to a comfortable but audible level inside the scanner room (around 75 dB). In the scanner, participants' only task was to attentively listen to the music delivered via high-quality MR-compatible insert earphones while keeping their eyes open.

fMRI Scanning and Preprocessing

Scanning was performed using a 3T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) and a standard 20-channel head-neck coil, at the Advanced Magnetic Imaging (AMI) Centre (Aalto University, Espoo, Finland). Using a single-shot gradient echo planar imaging (EPI) sequence, 33 oblique slices (field of view $= 192 \times 192$ mm; 64×64 matrix; slice thickness $= 4$ mm, interslice skip $= 0$ mm; echo time $= 32$ ms; flip angle $= 75^{\circ}$) were acquired every 2 s, providing whole-brain coverage. T1-weighted structural images (176 slices; field of view $= 256 \times 256$ mm; matrix $= 256 \times 256$; slice thickness $= 1$ mm; interslice skip $= 0$ mm; pulse sequence $=$ MPRAGE) were also collected for individual coregistration. Functional MRI scans were preprocessed on a Matlab platform using SPM8 (Statistical Parametric Mapping), VBM5 for SPM (Voxel Based Morphometry; Ashburner and Friston, 2000); Wellcome Department of Imaging Neuroscience, London, UK), and customized scripts developed by the present authors. For each participant, low-resolution images were realigned on six dimensions using rigid body transformations (translation and rotation corrections did not exceed 2 mm and 2°, respectively), segmented into gray matter, white matter, and cerebrospinal fluid, and registered to the corresponding segmented high-resolution T1-weighted structural images. These were in turn normalized to the MNI (Montreal Neurological Institute; Evans et al., 1994) segmented standard a priori tissue templates using a 12-parameter affine transformation. Functional images were then blurred to best accommodate anatomical and functional variations across participants as well as to enhance the signal-to-noise by means of spatial smoothing using an 8 mm full-width-at-half-maximum Gaussian filter. Movement-related variance components in fMRI time series resulting from residual motion artifacts, assessed by the six parameters of the rigid body transformation in the realignment stage, were regressed out from each voxel time series. Next, spline interpolation was used to detrend the fMRI data, followed by temporal filtering (Gaussian smoothing with kernel width $= 4$ s).

Brain responses to the three stimuli were concatenated making a total of $\sim$24 min worth of data. The rationale behind this was to combine stimuli representing a wide range of musical genres and styles in order to cancel out effects that the specific kinds of music may have on the phenomenon under investigation. The final time series had 702 samples after the four first samples of each of the three runs were removed to avoid artifacts due to magnetization effects.

Region of Interest (ROI) Selection

Because a ROI-based ICA approach improves the separation and anatomical precision of the identified spatial components (Formisano et al., 2004; Sohn et al., 2012; Beissner et al., 2014), we only included in the analysis regions that have been identified in previous research as relevant in pulse processing. Previous studies show substantial overlap of neural substrates underlying rhythm processing, namely auditory cortices, PMC, SMA, CER, and the BG (Schubotz and von Cramon, 2001; Mayville et al., 2002; Ullén et al., 2002; Lewis et al., 2004; Grahn and Brett, 2007; Chen J. L. et al., 2008; Bengtsson et al., 2009; Grahn,

2012). Those ROIs were consequently included in the analysis with the addition of other potentially interesting areas, such as the primary motor cortex (M1), primary and secondary somatosensory cortices (S1 and S2, respectively), and rolandic operculum (ROper). The cerebellar regions included were lobules V, VI, and VIII (lobV, lobVI, lobVIII, respectively), previously associated with motor control (Penhune et al., 1998; Salmi et al., 2010; Bernard and Seidler, 2013) and rhythm processing (Grahn and Brett, 2007; Chen J. L. et al., 2008; Bengtsson et al., 2009; Grahn and McAuley, 2009). The ROI contained a total of 25,047 voxels (see **Figure 1** for a map of selected ROI).

Dual PCA Reduction

Multi-subject ICA approaches are generally used in combination with dimensionality reduction methods in order to reduce the complexity for the subsequent ICA decomposition and avoid overfitting (Calhoun et al., 2001; Beckmann and Smith, 2005). Typically, dimension reduction is applied at both the individual and group levels. Performing subject-level principal component analysis (PCA) has the computational advantage of both reducing the dimensions of the data and denoising due to projecting the data onto their principal subspace. A second PCA at the group level is necessary prior to ICA to reduce the dimension of the data to the number of desired components estimated via ICA (Erhardt et al., 2011). This is required because the high dimensionality of the data from all subjects violates the ICA assumption of the determined mixture, where the number of fMRI images (mixtures) and sources match. Let Y_i denote the preprocessed, spatially normalized T -by- V data matrix for subject i out of M subjects, where $T =$ time points (fMRI scans collected during the course of the experiment) and $V =$ voxels. Y_i is subjected to PCA, resulting in the L-by-V PCA reduced data,

$$Y_i^* = F_i^- Y_i, \tag{1}$$

where F_i^- is the L-by-T reducing matrix of L number of principal components retained per subject. L is preferably chosen as a common value for all F_i^-, $i = 1, ..., M$ rather than separately for each F_i^-. The reason for this is that once in the back-reconstruction stage of subject-specific ICs, each subject has the same number of components determined by the ICA parameters.

Accordingly, the first 80 eigenvectors were retained in the subject-level PCA ($L = 80$), which preserved ∼93% of the variance for each participant. The subject-level dimensions were thus reduced from 702 to 80 time points per participant. Following this, the PCA-reduced subject data were concatenated in the temporal domain for all 36 participants into an LM-by-V aggregate data matrix $Y^* = [Y_1^{*T}, \ldots, Y_M^{*T}]^T$, which for our dataset was 2,880-by-25,047 ($LM = 80 \times 36 = 2,880$ concatenated time dimensions). The aggregate data were further reduced in a second PCA (group PCA) prior to ICA to N, the number of components to be estimated. Thus, the N -by- V group PCA-reduced matrix X was obtained,

$$X \equiv G^- Y^* = [G_1^T, \ldots, G_M^T] \begin{bmatrix} F_1^- Y_1 \\ \vdots \\ F_M^- Y_M \end{bmatrix} \tag{2}$$

where G is the LM-by-N group PCA reducing matrix, and G^- denotes its pseudo-inverse (see **Figure 2** for the variance as a function of number of principal components retained for both subject and group-level PCAs; and **Figure 3** for overall ICA pipeline).

The Question of Optimal Model Order Selection

The selection of the optimal N or ICA model order to analyze fMRI data is not an easy problem and is still a subject of ongoing debate (Manoliu et al., 2013). This is because of the lack of a priori knowledge about the ground truth of the underlying components for brain imaging data and their modulation profiles across subjects. Several rules of thumb on an upper bound for model order estimation have been suggested for robust estimation of number of sources (ICs; Särelä and Vigário, 2003; Onton and Makeig, 2006), which suggest that model orders above a certain upper bound are expected to deteriorate ICA

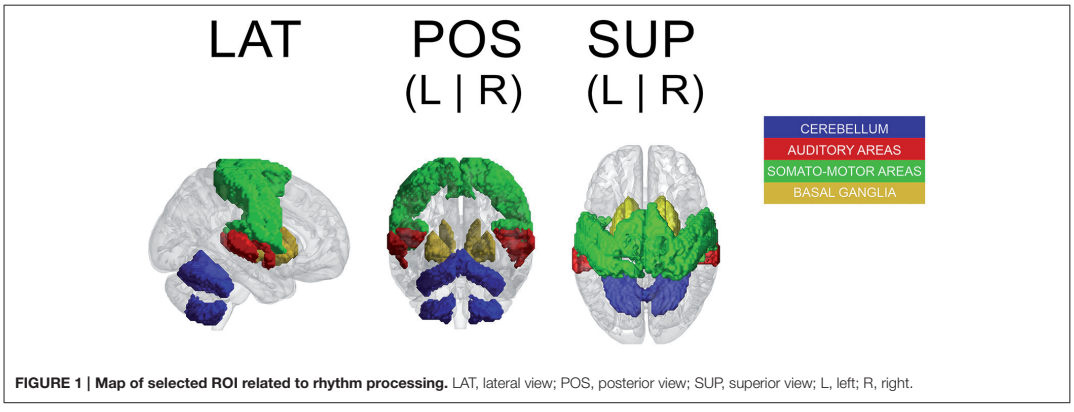

FIGURE 1 | Map of selected ROI related to rhythm processing. LAT, lateral view; POS, posterior view; SUP, superior view; L, left; R, right.

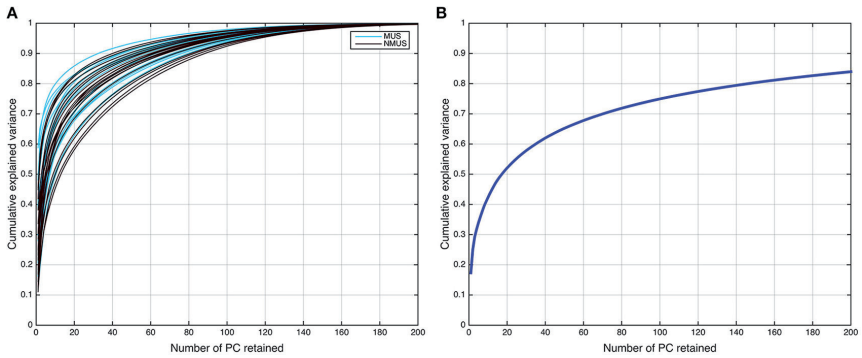

FIGURE 2 | Dual PCA reduction. (A) Cumulative variance explained as a function of principal components (PCs) retained for the subject-level and **(B)** group-level PCA.

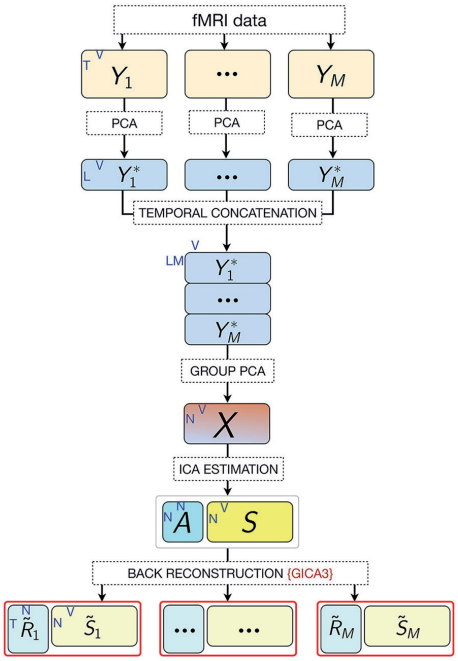

FIGURE 3 | Pipeline of the ICA approach. M, number of participants; L, size of subject-level PCA reduced time dimension; K, number of fMRI time points; V, number of voxels; N, number of estimated ICs (model order).

decomposition quality. According to the first of these rules (Särelä and Vigário, 2003), for robust estimation of N parameters (ICs) one needs $V = 5 \times N^2$, where $V =$ samples (voxels). According to this rule of thumb, the upper limit in our dataset ($V = 25,047$) corresponds to 70 ICs, explaining ∼70% of the variance. However, the second rule (Onton and Makeig, 2006) suggests that the number of data points needed to find N stable ICs from ICA is typically $V = kN^2$, where k is a multiplier with a recommended value of $k \geq 25$. Accordingly, the decomposition quality in our dataset would start to deteriorate above a model order of ∼35 ICs. It should be noted, however, that these upper limit rules do not guarantee the prevention of overfitting. Conversely, if the number of components were to be estimated based on the conventionally used 90–95% threshold of explained variance, the model order would have to be set to several 100 (see **Figure 2B**). Such a large estimate of model order will most likely lead to overfitting problems. Recent research (Abou-Elseoud et al., 2010; Allen et al., 2011) indicates that a model order around 70 components may represent an good heuristic estimate of model order to detect between-group differences and to avoid false positive results.

In order to avoid relying on prior model order assumptions given the divergent findings on model order estimation and the disparate model order selection approaches that currently exist, we aimed at decomposing our data into a varying number of assumed ICs, ranging from 10 to 100 in steps of 10, and examined the ensuing ICs derived from each decomposition.

Independent Component Analysis (ICA)

ICA in its most general, noise-free form assumes that

$$\mathrm{X} = AS, \tag{3}$$

where the measured signal $\mathrm{X} = [x_1, x_2 \ldots x_n]^T \in \mathbb{R}^n$ is a linear mixture of N statistically independent, non-normal, latent source signals $\mathrm{S} = [s_1, s_2 \ldots s_n]^T \in \mathbb{R}^n$ which are indirectly observed and called the independent components (ICs), where A, referred to as the mixing matrix, is unknown. ICA attempts to find an unmixing matrix $W \approx A^-$ to recover all source signals, such that $W\mathrm{X} = \hat{\mathrm{S}} \approx \mathrm{S}$, being the source signals optimized to be maximally

independent. The rows of $\hat{S}$ are the recovered ICs, each of which represents temporally coherent functional networks, i.e., brain regions with synchronized source signals.

According to the above noise-free model (Equation 3), in our current dataset, X denotes the N-by-V group PCA-reduced matrix with V signals (voxels), and thus there are N instances of each signal. A is an N-by-N mixing matrix and S is a N-by-V matrix containing the N independent components. The rows of S are spatially independent images, and the columns of A are spatially independent time courses associated with those images. ICs were estimated via ICASSO, a robustness analysis tool that ensures stability of the estimated components (Himberg and Hyvärinen, 2003; Himberg et al., 2004). It accomplishes this by running the same ICA algorithm several times under different random initial conditions and bootstrapping, after which it performs clustering on the obtained estimations. ICA was run 100 times using the FastICA algorithm (Hyvärinen, 1999), known to yield consistent results for fMRI data analysis (Correa et al., 2007), with a maximum number of 100 different randomly initialized unmixing matrices up to convergence. The decorrelation approach used was symmetric, i.e., the estimation of all ICs was run in parallel, with hyperbolic tangent (*tanh*) as the set non-linearity. The rest of parameters were left as their defaults specified in FastICA. ICASSO was run for each model order. This yielded a set of group ICs consisting of a spatial map reflecting the ICs' functional connectivity pattern across space and their associate temporal courses reflecting the ICs' activity across time. The spatial maps were scaled by z-scoring and thresholded at $p < 0.001$ by means of a one-sample Wilcoxon signed rank test ($N = 36$, $p < 0.001$, cluster-wise corrected, FEW = 0.05) to test their mean values against the null hypothesis of no significant difference from zero. These group-level IC maps defined relevant networks at the group level for the whole participant pool.

Subject-specific IC temporal courses ($\tilde{R}_i$) were then estimated via the back-reconstruction algorithm GICA3 (Erhardt et al., 2011), whereby the aggregate mixing matrix $A \approx W^{-}$ was back-projected to the subject space based on the PCA reducing matrices, such that

$$\tilde{R}_i = F_i(G_i^T)^{-}A, \tag{4}$$

where $\tilde{R}_i$ is the T-by-N matrix of IC temporal courses corresponding to subject i, F_i is the T-by-L PCA reducing matrix corresponding to subject i, and G_i^T is the i th subject partition of the transpose of LM-by-N group PCA reducing matrix G. GICA3 in combination with PCA reduction has been shown to produce accurate and robust results with the most intuitive interpretation in comparison to other back-projection procedures (Erhardt et al., 2011).

Identification of Pulse Clarity-Related Components

To identify the components of interest, i.e., those associated with the stimulus' pulse clarity, a model of the pulse clarity of the musical stimuli implemented in MIR Toolbox (Lartillot and Toiviainen, 2007) was used. It is based on the autocorrelation of the amplitude envelope of the audio waveform, and conveys how easily the underlying pulsation in music can be perceived by the listeners (Lartillot et al., 2008). Lartillot et al. (2008) evaluated this model of pulse clarity by means of a perceptual test where participants rated the pulse clarity of musical excerpts. Thus, it is perceptually grounded, representing the clarity of the beat as perceived by listeners, where low pulse clarity denotes that the metrical pulsation cannot be perceived easily because it is not strong or clear enough. The relevant components were then selected based on the highest correlation coefficient between the ICs' associated time courses (derived from Equation 2) and the predicted waveform of the stimulus' pulse clarity. Spearman's rank correlation coefficient was chosen as the suitable non-parametric measure of statistical dependence since the potential relationship between pulse clarity and neural time courses may be monotonic, but not necessarily linear. The significance of the correlation coefficients had to be estimated due to the intrinsic serial correlation between adjacent fMRI samples, which reduces the effective degrees of freedom in the data. These were estimated by computing the cross-correlation between the participants' IC time course and pulse clarity (Pyper and Peterman, 1998). The estimate of effective degrees of freedom was averaged across participants per model order and subsequently used to compute the significance by dividing the Fisher Z-transformed correlation coefficients by the standard error $\frac{1}{\sqrt{df-3}}$, where df represents the effective degrees of freedom. Z-transformed correlation coefficients were corrected for multiple comparisons within each model order using the false discovery rate (FDR)-criterion ($q = 0.05$). The most significant component with a significance of at least $p < 0.001$ was retained per model order (**Figure 6A** shows the 10 most significant correlations per model order). Following this, Fisher's combined probability tests (Fisher, 1950; musicians and non-musicians combined, FDR corrected, $q < 0.05$) were performed to identify, for each model order, the best predicted IC by pulse clarity for the whole participant pool (see **Figure 6B**). The identification of one pulse clarity-driven IC per model order for the combined pool of musicians and non-musicians has the advantage of enabling statistical inferences to be drawn from group comparisons. Once the best predicted ICs were identified, two extra analysis were performed: (a) Fisher's combined probability tests for each group to determine whether significance was reached for both groups separately, and (b) non-parametric t-tests performed on the individual Z-scores (two-sample Wilcoxon signed rank tests) to test for significance differences between groups.

To assess the reliability of the pulse clarity-driven ICs, their ICASSO stability indices were retrieved. ICASSO stability (quality) index (I_q; Himberg et al., 2004) is a criterion to validate the reliability and stability of ICA decomposition. It reflects the compactness and isolation of a cluster, which agglomerates similar ICs found in each ICASSO run. The I_q index scores the reliability of each extracted IC between zero and one. As the I_q approaches zero, it indicates that the IC is not reliable because its estimates from different ICA runs are not similar to each other. If it approaches one, the IC is reliably extracted, and therefore stable and robust.

IC Spatial Maps Associated with Pulse Clarity Processing

The spatial maps associated with each IC temporal course were obtained by means of a one-sample Wilcoxon signed rank test ($p < 0.001$). To account for multiple comparisons, a non-parametric cluster-wise correction approach was used, whereby participants' IC spatial maps were bootstrap resampled with replacement from the pool of back-projected IC maps within a given model order (i.e., for model order N, 36 IC spatial maps were randomly drawn from a total of 36^*N IC maps). The sample was then t-tested and thresholded (one-sample Wilcoxon signed rank test; $p < 0.001$). By running a sufficiently large number of iterations, a empirical distribution of cluster sizes was generated per model order. Bootstrap resampling within a given model order ensures not only that the spatial maps are uncorrelated, but also that the spatial autocorrelation structure is consistent among them. The maps were cluster-wise corrected using a FWE = 0.05.

GLM Analyses

For the purposes of comparing results between ICA and GLM, a voxelwise correlation analysis within the selected ROI was performed with pulse clarity separately for musicians and non-musicians to identify regions predicted significantly by it (Spearman's rho, $p < 0.001$, cluster-wise corrected, FWE = 0.05). We followed the same procedure of estimating the effective degrees of freedom explained in the previous section to correct the significance of the correlation coefficients, followed by a Fisher's combined probability test for each group ($p < 0.001$, cluster-wise corrected, FWE = 0.05).

RESULTS

IC Spatial Maps Associated with Pulse Clarity Processing

The spatial maps (one-sample Wilcoxon signed rank test, $p < 0.001$, cluster-wise corrected, FWE = 0.05) corresponding to the pulse clarity-driven ICs are shown in **Figure 4**. Their ICASSO I_q indices showed an $I_q > 0.90$, except for IC_{90}, with an $I_q = 0.72$, thus indicating that the ICs were reliably extracted. Overall, pulse-clarity networks were highly consistent across model orders in terms of polarity and topography, with auditory areas [Heschl's gyrus (HG), planum temporale (PT), and anterior and posterior superior temporal gyrus (aSTG and pSTG, respectively)] and somatomotor (M1, S1, S2, SMA, PMC, ROper) and CER areas exhibiting an inverse relationship. Generally auditory areas were positively associated with pulse clarity, whereas somatomotor areas and CER showed a negative association. Different sections of the ROper showed however both positive and negative relationships within the same IC spatial maps. The areas that were present in all ICs were the auditory cortices, ROper and S2, whereas large somatomotor areas were observed only in lower model orders. BG and CER were largely recruited only in low model order IC_{20}.

GLM Analyses

Results from the GLM analyses (Spearman's rho, $p < 0.001$, cluster-wise corrected, FWE = 0.05) yielded significant results only for non-musicians and only for auditory areas (HG, PT, pSTG; see **Figure 5**). Results overlapped with those from the ICA analyses.

Pulse Clarity Processing in Musicians and Non-musicians

Figure 6A shows the first 10 most significant correlations per model order between the temporal courses from each extracted spatial IC and pulse clarity for the whole participant pool (Fisher's combined probability test, FDR corrected, $q < 0.05$). For the purposes of group comparisons, we focused only the most significant IC driven by pulse clarity within model order that yielded a significance of at least $p < 0.01$ for the whole participant pool (in the following, IC_{10}, IC_{20},..., IC_{100}). Overall, for all ICs non-musicians' brain responses were notably better predicted by the pulse clarity of the music than musicians'. Non-musicians showed highly significant correlations at the group level for all pulse clarity-driven ICs ($p < 0.0001$) except for IC_{50} ($p = 0.06$), whereas musicians exhibited significant correlations only for the four highest model orders [IC_{70} ($p < 0.05$), IC_{80} ($p < 0.01$), IC_{90} ($p < 0.001$), and IC_{100} ($p < 0.005$); see **Figure 6B**]. Finally, the between-group comparisons (two-sample Wilcoxon signed rank tests) revealed significantly higher correlations for non-musicians compared to musicians for IC_{10} ($p < 0.01$), IC_{20} ($p < 0.005$), IC_{30} ($p < 0.005$), IC_{40} ($p < 0.005$), IC_{60} ($p < 0.05$), IC_{70} ($p < 0.05$), IC_{80} ($p < 0.05$), IC_{100} ($p = 0.05$), with IC_{90}, showing a higher non-significant trend also in favor of non-musicians.

DISCUSSION

The aim of this study was to investigate the processing of an aspect of rhythm, namely clarity of the pulse (extracted computationally), during naturalistic music listening, and potential differences in pulse clarity processing between professional musicians and non-musicians. To this end, ICA was used to decompose participants' brain responses into ICs in a targeted hypothesis-driven ROI related to rhythm processing. An advantage of using a ROI-based ICA approach is that it improves the separation and anatomical precision of the identified spatial ICs as it includes in the analysis informative signals with respect to potentially interesting sources. The ICA decomposition was performed under a range of model orders, i.e., assuming a different number of sources (ranging from 10 to 100 in steps of 10) in order to avoid relying on prior model order assumptions. A total of 10 ICs (one per model order) were selected based on the highest correlation coefficient between IC temporal courses and a continuous measure of the pulse clarity of the music, obtained from the stimulus using computational acoustic feature extraction. Additionally, the associated spatial networks across model orders were examined.

IC Spatial Maps Associated with Pulse Clarity Processing

Because non-musicians' brain responses at all model orders were overall significantly better predicted by pulse clarity than musicians' (see Section Pulse Clarity Processing in Musicians

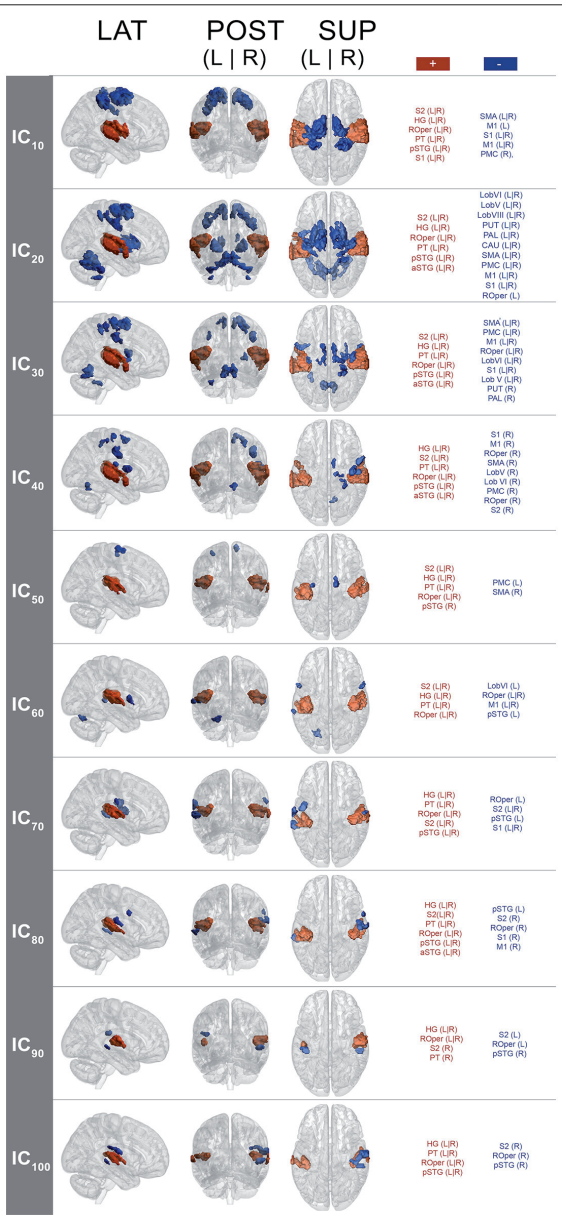

FIGURE 4 | Group-level IC spatial maps corresponding to IC temporal courses with maximal correlation with pulse clarity per model order (one-sample Wilcoxon signed rank test, $N = 36$, $p < 0.001$, cluster-wise corrected, FWE = 0.05). + (plus), positive correlation; – (minus), negative correlation; LAT, lateral view; POS, posterior view; SUP, superior view; L, left; R, right; S1, primary somatosensory cortex; S2, secondary somatosensory cortex; HG, Heschl's gyrus; PT, planum temporale; ROper, Rolandic operculum; pSTG, superior temporal gyrus (posterior); aSTG, superior temporal gyrus (anterior); PMC, premotor cortex.

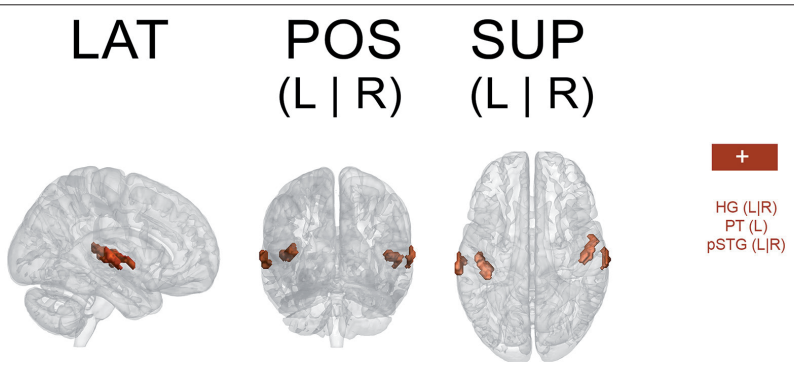

FIGURE 5 | Spatial map resulting from the GLM analysis (only non-musicians). Regions predicted by pulse clarity (Spearman's rho, $p < 0.001$, cluster wise corrected (FWE = 0.05). + (plus), positive correlation; – (minus), negative correlation; LAT, lateral view; POS, posterior view; SUP, superior view; L, left; R, right; HG, Heschl's gyrus; PT, planum temporale; pSTG, superior temporal gyrus (posterior).

and Non-musicians), the associated spatial maps here discussed reflect to a higher degree the brain networks underpinning pulse clarity processing in non-musicians than in musicians.

The spatial maps associated with pulse clarity processing during continuous, real-world music listening revealed consistent action-perception networks across model orders. This observed consistency supports the idea that they respond to the same phenomenon (i.e., pulse clarity processing). The observed networks at lower model orders comprised auditory-motor areas, while at higher model orders they recruited mainly auditory areas. One feature of the observed networks was their consistent polarity for all decompositions. Positive sign was largely found for the auditory areas, whereas negative sign was observed for somatomotor and CER areas. This polarity may be construed as an action-perception functional network during pulse clarity processing, denoting an excitatory-inhibitory relationship. A tentative interpretation of this polarity may be that when the pulse is stable and clear in the music, auditory cortices engage as motor areas disengage, and when the pulse is less clear, the major engagement of the motor system, as auditory areas disengage, could respond to the demand to organize temporally complex auditory information. Although the question of this polarity remains unresolved, it is a relevant question and remains open for further study.

Interestingly, an aspect of this polarity, namely the activation of motor areas with decreasing pulse clarity, seems to be in disagreement with previous neuroimaging results on rhythm perception, which found increased regional activity in motor areas as rhythmic saliency increased (i.e., high pulse clarity). For instance, Bengtsson et al. (2009) found cortico-motor areas to be activated when listening to metrically less complex rhythm (isochronous sequence) compared to more complex sequences (non-metric and random) during a listening task. Similarly, Grahn and Brett (2007) observed that a simple rhythmic sequence elicited increased activity in BG and SMA to a greater extent than complex or non-metric rhythmic sequences. Moreover, in their study complex metric and non-metric sequences did not statistically differ in terms of their activation in all areas, which could mean that the contrast between these two conditions was not sufficient to observe significant differences.

We argue that previous studies assume motor-related activity to be a direct linear, or at least monotonic, function of complexity. However, the extent to which pulse prediction is engaged may exhibit an inverted *U*-shape as a function of rhythmic complexity, and such a continuum may not be captured by the stimuli used in controlled experiments. This is in line with previous work on groove (Witek et al., 2014) on movement propensity vs. rhythmic complexity (Burger et al., 2013), and more generally with aesthetic experience as a function of complexity (Berlyne, 1971, 1974; Nasar, 2002; Akalin et al., 2009). Thus, previous work may lack conditions that account for different degrees of rhythmic complexity (from simple to random sequences) that allow for increasingly challenging sequences. For instance, random sequences designed to represent highly complex rhythms are unpredictable. This may explain less involvement of BG or corticomotor activation than in simpler rhythms in previous work, as no predictions are available. A condition that represents a compromise between high predictability and unpredictability could show perhaps an increased activation of the motor system in response to increasing compared to decreasing complexity.

In addition, non-temporal aspects of the musical structure (i.e., melody, harmony, timbre, pitch) can impact the perception of pulse (Temperley, 1963; Dawe et al., 1993; Parncutt, 1994; Huron and Royal, 1996). This becomes relevant in real music, as it presents higher variability in a higher number of dimensions (i.e., dynamics, timbre, harmony, melody) than simple, controlled auditory stimuli. This characteristic of real music may facilitate the prediction of subsequent beats, even in highly complex sequences. Because this multidimensionality

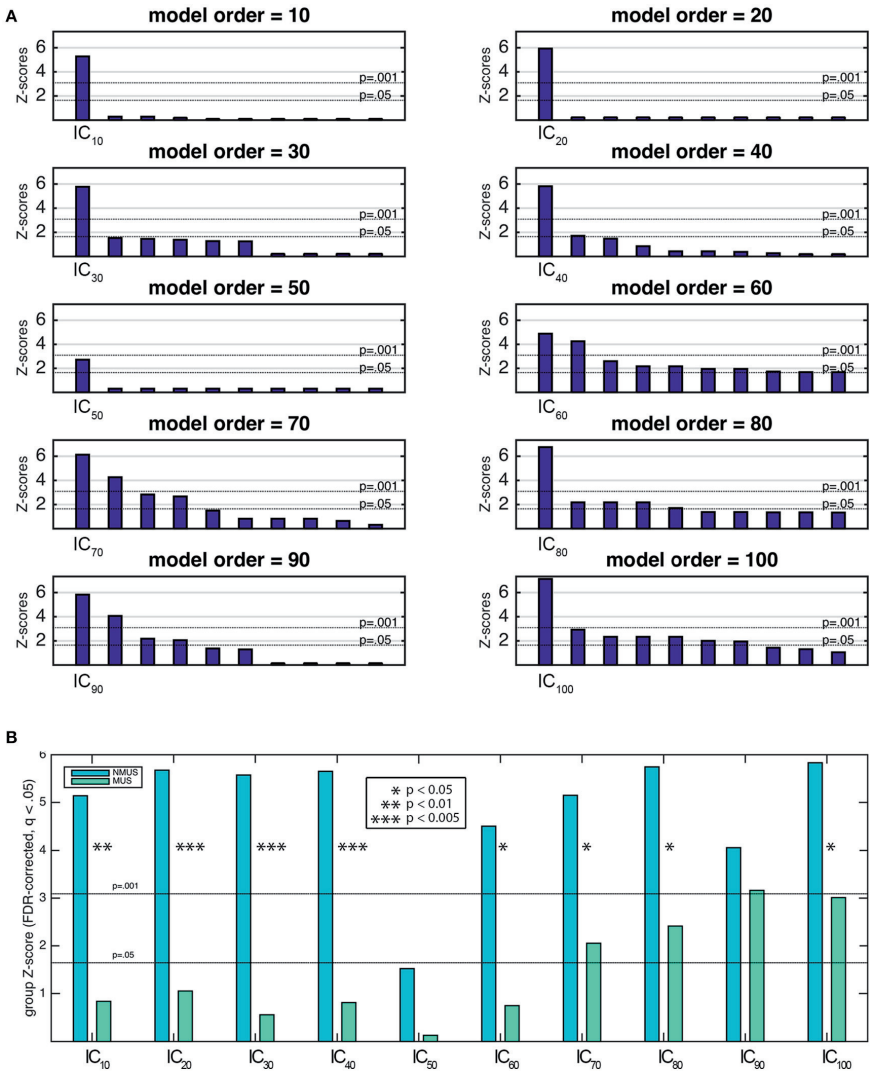

FIGURE 6 | (A) Ten most significant correlations per model order resulting from the Fisher's combined probability test for the whole participant pool (musicians and non-musicians combined; Spearman's rho, FDR corrected, $q = 0.05$, within model order). **(B)** Musicians' and non-musicians' results from the Fisher's combined probability test (Spearman's rho, FDR corrected, $q = 0.05$, within model order). Additionally, significant between-group differences are indicated as a result of a nonparametric *t*-test performed on the individual *Z*-scores (Wilcoxon signed rank test).

is missing in controlled stimuli, pulse tracking in controlled auditory conditions in may pose an additional challenge. Thus, the use of real music and a continuous measure of pulse clarity in our study may be one of the reasons for this discrepancy of results.

Furthermore, in contrast to previous work on rhythm processing based on mass-univariate analyses (GLM) targeted at findings regionally specific effects, ICA is a multivariate approach, which explicitly accounts for inter-regional dependencies. This makes multivariate inference more powerful

than mass-univariate topological inference, because it does not depend on focal responses that survive a given threshold (Friston et al., 2008). Additionally, ICA seems to have a higher sensitivity for detecting task-related changes in fMRI signal compared to the widely used mass-univariate GLM-based approach as a consequence of a stricter criterion for spatial independence between spatial maps (ICs), which reduces noise in the final solution by separating artifactual and other physiological fluctuations from the fMRI signal of interest (McKeown et al., 1998). Similarly, because GLM-based approaches cannot segregate the signal mixture from each voxel into source signals, they are not suited to detect overlaps of functional networks and their temporal course modulation by cognitive tasks. ICA methods, conversely, are capable of disentangling signal mixtures. In this regard, how functional networks overlap with different temporal courses and their modulation by cognitive tasks is critical for understanding brain functional organization (Quintana and Fuster, 1999; Fuster, 2009).

If we consider the aforementioned observations, the discrepancy of our findings in view of previous work may be reconcilable as it may result from a combination of methodological factors. However, further research is required to determine the reason for this inconsistency of results.

Disregarding the polarity issue, previous research on rhythm processing found a similar network including BS areas, PMC, SMA, and auditory cortices responding to salient rhythms, which was observed in the current study at the model orders of 20 and 30 (see **Figure 4**) among other areas (such as S1, S2, ROper, M1, and CER). Conversely, at higher model orders, networks were mainly constrained to the auditory areas with minor encroachments into the S2 and ROper. This can be explained by the fact that at low model orders, signal sources tend to merge into singular ICs, which then split into several subcomponents at higher model orders (Abou-Elseoud et al., 2010). Thus, different choices of model order lead to the identification of different networks or subdivisions of networks (Kalcher et al., 2012).

There is a lack of knowledge on the neurophysiological reasons as to why some components tend to branch into more fine-tuned components while others remain stable. It is speculated that low model orders may group larger networks which are sparsely connected (van den Heuvel et al., 2008), whereas higher order would seem to group non-branching components which are more functionally independent from each other. Thus, low model orders may provide a general picture of large-scale brain networks (Abou-Elseoud et al., 2010). This hierarchical structure of functional brain networks would be organized in a highly efficient small-world manner (Sporns and Zwi, 2004; Stam, 2004; Achard et al., 2006), i.e., with a dense neighborhood clustering sensible to local information processing and sparse, long-distance connections in order to both target and integrate global communication across the network. At very high model orders, however, ICs' repeatability is known to decline (Abou-Elseoud et al., 2010). In the current study, the high quality index (I_q) of all the pulse clarity-driven ICs guaranteed their stability and robustness, suggesting relatively good repeatability for all model orders. Moreover, between-group differences in functional connectivity measured with ICA might be affected by model order selection (Abou Elseoud et al., 2011). This was apparent in the current analyses, as group differences were more striking in low model orders (IC_{10-40}) than in higher model orders (IC_{60-100}; see **Figure 6B**). A hypothetical explanation of these results would be that, at low model orders, large-scale networks emerge which represent the functional footprint for pulse clarity processing specific of a particular population (e.g., musicians or non-musicians). Conversely, high model orders may uncover small-scale networks, which would constitute subcomponents or main functional hubs of the broader low model order networks. Accordingly, these main hubs may be conceived as more universal and hence characteristic of a wider population, common to individuals with and without professional musical training. In the current study, the auditory cortex would act as the main hub of the networks subserving pulse clarity processing, common to both musicians and non-musicians.

Given the trade-off between number of ROIs and resolution of the ICA solution, the current work was focused only on how action-perception networks sustain pulse clarity. Future work will use current findings to include additional regions, e.g., cortical areas associated with exogenous temporal expectation, so as to investigate top-down aspects of rhythm processing.

Finally, a key strength of the present approach was the inclusion of results from a continuum of model orders, rather than assuming a fixed number of sources, whereby different hierarchies are exposed in the functional brain organization of pulse clarity processing during continuous, real-world music.

Comparison with GLM

The complementary GLM analyses were only significant for non-musicians, indicating that only non-musicians' brain responses to pulse clarity fitted the pulse clarity model used in the analyses. Thus, similarly to the ICA results, GLM results could be explained by the idea that musicians possess different models for predicting the pulse of the music. Furthermore, functional brain correlates underlying pulse clarity processing from the GLM approach evidenced a positive relationship between the stimulus' pulse clarity and non-musicians' auditory cortical activity, a result consistent with ICA analyses, especially at the higher model orders (see previous Section IC Spatial Maps Associated with Pulse Clarity Processing. IC spatial maps associated with pulse clarity processing). In sum, extra GLM analyses provided a framework against which to compare and validate the reliability of the ICA findings, and thus convergence of results denoted the robustness of the ICA approach. The complementary GLM analysis served as an additional reliability check, by demonstrating the power of the ICA approach, which enabled the detection of networks undetectable through GLM.

Pulse Clarity Processing in Musicians and Non-musicians

Examination of the IC time courses across model orders indicated that non-musicians' brain activity was overall significantly better predicted by the stimulus' pulse clarity than musicians' (see **Figure 6**). Thus, the computational model of pulse clarity based on acoustical descriptors alone was insufficient in predicting the temporal evolution of activations to pulse clarity in musicians compared to non-musicians. These

results would be in line with the notion that non-musicians' internal model of pulse clarity relies on the acoustical content of the stimulus to a greater extent than musicians', whose pulse clarity model would rely more on cognitive processes and top-down rules of metricality, facilitating enhanced internal beat generation. Tapping experiments indicate an advantage in synchronization abilities for musically trained individuals as opposed to controls (Drake et al., 2000; Hove et al., 2007; Repp and Doggett, 2007; Krause et al., 2010; Repp, 2010). These experiments indicate that musicians show smaller mean negative asynchrony (MNA; the tendency for taps to precede the pacing tones) than untrained individuals (Aschersleben, 2002). Supporting this, previous evidence highlights intense, lifelong musical training as an important factor influencing beat processing, either by enabling better predictions due to a stronger internal representation of the beat, via enhanced working memory abilities (Zatorre et al., 2010; Kung et al., 2011), or by creating a richer internal model stemming from explicit knowledge of musical rules (Grahn and Rowe, 2013). In sum, significant between-group differences may be attributed to the musicians' improved accuracy to internally keep the temporal regularities in the music. As such, this is a *post-hoc* explanation of the present results which would need further support from future experiments to determine its validity.

CONCLUSION

The present study used a novel approach in the study of musical pulse processing by combining ROI-based ICA, a naturalistic auditory stimulation paradigm (free-listening to continuous real-world music), and acoustic feature extraction. The approach of relating brain responses during continuous music listening to computationally extracted acoustic features has been first applied by Alluri et al. (2012), replicated by Burunat et al. (2016) for fMRI, and by Poikonen et al. (2016a,b) for electroencephalography. Here, data decomposition at different assumed dimensionalities revealed the hierarchical organization of the functional networks subserving pulse clarity processing, hidden from GLM analyses. This networks exposed a strengthened functional action-perception network (auditory cortices, motor-related areas, BG and CER) consistent with previous neuroimaging work on rhythm processing. In addition, the fact that the associated spatial maps were spatially consistent across dimensionalities further supported the reliability of the approach. Results additionally revealed that non-musicians' internal model of pulse clarity relies on the acoustical content to a greater extent than musicians', which may be explained by musicians' improved predictive models of beat induction. These inferences are in line with evidence stressing intense musical training as a crucial factor that shapes beat processing.

AUTHOR CONTRIBUTIONS

IB and PT conceived and developed the hypotheses, designed the analysis methodology of the study, and interpreted the results. VT contributed to the methodological design and performed the PCA reduction and preliminary ICA analyses. IB preprocessed the data, wrote the manuscript, performed ICA, subject back-reconstruction, and GLM analyses. EB coordinated the data collection, contributed to the design of the stimulation paradigm, the fundraising for the Tunteeet database, the preprocessing of the data, and obtained the ethics and research permissions. All authors revised the manuscript critically and approved the final version.

ACKNOWLEDGMENTS

Authors want to thank Brigitte Bogert, Marina Kliuchko, Ben Gold, Johanna Nohrström, Taru Numminen-Kontti, Mikko Heimola, Marita Kattelus, Jyrki Mäkelä, Alessio Falco, Anja Thiede, Suvi Heikkilä for the data collection, Emilia Tuovinen for the help with the behavioral data, Jussi Numminen for checking the MRI scans, Martin Hartmann for the writing of the Supplementary Material. We thank Emily Carlson, who proofread the manuscript. We are also grateful to Mikko Sams, Peter Vuust, Mari Tervaniemi, and Asoke K. Nandi for their support on the project. The contents of this scientific paper are included in the first author's Ph.D. dissertation (Burunat, 2017), which can be accessed online here: https://jyx.jyu.fi/dspace/handle/123456789/52606. Simulations were done using taito.csc.fi HP super cluster at CSC—IT Center for Science Ltd., administered by the Ministry of Education, Science, and Culture of Finland. This work was financially supported by the Academy of Finland (project numbers 272250 and 274037), and by the Danish National Research Foundation (DNRF 117).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2017.00230/full#supplementary-material

REFERENCES

Abou Elseoud, A., Littow, H., Remes, J., Starck, T., Nikkinen, J., Nissilä, J., et al. (2011). Group-ICA model order highlights patterns of functional brain connectivity. *Front. Syst. Neurosci.* 5:37. doi: 10.3389/fnsys.2011.00037

Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., and Kiviniemi, V. (2010). The effect of model order selection in group PICA. *Hum. Brain Mapp.* 31, 1207–1216. doi: 10.1002/hbm.20929

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. T. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006

Akalin, A., Yildirim, K., Wilson, C., and Kilicoglu, O. (2009). Architecture and engineering students' evaluations of house façades: preference, complexity and impressiveness. *J. Environ. Psychol.* 29, 124–132. doi: 10.1016/j.jenvp.2008.05.005

Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5:2. doi: 10.3389/fnsys.2011.00002

Alluri, V., Brattico, E., Toivianen, P., Burunat, I., Bogert, B., Numminen, J., et al. (2015). Musical expertise modulates functional connectivity of limbic regions during continuous music listening. *Psychomusicology* 25, 443–454. doi: 10.1037/pmu0000124

Alluri, V., Toiviainen, P., Burunat, I., Kliuchko, M., Vuust, P., and Brattico, E. (2017). Connectivity patterns during music listening: evidence for action-based processing in musicians. *Hum. Brain Mapp*. doi: 10.1002/hbm.23565. [Epub ahead of print].

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage* 59, 3677–3689. doi: 10.1016/j.neuroimage.2011.11.019

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain Cogn.* 48, 66–79. doi: 10.1006/brcg.2001.1304

Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry–the methods. *Neuroimage* 11, 805–821. doi: 10.1006/nimg.2000.0582

Bartels, A., and Zeki, S. (2004). The chronoarchitecture of the human brain–natural viewing conditions reveal a time-based anatomy of the brain. *Neuroimage* 22, 419–433. doi: 10.1016/j.neuroimage.2004.01.007

Bartels, A., and Zeki, S. (2005). Brain dynamics during natural viewing conditions–a new guide for mapping connectivity *in vivo*. *Neuroimage* 24, 339–349. doi: 10.1016/j.neuroimage.2004.08.044

Beckmann, C. F., and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage* 25, 294–311. doi: 10.1016/j.neuroimage.2004.10.043

Beissner, F., Schumann, A., Brunn, F., Eisenträger, D., and Bär, K.-J. (2014). Advances in functional magnetic resonance imaging of the human brainstem. *Neuroimage* 86, 91–98. doi: 10.1016/j.neuroimage.2013.07.081

Bengtsson, S. L., Ullén, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., et al. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex* 45, 62–71. doi: 10.1016/j.cortex.2008.07.002

Berlyne, D. E. (1971). *Aesthetics and Psychobiology, Vol. 336*. New York, NY: Appleton-Century-Crofts.

Berlyne, D. E. (1974). *Studies in the New Experimental Aesthetics: Steps toward an Objective Psychology of Aesthetic Appreciation*. Washington, DC: Hemisphere.

Bernard, J. A., and Seidler, R. D. (2013). Cerebellar contributions to visuomotor adaptation and motor sequence learning: an ALE meta-analysis. *Front. Hum. Neurosci.* 7:27. doi: 10.3389/fnhum.2013.00027

Burger, B., Ahokas, R., Keipi, A., and Toiviainen, P. (2013). "Relationships between spectral flux, perceived rhythmic strength, and the propensity to move," in *Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Stock* (Sweden), 179–184.

Burunat, I. (2017). *Brain Integrative Function Driven by Musical Training during Real World Music Listening*. Ph.D. dissertation, Jyväskylä Studies in Humanities, University of Jyväskylä, Jyväskylä, Finland, 302.

Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., and Toiviainen, P. (2015). Action in perception: prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE* 10:e0138238. doi: 10.1371/journal.pone.0138238

Burunat, I., Toiviainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., et al. (2016). The reliability of continuous brain responses during naturalistic listening to music. *Neuroimage* 124, 224–231. doi: 10.1016/j.neuroimage.2015.09.005

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. doi: 10.1002/hbm.1048

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042

Chen, J., Penhune, V. B., and Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* 20, 226–239. doi: 10.1162/jocn.2008.20018

Chen, J., Zatorre, R. J., and Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage* 32, 1771–1781. doi: 10.1016/j.neuroimage.2006.04.207

Clarke, E. (1989). "Considérations sur le langage et la musique," in *La Musique Et Les Sciences Cognitives*, eds S. McAdams and I. Deliège (Liège: Pierre Mardaga), 23–44.

Cooper, G., and Meyer, L. (1960). *The Rhythmic Structure of Music*. Chicago, IL: University of Chicago Press.

Correa, N., Adali, T., and Calhoun, V. D. (2007). Performance of blind source separation algorithms for fMRI analysis using a group ICA method. *Magn. Reson. Imaging* 25, 684–694. doi: 10.1016/j.mri.2006.10.017

Dawe, L. A., Platf, J. R., and Racine, R. J. (1993). Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Attent. Percept. Psychophys.* 54, 794–807. doi: 10.3758/BF03211804

Drake, C., Penel, A., and Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Percept. Interdiscip. J.* 18, 1–23. doi: 10.2307/40285899

Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., and Calhoun, V. D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Hum. Brain Mapp.* 32, 2075–2095. doi: 10.1002/hbm.21170

Evans, A., Kamber, M., Collins, D., and MacDonald, D. (1994). "An MRI-based probabilistic atlas of neuroanatomy," in *Magnetic Resonance Scanning and Epilepsy*, ed S. D. Shorvon (New York, NY: Plenum Press), 263–274. doi: 10.1007/978-1-4615-2546-2_48

Fisher, R. A. (1950). *Statistical Methods for Research Workers, 11th Edn*. London: Oliver and Boy.

Formisano, E., Esposito, F., Di Salle, F., and Goebel, R. (2004). Cortex-based independent component analysis of fMRI time series. *Magn. Reson. Imaging* 22, 1493–1504. doi: 10.1016/j.mri.2004.10.020

Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., et al. (2008). Bayesian decoding of brain images. *Neuroimage* 39, 181–205. doi: 10.1016/j.neuroimage.2007.08.013

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072. doi: 10.1162/jocn.2009.21280

Gabrielsson, A. (1987). Once again: the theme from Mozart's piano sonata in A major (K. 331). *Action Percept. Rhythm Music* 55, 81–103.

Gold, B. P., Frank, M. J., Bogert, B., and Brattico, E. (2013). Pleasurable music affects reinforcement learning according to the listener. *Front. Psychol.* 4:541. doi: 10.3389/fpsyg.2013.00541

Grahn, J. A. (2009). The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. *Ann. N.Y. Acad. Sci.* 1169, 35–45. doi: 10.1111/j.1749-6632.2009.04553.x

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* 4, 585–606. doi: 10.1111/j.1756-8765.2012.01213.x

Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906. doi: 10.1162/jocn.2007.19.5.893

Grahn, J. A., and McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *Neuroimage* 47, 1894–1903. doi: 10.1016/j.neuroimage.2009.04.039

Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009

Grahn, J. A., and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb. Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083

Himberg, J., and Hyvärinen, A. (2003). "Icasso: software for investigating the reliability of ICA estimates by clustering and visualization," *IEEE XIII Work. Neural Networks for Signal Processing (IEEE Cat. No. 03TH8718)* (Toulouse), 259–268. doi: 10.1109/NNSP.2003.1318025

Himberg, J., Hyvärinen, A., and Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage* 22, 1214–1222. doi: 10.1016/j.neuroimage.2004.03.027

Hove, M. J., Keller, P. E., and Krumhansl, C. L. (2007). Sensorimotor synchronization with chords containing tone-onset asynchronies. *Percept. Psychophys.* 69, 699–708. doi: 10.3758/BF03193772

Huron, D., and Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Percept. Interdiscip. J.* 13, 489–516. doi: 10.2307/40285700

Hyvärinen, A. (1999). Fast and robust fixed-point algorithm for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634. doi: 10.1109/72.761722

Kalcher, K., Huf, W., Boubela, R. N., Filzmoser, P., Pezawas, L., Biswal, B., et al. (2012). Fully exploratory network independent component analysis

of the 1000 functional connectomes database. *Front. Hum. Neurosci.* 6:301. doi: 10.3389/fnhum.2012.00301

Kliuchko, M., Heinonen-Guzejev, M., Monacis, L., Gold, B. P., Heikkilä, K. V., Spinosa, V., et al. (2015). The association of noise sensitivity with music listening, training, and aptitude. *Noise Heal.* 17:350. doi: 10.4103/1463-1741.165065

Krause, V., Pollok, B., and Schnitzler, A. (2010). Perception in action: the impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychol.* 133, 28–37. doi: 10.1016/j.actpsy.2009.08.003

Kung, S.-J., Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* 25, 401–420. doi: 10.1162/jocn_a_00325

Kung, S.-J., Tzeng, O. J. L., Hung, D. L., and Wu, D. H. (2011). Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. *Exp. Brain Res.* 210, 269–282. doi: 10.1007/s00221-011-2630-2

Large, E. W., and Palmer, C. (2002). Perceiving temporal regularity in music. *Cogn. Sci.* 26, 1–37. doi: 10.1207/s15516709cog2601_1

Large, E. W., and Snyder, J. S. (2009). Pulse and meter as neural resonance. *Ann. N.Y. Acad. Sci.* 1169, 46–57. doi: 10.1111/j.1749-6632.2009.04550.x

Lartillot, O., and Toiviainen, P. (2007). "A Matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects* (Bordeaux).

Lartillot, O., Eerola, T., Toiviainen, P., and Fornari, J. (2008). "Multi-feature modeling of pulse clarity: design, validation and optimization," in *ISMIR (Citeseer)* (Sapporo), 521–526.

Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., and Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia* 42, 1301–1312. doi: 10.1016/j.neuropsychologia.2004.03.001

Malinen, S., Hlushchuk, Y., and Hari, R. (2007). Towards natural stimulation in fMRI—issues of data analysis. *Neuroimage* 35, 131–139. doi: 10.1016/j.neuroimage.2006.11.015

Manoliu, A., Meng, C., Brandl, F., Doll, A., Tahmasian, M., Scherr, M., et al. (2013). Insular dysfunction within the salience network is associated with severity of symptoms and aberrant inter-network connectivity in major depressive disorder. *Front. Hum. Neurosci.* 7:930. doi: 10.3389/fnhum.2013.00930

Mayville, J. M., Jantzen, K. J., Fuchs, A., Steinberg, F. L., and Kelso, J. A. (2002). Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI. *Hum. Brain Mapp.* 17, 214–229. doi: 10.1002/hbm.10065

McKeown, M. J., Jung, T. P., Makeig, S., Brown, G., Kindermann, S. S., Lee, T. W., et al. (1998). Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc. Natl. Acad. Sci. U.S.A.* 95, 803–810. doi: 10.1073/pnas.95.3.803

Nasar, J. L. (2002). What design for a presidential library? Complexity, typicality, order, and historical significance. *Empir. Stud. Arts* 20, 83–99. doi: 10.2190/286Y-5VLW-G05W-RAQG

Onton, J., and Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Prog. Brain Res.* 159, 99–120. doi: 10.1016/S0079-6123(06)59007-7

Palmer, C. (1989). Mapping musical thought to musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 331–346. doi: 10.1037/0096-1523.15.2.331

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* 11, 409–464. doi: 10.2307/40285633

Patel, A. D. (2014). The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol.* 12:e1001821. doi: 10.1371/journal.pbio.1001821

Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057

Penhune, V. B., Zatorre, R. J., and Evans, A. C. (1998). Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* 10, 752–765. doi: 10.1162/089892998563149

Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M., and Huotilainen, M. (2016a). Event-related brain responses while listening to entire pieces of music. *Neuroscience* 312, 58–73. doi: 10.1016/j.neuroscience.2015.10.061

Poikonen, H., Toiviainen, P., and Tervaniemi, M. (2016b). Early auditory processing in musicians and dancers during a contemporary dance piece. *Nat. Publ. Gr.* 35, 1–11. doi: 10.1038/srep33056

Pyper, B. J., and Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Can. J. Fish. Aquat. Sci.* 55, 2127–2140.

Quintana, J., and Fuster, J. M. (1999). From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex* 9, 213–221. doi: 10.1093/cercor/9.3.213

Repp, B. (1990). Patterns of expressive timing in performances of a Beethoven Minuet by nineteen famous pianists. *J. Acoust. Soc. Am.* 93, 622–641. doi: 10.1121/1.399766

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Hum. Mov. Sci.* 29, 200–213. doi: 10.1016/j.humov.2009.08.002

Repp, B. H., and Doggett, R. (2007). Tapping to a very slow beat: a comparison of musicians and nonmusicians. *Music Percept. Interdiscip. J.* 24, 367–376. doi: 10.1525/mp.2007.24.4.367

Repp, B. H., and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2

Salmi, J., Pallesen, K. J., Neuvonen, T., Brattico, E., Korvenoja, A., Salonen, O., et al. (2010). Cognitive and motor loops of the human cerebro-cerebellar system. *J. Cogn. Neurosci.* 22, 2663–2676. doi: 10.1162/jocn.2009.21382

Särelä, J., and Vigário, R. (2003). Overlearning in marginal distribution-based ICA: analysis and solutions. *J. Mach. Learn. Res.* 4, 1447–1469.

Schubotz, R. I., and von Cramon, D. Y. (2001). Interval and ordinal properties of sequences are associated with distinct premotor areas. *Cereb. Cortex* 11, 210–222. doi: 10.1093/cercor/11.3.210

Sohn, W. S., Yoo, K., and Jeong, Y. (2012). Independent component analysis of localized resting-state functional magnetic resonance imaging reveals specific motor subnetworks. *Brain Connect.* 2, 218–224. doi: 10.1089/brain.2012.0079

Sporns, O., and Zwi, J. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145–162. doi: 10.1385/NI:2:2:145

Stam, C. J. (2004). Functional connectivity patterns of human magnetoencephalographic recordings: a small-world network? *Neurosci. Lett.* 355, 25–28. doi: 10.1016/j.neulet.2003.10.063

Temperley, N. M. (1963). Personal tempo and subjective accentuation. *J. Gen. Psychol.* 68, 267–287. doi: 10.1080/00221309.1963.9920534

Ullén, F., Forssberg, H., and Henrik Ehrsson, H. (2002). Neural networks for the coordination of the hands in time. *J. Neurophysiol.* 89, 1126–1135. doi: 10.1152/jn.00775.2002

van den Heuvel, M. P., Stam, C. J., Boersma, M., and Hulshoff Pol, H. E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528–539. doi: 10.1016/j.neuroimage.2008.08.010

Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE* 9:e94446. doi: 10.1371/journal.pone.0094446

Wolf, I., Dziobek, I., and Heekeren, H. R. (2010). Neural correlates of social cognition in naturalistic settings: a model-free analysis approach. *Neuroimage* 49, 894–904. doi: 10.1016/j.neuroimage.2009.08.060

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152

Zatorre, R. J., Halpern, A. R., and Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *J. Cogn. Neurosci.* 22, 775–789. doi: 10.1162/jocn.2009.21239

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107:5768. doi: 10.1073/pnas.1000121107

Conflict of Interest Statement: EB is currently an associate Editor of Frontiers. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PV

ON APPLICATION OF KERNEL PRINCIPAL COMPONENT ANALYSIS FOR GENERATING STIMULUS FEATURES FOR FMRI DURING CONTINUOUS MUSIC LISTENING

by

Valeri Tsatsishvili, Iballa Burunat, Frngyu Cong, Petri Toiviainen, Vinoo Alluri, Tapani Ristaniemi, 2017

Submitted to Journal of Neuroscience Methods, Elsevier

On application of Kernel PCA for Generating Stimulus Features for fMRI during Continuous Music Listening

Valeri Tsatsishvili, Iballa Burunat, Fengyu Cong, Petri Toiviainen, Vinoo Alluri, Tapani Ristaniemi

Abstract

Recently there has been an increased interest towards the brain imaging studies with high ecological validity, which deepen our understanding of how human brain processes and integrates incoming streams of multifaceted sensory information as commonly happens in real world. Music is a good example of such common stimulus having complex continuous nature. Several functional magnetic resonance imaging (fMRI) studies attempted to study neural correlates of music in a task-free, continuous music listening settings. Multiple perceptual attributes of music stimulus were represented by a set of high-level features, produced as the linear combination of the low-level descriptors computationally extracted from the stimulus audio. The present study exploited an advantage of kernel principal component analysis (KPCA) in capturing nonlinear patterns for generating a new set of nonlinear stimulus features. Results revealed interesting perceptual and neural correlates of the nonlinear features, some of which were hidden from the linear features.

1. Introduction

Music, as an important part of human experience, has long been of interest for neuroimaging community. Traditionally, temporal structure of stimulus presentation has been controlled and stimuli have been simplified in fMRI experiments, in order to maintain control over independent variables. Such experimental design limits replicating experience of music and other multifaceted phenomena as they occur in real life and therefore, do not provide an optimal framework for studying sensory or cognitive processing mechanisms of continuous information flow. Consequently, interest towards naturalistic neuroimaging experiments, where participants are exposed to continuous complex stimuli (e.g. see [1]-[3]), is growing. Music possesses several inter-related perceptual attributes including loudness, timbre, tonality, and rhythm. Many excellent works have examined neural correlates of each attribute in isolation to understand neural processing of specific musical features [4]-[6]. However, musical dimensions are rarely processed independently from each other and growing evidence suggests that the brain areas involved in music processing is not a simple integration over the networks associated to its isolated attributes [7], [8]. Therefore, segregation of neural correlates of the musical features should be achieved in natural music listening contexts in order to observe simultaneous processing of multiple features as it typically takes place in real life [8].

Functional MRI studies investigating activations in naturalistic music listening settings are scarce. Nevertheless, few attempts have been made recently. Abrams et al. [7] examined regions involved in music processing by comparing of synchronization between non-musician participants' brain responses while listening to the recordings of classical music (musical stimuli) as well as their phase scrambled and spectrally rotated versions (pseudo-musical stimuli). Results revealed synchrony across subjects in large network of cortical and subcortical regions was significantly greater in response to musical compared to pseudo-musical stimuli. With different objectives, Alluri and colleagues in [8] studied neural correlates of multiple musical features in an experiment where 11 musicians listened to a piece of modern tango in the fMRI scanner. A set of 25 descriptors were computationally extracted

from the music stimulus. The descriptors were subsequently summarized in nine principal components, or stimulus features, capturing high-level timbral, tonal, and rhythmic aspects of music. Perceptual relevance of the components was validated in the perceptual experiment. Neural correlates of the stimulus features revealed large-scale consistent foci of activations in cortical and sub-cortical areas. Analogous stimulus feature generation scheme was employed in several subsequent studies with similar experimental setup [9]-[11].

Interestingly, in [8] several brain areas showing high inter-subject consistency that failed to correlate with any of the stimulus features were also reported. Authors suggested an incomplete representation of stimulus as one possible explanation. In addition, two of the nine stimulus features had to be excluded in early stages of the analysis as they failed to significantly correlate with perceptual ratings from the validation experiment. These limitations might, among other possible reasons, stem from the existence of nonlinear patterns between the initially extracted 25 descriptors, which cannot be captured by linear PCA. In other domains, exploiting nonlinear patterns using KPCA-based features has shown to improve classification accuracy when compared to the performance of linear features [12]-[14]. This motivated us to explore application of nonlinear kernel PCA for generating perceptually relevant high-level features. This study employed the data from [8] (in the following will be referred as original study) to generate alternative stimulus representation. Perceptual and neural correlates of the new features were examined, and the results were also compared to the findings in the original study.

Section 2 details methodology for generating nonlinear stimulus features. Section 3 discusses perceptual correlates of the generated features. Section 4 presents neural correlates of and compares the spatial activation maps with findings from the original study. Section 5 highlights findings and limitation of the present study.

2. Generation of stimulus features

2.1. Acoustic feature extraction and preprocessing

Overall, 25 descriptors representing timbral, tonal, and rhythmic information were extracted from the overlapping windows. Window length was varied to facilitate different time scales of processing needed for different musical features. The shorter window length of 25ms with 50% overlap was selected for 20 low-level descriptors capturing polyphonic timbre of music. The longer windows spanning three seconds with 67% overlap were employed for the remaining more complex features capturing higher-level concepts, such as tonality and rhythm. These descriptors were adopted from the original study [8] where the complete list and descriptions are provided. All 25 features in the initial set were centered and normalized with respect to their standard deviation, and downsampled to 0.5 Hz to match fMRI sampling rate. After the initial preprocessing, the features were subjected to KPCA. The generated kernel features were finally convolved with the canonical hemodynamic response function.

2.2. Kernel PCA

Kernel PCA is an extension of the linear PCA for nonlinear data distributions where mapping into linear subspace is not useful [15], [16]. The method has gained popularity as a tool for finding useful image representations with wide variety of applications such as edge detection, handwriting recognition, and classification of various medical images.

The most straightforward way to do nonlinear extension of PCA is to introduce a nonlinear mapping to the (generally) higher dimensional feature space: $X \to \phi\ x$, calculate covariance $V = \phi\ x\ \phi\ x\ ^T$ in the feature space and then solve the following eigenvalue problem:

$$Cv = \lambda v \tag{1}$$

Such mapping can become very expensive or infeasible as the dimensionality of X increases. However, in certain cases it is possible to bypass the mapping step and directly calculate the dot product in the feature space. This is achieved by introducing a kernel function that replaces dot product in the feature space:

$$K(x,x) = E\{\phi\ x\ \phi\ x\ ^T\} = \frac{1}{m}\sum_{i=1}^{m} \phi\ x_i\ \phi\ x_i\ ^T \quad (2)$$

This procedure, called 'Kernel trick' reduces mapping and dot product operations to kernel function estimation and can be applied as nonlinear extension of any linear method that depends exclusively dot products. Once kernel function is estimated, inserting (2) in (1) reformulates the eigenvalue problem:

$$\frac{1}{m}\sum_{i=1}^{m} \phi\ x_i\ \phi\ x_i\ ^T v = \lambda v \quad (3)$$

By considering that all solutions of v in (3) lie in the span of $\phi\ x\ $:

$$v = \sum_{i=1}^{m} a_i \phi\ x_i \quad (4)$$

and substituting (4) into (3), the eigenvalue problem can be expressed as:

$$Ka = \lambda a \quad (5)$$

Finally, principal components can be obtained by projecting the sample onto the eigenvectors of the covariance matrix in the feature space:

$$\langle v, \phi\ x\ \rangle = \sum_{i=1}^{m} a_i\, K(x_i, x) \quad (6)$$

Many nonlinear functions are suitable as the kernel functions. The most common, Gaussian kernel was selected here:

$$k_{ij} = \exp(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\varepsilon^2}) \quad (7)$$

where k_{ij} are the elements of the kernel matrix, and $\boldsymbol{x}_i, \boldsymbol{x}_j$ are d dimensional instances of X. The ε parameter was set as median of the minimum value of the distances between the data points i.e. $\varepsilon = \underset{i}{\mathrm{median}} \min_{i \neq j} \left|\boldsymbol{x}_i - \boldsymbol{x}_j\right|$.

Consequently, 10 kernel PCA features were selected for further analysis, retaining about 90% of the variance in the data. More thorough derivation of kernel PCA can be found in e.g. [14].

3. Perceptual correlates of KPCA features

Perceptual interpretation of the generated kernel PCA features is not straightforward. As mentioned above, in the original study Alluri et al. applied PCA to the 25 initially extracted descriptors. The relevant musical percepts corresponding to each produced principal component (PC) was estimated observing principal component loadings, i.e. contribution of each of the 25 initial features to the given PC. Consequently, the nine PC-s were associated with the following musical percepts: Fullness, Brightness, Timbral Complexity, Activity, Rhythmic Complexity, Event Synchronicity, Pulse Clarity Key Clarity, and Mode. First four can be loosely defined as four dimensions of polyphonic timbre, following three represent different aspects of musical rhythm, and the remaining two relates to the tonality of music. The applied perceptual labels except Mode (excluded early from the analyses) were validated in the subsequent listening experiment, where 21 musicians rated short musical excerpts representing varying levels of each percept on scales from 1 to 9. To select the musical excerpts, entire stimulus piece was divided by six second moving windows with one-second step size, producing over 470 windows. Next, average PC values from all windows were extracted, sorted and 30 values were equidistantly sampled from the entire range. Perceptual ratings of Rhythmic Complexity and Event Synchronicity failed to correlate corresponding principal components and excluded, leaving six validated features that were used in subsequent analyses of the original study.

Unlike PCA, KPCA operates in higher dimensional feature space without explicitly mapping the input data in it. Thus, interpreting nonlinear acoustic components by observing the weights of the initial 25 descriptors in the resulting kernel PC-s is not applicable here. Following a more heuristic approach, we employed the 9 sets of 30 short stimulus excerpts and human ratings from the listening experiment conducted in the original study. To this end, average values of the 10 kernel PCs were extracted from the musical excerpts and correlated to the corresponding perceptual ratings. Statistical significance of correlations was estimated by permutation tests: for each feature, average feature values from randomly sampled 30 windows were correlated with randomly selected rating multiple (100 000) times. The threshold of significance for correlation coefficients was derived from the resulting distribution.

In addition to the significance of correlations, we required for a given set of 30 samples to capture majority (at least 80%) of the entire range of feature values. Otherwise, correlation (even significant) was rejected. This constraint was introduced to be concordant with the sampling criterion originally applied to the linear features. As mentioned above, Alluri et al. [8] selected stimulus excerpts such that the entire dynamic range of the PCA features was equidistantly sampled, and consequently participants of the experiment rated the excerpts spanning all levels of the given feature value. The same set of audio excerpts represent arbitrary sampling for KPCA features, which in turn has two limitations if directly correlated with human ratings: 1) it increases chance of spurious correlation and 2) by possibly capturing a fraction of the feature range, it violates existing systematic relationship between the ratings and corresponding acoustic components.

Table 1 contains correlations between the nonlinear features extracted from the audio excerpts and corresponding perceptual ratings. As evident from the table, majority of KPCA features correlated significantly with the ratings of at least one musical percept. However, the constraint on spanned dynamic range was not satisfied in all cases and accordingly, K_4, K_5, and K_{10} were eliminated from further analysis. From the remaining features, K_1 seem to capture all four dimensions of polyphonic timbre and Pulse Clarity, K_2 correlated with Brightness and Event Synchronicity, K_3 and K_6 were associated with the perceived Pulse Clarity and Key Clarity respectively, whereas K_9 was correlated to Brightness and Rhythmic Complexity. Finally, no significant correlations were observed for K_7 and K_8, which might indicate that they relate to the perceptual attributes that are not among the eight ones available.

Strong association of the kernel features with the Event Synchronicity and Rhythmic Complexity ratings is very interesting finding, since these were among the musical percepts that did not validate in the perceptual experiment in the original study (see Table 2 in [8]). Neural correlates of these 5 features are introduced in the following section.

Table 1. Correlations between KPCA and perceptual ratings. Significance thresholds r=±0.37 (p<0.05); r=±0.56 (p<0.01). Double asterisk indicates cases where correlation is significant and the additional constraint holds. Single asterisk shows only significant correlation for which the constraint was not satisfied.

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}
Fullness	0.78**	-0.23	0.33	0.21	-0.14	-0.27	0.09	-0.2	-0.09	0.13
Brightness	0.39**	0.48**	-0.02	0.1	-0.47*	0.02	-0.03	0.33	0.39**	-0.17
T. Complexity	0.73**	0.08	0.22	0.33	-0.05	-0.07	-0.02	-0.22	0.11	-0.28
R. Complexity	-0.1	0.32	0.12	0.1	-0.28	-0.35	-0.33	0.31	0.38**	-0.03
Key Clarity	0.31	0.04	0.2	-0.17	-0.1	-0.53**	0.29	-0.15	-0.22	-0.21
Pulse Clarity	0.74**	-0.14	-0.38**	0.27	0.28	0.03	0.01	-0.04	-0.2	0.29
E. Synchronicity	0.16	-0.65**	0.21	0.13	0.38*	0.21	0.11	-0.32	-0.32	-0.38*
Activity	0.8**	-0.11	0.23	0.44*	-0.01	-0.1	-0.08	0.08	0.28	-0.3

4. Finding stimulus related brain activations

4.1. Data Analysis

The fMRI data were collected, preprocessed, and analyzed in the original study and comprises brain images of eleven healthy musicians (mean age: 23.2; SD: 3.7; 5 females), elicited during listening to a piece of modern Argentinian tango without any specific task. Since the process of finding stimulus-driven activation maps closely followed to the original study, we only briefly introduce it here. Activation maps were first obtained per subject and per stimulus feature by correlating fMRI voxel time series to the KPCA features. The subject-specific maps were then pooled to group maps (continuous valued maps) such that one functional map was associated to each KPCA feature. The group maps were thresholded at $p<0.001$ to obtain binary maps of activations (positive significant correlations) and deactivations (negative significant correlations). The activation maps of the selected KPCA features are provided in Figure 2. The activation maps driven by the KPCA features (K-maps) were also compared with the PCA-driven maps (PC -maps) reported in the original study. Similarity between the spatial maps was evaluated at global and individual level. At global level, a thresholded individual K-maps and PC-maps were combined, and the overlap was estimated by:

$$O_{kp} = \frac{K \cap P}{|K|} \tag{8}$$

where K and P represent the voxels showing above-threshold activity in K-maps and PC-maps respectively, $|K|$ is the number of active voxels in K-map. This provided an overview to the shared and exclusive regions in above-threshold activations. For the individual analysis, spatial correlation was calculated between the continuous valued (unthresholded) maps. Significance of correlations was assessed by permutation tests. The procedure was similar to the one implemented in [11]: a randomly selected and phase scrambled pair of features from the two sets were voxel-wise correlated to the fMRI responses from all participants. The individual maps were subsequently averaged across participants, producing one group map per feature, and finally correlation between the pair of group maps was calculated. The process was repeated about 11 000 times and the significance threshold was estimated from the resulting distribution.

4.2. Results

Similarity at the global level is summarized in Table 2, where the sizes of combined binary maps and the overlap is provided. The size of the spatial map is defined by the number of voxels above the threshold of activation (in practice simply the number of ones in the thresholded binary map). Therefore, as the table shows, 76% of 9863 active voxels in the combined K -map are shared with the combined PC-map. Figure 1 depicts the shared and exclusive anatomical regions of the two combined activation maps. As can be seen in the figure, auditory areas occupied the largest portion of the shared areas. Elsewhere, overlapping clusters were observed in cerebellum, left precentral and bilateral postcentral gyri, left precuneus, bilateral inferior parietal gyrus, and right putamen. The unique areas associated with KPCA features (denoted by blue color) were mostly concentrated in frontal lobe, including bilateral superior frontal gyrus (SFG), middle frontal gyrus (MFG), inferior frontal gyrus (IFG), right precentral gyrus, as well as in cerebellum (crus I and II, vermis III, VII and VIII). In addition, activations in visual processing-related areas (right lingual gyrus and left middle and superior occipital gyri), bilateral anterior cingulate and paracingulate gyrus, and right paracentral lobule were also present.

Figure 2 depicts activation maps corresponding to the five KPCA features that were significantly correlated to the perceptual ratings, whereas correlations between K-maps and PC-maps are provided in Table 3. As mentioned in Section 3, features K_2 and K_9 showed significant correlations with event synchronicity and rhythmic complexity ratings. These musical percepts were not represented in PCA features in the original study as corresponding principal components failed correlate significantly with the perceptual ratings. Interestingly, both features also predicted perceived brightness of overall timbre. Activation pattern of K_2 map consists of activations in auditory cortices (in the left hemispheric prominently), visual-processing related areas (precuneus) and some cerebellar

areas. Left SMA and the right IFG tend to deactivate in response to this feature. As shown in Table 3, no significant correlations were observed between K_2-map and any of the PC-maps, although thresholded map shared predominantly auditory areas with timbral maps. In addition, K_2 map featured regions in right medial inferior orbitofrontal gyrus (negative correlation) and the left middle and superior occipital gyri (positive correlation), not reported in the original study. Similarly, thresholded K_9-map featured some overlap with timbral maps in auditory regions. Bilateral STG, MTG, right HG, and right superior temporal pole were observed to activate in response to K_9, while small areas in left inferior parietal supramarginal and angular gyri, right postcentral gyrus, and right supramarginal gyrus showed deactivation.

From the remaining features, the largest activation map was found to be involved in processing of K_1. Large scale activations included areas bilateral STG, Heschl's gyrus (HG), rolandic operculum (RO), supramarginal gyrus, superior temporal pole, insula, MTG as well as some cerebellar regions. These areas were associated to processing of polyphonic timbre in the original study. Indeed, Table 3 reveals high similarity of with the activation maps associated to timbral features and Pulse clarity. Negatively correlated regions to K_1 were located mainly in frontal and parietal lobes. We also observed few active clusters in occipital gyrus and orbitofrontal cortex that were not reported in the previous study.

Functional map of K_3 consisted of only a few hundred voxels, predominantly showing auditory area (bilateral STG and left MTG) and small cluster in postcentral gyrus.

Finally, feature K_6 was negatively correlated with perceived Key Clarity. Increased activation in clusters located in the left precuneus, left angular gyrus, left supramarginal gyrus and left lateralized deactivations in cerebellum, inferior frontal gyrus, median cingulate and paracingulate gyrus were observed in response to this feature. In the right hemisphere, deactivations in only two small clusters in right insula and right putamen were found.

5. Discussion

The present study explored application of kernel PCA for finding new perceptually meaningful auditory stimulus features for naturalistic fMRI experiment. An fMRI data from the music listening experiment adopted from the previously published study [8] and the same the methodological design has been employed for finding neural correlates of the new stimulus features, which also enabled comparison of results. However, it should be noted that since the stimulus features in the two studies were different, replication of the previously reported spatial maps was not expected.

First, perceptual correlates of the stimulus features were examined. We found significant correlations between the majority of kernel PC-s and the perceptual ratings. More interestingly, the ratings of Event Synchronicity and Rhythmic Complexity were explained better by the kernel PC-s than by the linear PC-s. This finding indicates that the KPCA features are capable of capturing high-level musical percepts hidden from the linear PCA counterparts. Interestingly, it was also observed that KPCA features were significantly correlated with multiple perceptual ratings (see Table 1). While this can be beneficial in some cases, it might make features difficult to interpret in others. For example, as observed above K_1 feature was simultaneously associated to perceived timbre and rhythm, which is not straightforward to explain. Nonetheless, existence of a latent link between perceived clarity of the main beat and color of the timbre is possible. Intriguingly, multiple associations were also found when linear acoustic components were correlated to all perceptual ratings. For example, in the original study acoustic component labeled as Fullness was reported to be significantly correlated with Fullness rating ($r = 0.80; p < 0.001$). We found that Fullness can as successfully predict Timbral Complexity, Pulse Clarity and Activity ratings ($r = 0.68,\ p < 0.001;\ r = 0.74,\ p < 0.001; r = 0.79,\ p < 0.001$). Lack of data is the most likely reason. As mentioned in Section 3, in the perceptual experiment only 30 short audio samples were rated per scale. For the stimulus lasting over seven minutes, 30 samples might capture large-scale dynamics of acoustic features but will miss more fine

details. This coarse representation is then further smoothed when the ratings are averaged across participants. Consequently, the ratings of different musical percepts will be more likely to exhibit very similar dynamics capturing only the most significant changes in music stimulus. Therefore, while served well for validation of already applied labels in the original study, the available behavioral rating data are probably not optimal to use for our purposes. The solution would be to collect more ratings in a new perceptual experiment, where the new sets of representative music excerpts are selected by equidistantly sampling KPCA features. Furthermore, the obtained excerpts should be rated on more perceptual scales than was available here, and more rigorous testing on inter-subject agreement should be performed before averaging the ratings. Alternatively, averaging the ratings can be avoided and instead correlation between a given feature and individual ratings can be averaged.

Table 2. Overlap of global thresholded KPCA map with the global thresholded PCA map

		Global map size (voxels)	Overlap
KPCA		9863	0.76
PCA		11468	

Table 3. Correlations between continuous K-maps and PC-maps (thr r=0.47, p<0.05, r=0.63, p<0.01)

	K_1	K_2	K_3	K_6	K_9
Fullness	0.98**	-0.19	0.2	-0.09	0.42
Brightness	0.9**	0.41	0.28	-0.12	0.46
T. Complexity	-0.69**	-0.24	-0.8**	-0.16	-0.67**
Key clarity	-0.04	-0.03	0.02	-0.41	-0.17
Pulse clarity	0.68**	-0.16	0.27	0.45	0.33
Activity	0.99**	-0.13	0.12	-0.12	0.44

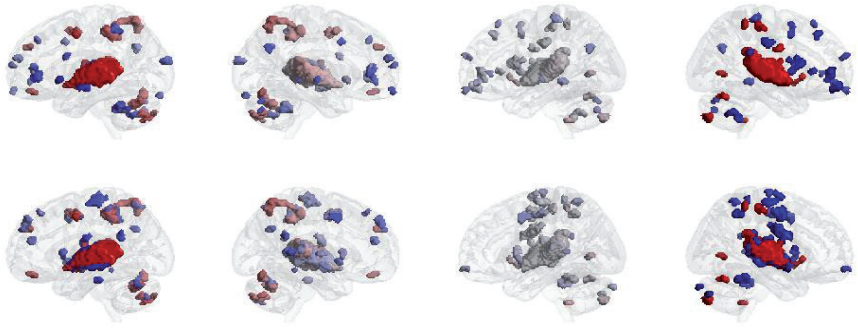

Figure 1.Combined K-map (top) and combined PC-map (bottom). Red and blue colors denote overlapping and unique areas, respectively. Note that polarity information is discarded here.

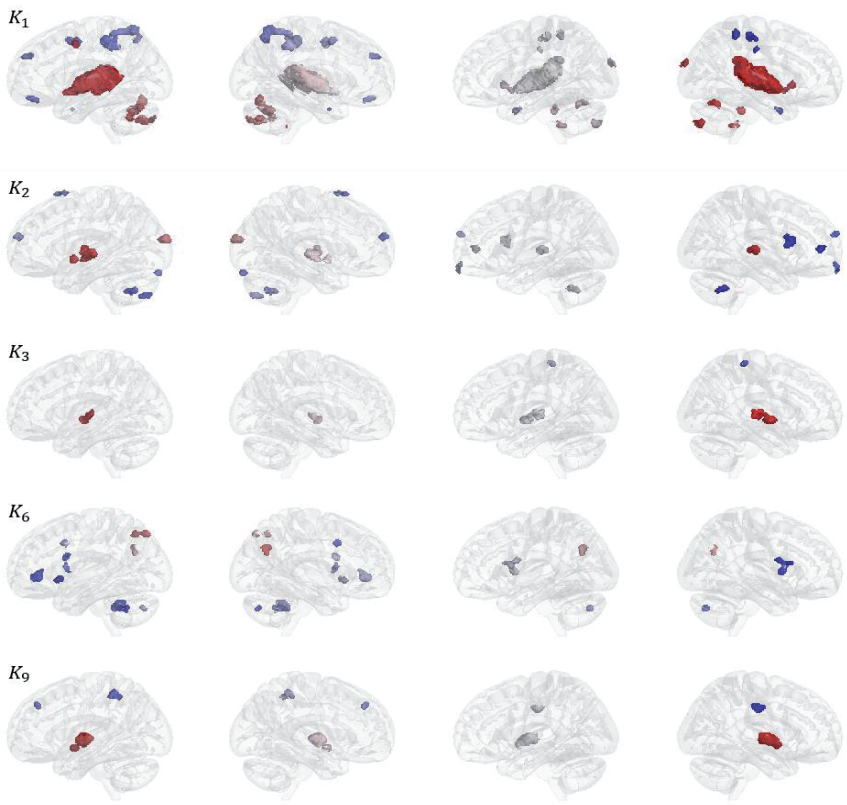

Figure 2. Spatial maps of the selected KPCA features. Note that polarity information of activations is preserved here. Red color denotes positive correlations and blue – negative correlations.

Another important issue related to the feature generation is the selection of the kernel function and its parameters. Due to the absence of the evaluation strategy that would quantify 'interestingness' of the features in early stages of the study, the most commonly applied Gaussian kernel was chosen as a starting point. Kernel width for the Gaussian kernel was also estimated based on the commonly applied heuristics. There is a plenty of room for improvements in this regard. For instance, both the kernel and parameter selection can be defined as an optimization problem where objective is to maximize correlations between the generated features and the behavioral data. This should also reduce the number of discarded features that fail to correlate with the perceptual ratings.

As mentioned before, K_2 and K_9 predicted Event Synchronicity and, Rhythmic Complexity ratings better than the linear PCA features. Both are high-level musical percepts characterized by rhythmic fluctuations of energy in different frequency bands (fluctuation spectrum). Event synchronicity is the centroid of the fluctuation spectrum and roughly relates to the average of periodicities in different bands, while Rhythmic Complexity relates to the noisiness of the fluctuation spectrum. In addition, both kernel features were correlated with Brightness ratings. As these features were absent in the original study, neural correlates cannot be compared with the previous findings. Nevertheless, auditory areas comprise the majority of positive correlates for both feature maps. From the regions

associated with K_2, previous studies on neural processing of rhythm implicate auditory cortex, SMA and cerebellum to be involved in rhythm perception [17]-[20] . Recruitment of frontal areas such as the right IFG may denote aspects of temporal attention processing during the perception of complex rhythms. Work by Chen et al. [20] revealed that frontocortical areas including the IFG were modulated by complexity of the rhythm during an auditory-motor synchronization task, and that they were more extensively recruited in musicians than in non-musicians. They argued this finding to underlie a greater involvement of working memory resulting from musicians' superior ability to organize temporal structure. Interesting to highlight is also activation of visual processing regions (left middle and superior occipital gyri) in response of increased K_2 and deactivation of right supramarginal gyrus in response to K_9, since visual areas have not been reported active in response to rhythm processing thus far. Nevertheless, the precuneus, another area associated with visuo-spatial processing and attention, was found to be negatively correlated with the clarity of the pulse in the original study. Given the lack of studies on rhythm complexity, and above-mentioned issues on reliability of perceptual interpretation, further investigation is needed in order to achieve a proper interpretation behind the functional neuroanatomy of these musical percepts.

To conclude, this study demonstrated that kernel PCA has the potential to produce features that captures novel musical percepts hidden from the linear features. Namely, two new rhythmic features representing periodicities in different bands of the frequency spectrum were identified. In addition to previously known anatomical regions involved in rhythm processing, visual processing areas were found to be associated with the new features. However, we recommend the highlighted limitations of this study to be addressed in future investigations before generalizing findings.

Aknowledgements

The first author wishes to thank Fabian Prezja and Virpi-Liisa Kykyri for their contribution and support. Part of this work was financially supported by the Academy of Finland (project numbers 272250 and 274037)

6. References

[1] U. Hasson *et al*, "Intersubject synchronization of cortical activity during natural vision," *Science,* vol. 303, *(5664),* pp. 1634-1640, Mar 12, 2004.

[2] H. J. Spiers and E. A. Maguire, "Decoding human brain activity during real-world experiences," *Trends Cogn. Sci. (Regul. Ed.),* vol. 11, *(8),* pp. 356-365, 2007.

[3] U. Hasson and C. J. Honey, "Future trends in Neuroimaging: Neural processes as expressed within real-life contexts," *Neuroimage,* vol. 62, *(2),* pp. 1272-1278, 2012.

[4] P. Janata *et al*, "The cortical topography of tonal structures underlying Western music," *Science,* vol. 298, *(5601),* pp. 2167-2170, Dec 13, 2002.

[5] J. D. Warren *et al*, "Separating pitch chroma and pitch height in the human brain," *Proc. Natl. Acad. Sci. U. S. A.,* vol. 100, *(17),* pp. 10038-10042, Aug 19, 2003.

[6] A. Caclin *et al*, "Separate neural processing of timbre dimensions in auditory sensory memory," *J. Cogn. Neurosci.,* vol. 18, *(12),* pp. 1959-1972, 2006.

[7] D. A. Abrams *et al*, "Inter-subject synchronization of brain responses during natural music listening," *Eur. J. Neurosci.,* vol. 37, *(9),* pp. 1458-1469, 2013.

[8] V. Alluri *et al*, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *Neuroimage,* vol. 59, *(4),* pp. 3677-3689, Feb 15, 2012.

[9] V. Alluri *et al*, "From Vivaldi to Beatles and back: Predicting lateralized brain responses to music," *Neuroimage,* vol. 83, pp. 627-636, 2013.

[10] P. Toiviainen *et al*, "Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data," *Neuroimage,* vol. 88, pp. 170-180, 2014.

[11] I. Burunat *et al*, "The reliability of continuous brain responses during naturalistic listening to music," *Neuroimage,* vol. 124, pp. 224-231, 2016.

[12] G. S. Sidhu *et al*, "Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD," *Frontiers in Systems Neuroscience,* vol. 6, 2012.

[13] 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, *(8),* pp. 769-780, 2007.

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

[15] K. Muller *et al*, "An introduction to kernel-based learning algorithms," *Neural Networks, IEEE Transactions On,* vol. 12, *(2),* pp. 181-201, 2001.

[16] B. Scholkopf *et al*, "Input space versus feature space in kernel-based methods," *Neural Networks, IEEE Transactions On,* vol. 10, *(5),* pp. 1000-1017, 1999.

[17] S. L. Bengtsson *et al*, "Listening to rhythms activates motor and premotor cortices," *Cortex,* vol. 45, *(1),* pp. 62-71, 2009.

[18] J. A. Grahn and J. B. Rowe, "Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception," *J. Neurosci.,* vol. 29, *(23),* pp. 7540-7548, Jun 10, 2009.

[19] J. A. Grahn and M. Brett, "Rhythm and beat perception in motor areas of the brain," *J. Cogn. Neurosci.,* vol. 19, *(5),* pp. 893-906, 2007.

[20] J. L. Chen, V. B. Penhune and R. J. Zatorre, "Listening to musical rhythms recruits motor regions of the brain," *Cereb. Cortex,* vol. 18, *(12),* pp. 2844-2854, Dec, 2008.