Consensus Clustering Framework for Analysing fMRI Datasets

Computational Framework & Neuroscientific Insights

A thesis submitted for the degree of

Doctor of Philosophy

by

Chao Liu

Department of Electronic and Computer Engineering

College of Engineering, Design and Physical Sciences

Brunel University London

June 2017

Abstract

Neuroimaging of humans has gained a position of status within neuroscience. Modern functional magnetic resonance imaging (fMRI) technique provides neuroscientists with a powerful tool to depict the complex architecture of human brains. fMRI generates large amount of data and many analysis methods have been proposed to extract useful information from the data. Clustering technique has been one of the most popular data-driven techniques to study brain functional connectivity, which excels when traditional model-based approaches are difficult to implement. However, the reliability and consistency of many findings are jeopardised by too many analysis methods, parameters, and sometimes too few samples used. In this thesis, a consensus clustering analysis framework for analysing fMRI data has been developed, aiming at overcoming the clustering algorithm selection problem as well as reliability issues in neuroimaging. The framework is able to identify groups of voxels representing brain regions that consistently exhibiting correlated BOLD activities across many experimental conditions by integrating clustering results from multiple clustering algorithms and various parameters such as the number of clusters K. In the framework, the individual clustering result generation is aided by high performance grid computing technique to reduce the overall computational time. The integration of clustering results is implemented by a technique named binarisation of consensus partition matrix (Bi-CoPaM) adapted and enhanced for fMRI data analysis. The whole framework has been validated and is robust to participants' individual variability, yielding most complete and reproducible clusters compared to the traditional single clustering approach. This framework has been applied to two real fMRI studies that investigate brain responses to listening to the emotional music with different preferences. In the first fMRI study, three brain structures related to visual, reward, and auditory processing are found to have intrinsic temporal patterns of coherent neuroactivity during affective processing, which is one of the few data-driven studies that have observed. In the second study, different levels of engagement, i.e. intentional to unintentional, with music have unique effects on the auditory-limbic connectivity when listening to music, which has not been investigated and understood well in neuroscience of music field. We believe the work in this thesis has demonstrated an effective and competent approach to address the reliability and consistency concerns in fMRI data analysis.

Acknowledgment

First, I would like to express my gratitude to my mother Shengxu Gao, father Deyun Liu, and other family members without whose support I would not have been able to concentrate and complete this thesis. I also would like to thank my first supervisor Professor Asoke K. Nandi for his excellent guidance and supervision through my PhD study and all the opportunities, such as attending conference and academic visiting, he has provided me with. Additionally, I would like to thank Professor Elvira Brattico, who is the main collaborator, for her guidance on neuroscience knowledge. I would like to thank my second supervisor Dr. Hongying Meng for his kind support and help. I also would like to thank Prof. Rick Li for his career advice.

I would like to thank Brunel University London for the research student funding that helps me focus on the research, as well as all the academic supports from the University, including the computing facilities and other hardware. I also thank Tongji University (China) for its financial support towards several academic meetings.

I would like to thank Dr. Rui Fa for his invaluable advices on my research. I thank Dr. Basel Abu-jamous and Dr. Zhechen Zhu for their guidance at the early stage of my PhD study. I would like to thank Dr. Ruoqi Geng for her support and advice on research and leisure activities. Last but not the least, I would like to thank all my friends and colleagues who have filled my research life with joy and fun.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background and motivation

The invention of magnetic resonance imaging (MRI) technique provides a non-invasive way of imaging the various tissues of human and animal body in a very high spatial resolution. The MRI technique has been widely applied to research and clinical applications. In 1990, Ogawa et al. published a new blood oxygen level dependent contrast in MRI (Ogawa et al., 1990), providing a way of measuring brain neural activities from the blood oxygen level within the blood vessels spreading across the whole brain, depending on the fact that cerebral blood flow and neuronal activation are coupled. Instead of imaging the static structural information of body tissues, the blood oxygen level dependent (BOLD) signal reflects the level of neural activation within imaged brain areas at a particular time, and by scanning these brain areas every several seconds, the whole MRI datasets reflect dynamic changes of neural activities during a period of time. The imaging of neural activities through measuring the BOLD changes is called functional MRI (fMRI). Since the discovery of BOLD contrast, fMRI has become a dominant brain imaging method to study the functions and interactions of different human or animal brain areas, due to the non-invasive features of the technique and relatively safe environment for participants, i.e., it does not require participants to undergo surgery or to be exposed to radiation like X-ray imaging.

As the acquisition and reconstruction of fMRI signals are complicated and BOLD signals have a very low signal to noise ratio (SNR), preprocessing is an important step for fMRI data analysis. Standard preprocessing includes the slice timing correction (important for fast event-related experimental design), head motion correction, spatial normalisation to standard structure brain template, and spatial smoothing. These steps will be introduced in Chapter 2 in more details. After the preprocessing of fMRI data, next task is how to analyse them. In order to extract meaningful information from fMRI data, many computational methods have been designed and applied to analyse fMRI data. In general, these methods can be divided into two groups, namely model-driven and data-driven approaches.

General linear model (GLM) is one of the most popular model-driven methods for analysing fMRI data. A model is first constructed to describe the BOLD signal changes corresponding to the stimulation during fMRI scanning. Statistical test is carried out on each voxel to determine whether the voxel's time-series can match with the pre-defined model that indicates the activated BOLD patterns. While GLM has helped neuroscience researchers to discover many important findings, it is hard to utilise this technique in some fMRI experiments when the paradigm is very complex or even impossible to model, for example, the experiment that researchers do not know when the responses elicited by stimulus will happen (Liu et al., 2000). Resting-state fMRI data for studying the functional connectivity during rest, where participants do not perform any explicit cognitive activities during scanning, also pose challenges to use traditional GLM approach.

To complement the model-based approaches and provide the researchers with more choices to analyse the neuroimaging data, various data-driven methods also have been proposed including principal component analysis (PCA), independent component analysis (ICA), clustering analysis, and so on. PCA and ICA are decomposition-based methods, meaning the results are on a different space from the original data. Clustering technique also have been widely used in various neuroimaging study, mainly in the resting-state fMRI data. Clustering analysis belongs to the unsupervised classification in machine learning domain, in contrast to the supervised classification that the data used to train the classifier have known category labels. Note that in machine learning field, there is another prediction technique called regression that is used to predict continuous values rather than the discrete category labels used in classification methods. Clustering technique involves the grouping of objects into a set of groups where objects within the same group show a high level of similarity and objects in the different groups show a low level of similarity. Compared to decomposition-based methods, clustering results may be more comparable to traditional functional connectivity maps from correlation with the seed regions, as they reflect functional connections between brain regions more directly. The main topic of this thesis is about the advanced development of clustering framework of analysing fMRI data. The rest part of the introduction will describe the motivation for the method development and validation in the thesis.

George E. P. Box, who is considered to be one of the great statistical minds of the 20th century (DeGroot, 1987; Peña, 2001), once said "Essentially, all models are wrong, but some are useful".

This statement can also apply to the clustering field. Since the concept of clustering was introduced, a large number of clustering algorithms have been proposed (Estivill-Castro, 2002). Regarding the fMRI experiment side, clustering is model-free, meaning no explicit assumptions need to be made on the experiment paradigm when running clustering analysis. However, in terms of algorithms themselves, each clustering process more or less asserts some underlying assumptions such as the data distribution. Nevertheless, these emerging data-driven clustering methods provide more possibilities and perspectives of investigating and understanding the neuroimaging data that convey some interesting and crucial information for understanding brain mechanisms underlying cognitive behaviours.

Typically, when using clustering analysis in an fMRI study on a certain stimulation paradigm an algorithm is chosen based on the authors' preference or some empirical experience. This raises the question of the generalisability of the results from a single method of analysis. Yet, if a second study on the same stimulation paradigm would utilize another method of analysis, a discordance is very likely to occur. Considering that in neuroimaging field methods of analysis and statistics have proliferated, it is inevitable that a somewhat confusing picture of the scientific progress gained by fMRI research would emerge. Call for consistent analysis of fMRI data has been made based on the finding "Here, we show that the average statistical power of studies in the neurosciences is very low. The consequences of this include overestimates of effect size and low reproducibility of results." (Button et al., 2013). Moreover, a very recent study points out that some of the well-known data analysis software packages (e.g. SPM, FSL, AFNI) are exposed to "bugs" or very high false-positive rates especially when not properly implemented or used (Eklund et al., 2016).

Motivated by the above situations in fMRI data analysis, we intend to explore the feasible way of addressing current issues. We set the scope of this thesis to the clustering analysis in analysing fMRI data, aiming to address the discordance of results from different clustering algorithms, which is a common phenomenon existing not only in fMRI data analysis but also other applications where clustering technique is used.

In this thesis, a comprehensive consensus clustering analysis of fMRI data framework is designed, including a series of methods from initial data processing, individual clustering, high performance

grid computing, to tunable consensus clustering results integrations from individual clustering and following statistical tests for extracting interesting information from clustering results. The framework is capable of merging results from many analysis methods in order to obtain robust and reproducible clusters from various datasets, which has been systematically validated using real fMRI datasets. The comparisons between the results obtained from proposed consensus clustering analysis framework and those from each individual clustering algorithm demonstrate that our framework has notable advantages over traditional single clustering algorithms in being able to evidence robust connectivity patterns even with complex neuroimaging data, involving a variety of stimuli. We believe this can greatly increase the consensual level among the clustering results from various clustering methods and neuroimaging data.

We claim this study to be highly application oriented as intensive collaboration with neuroscience researchers went through the whole procedure of framework development and validation. The collaborative study is between Brunel University London and Aarhus University in Denmark, with the fMRI data collected at the Advanced Magnetic Imaging centre of Aalto university, Finland. The valuable feedback from them are crucial to the successful application of the work in this thesis.

We have applied consensus clustering analysis framework on two real fMRI datasets, namely **Affect** and **Affect 2**. Affect fMRI study is to investigate the brain responses to music with emotions and listeners' preferences (affective processing). Several brain structures associated with visual, reward, and auditory processing consistently exhibiting intrinsic temporal patterns of coherent neuroactivity during affective processing have been identified without introducing any explicit model regarding the fMRI experiment paradigm. Affect 2 fMRI study is a further step towards understanding the neural correlates under different intentional engagement of music listening. The results show that the intentionality has different effects on auditory-limbic connectivity during affective processing of music, which enriches the relative literature on the neuroscience of music research field. The investigation stages of these studies vary. The Affect study has been published in a high impact factor journal International Journal of Neural Systems (IJNS) and has been shown on Brunel University London news webpage (https://www.brunel.ac.uk/news/articles/Dead-salmon-bugs-brain-scans-can-we-ever-agree-on-neuroscience-research). The results of Affect 2 study are currently under peer review.

1.2 Summary of major contributions

The original contributions of the works presented in this thesis can be grouped into the consensus clustering analysis framework development and neuroscientific insights. There are close interactions between computational methods development and application in real neuroimaging study throughout the research study. The framework proposed in the thesis has been employed on two real fMRI datasets namely Affect and Affect 2. The framework details, validation, and neuroscientific explanation and insights are described in Chapter 3, 4, and 5. The rest of this section will list some key contributions from this thesis.

1.2.1 Consensus clustering analysis of fMRI data framework development

a) **Main framework**. The consensus clustering analysis of fMRI data framework includes a series of methods and techniques that work together to function as a novel and powerful data-driven approach for clustering analyse of fMRI data. By applying the framework on fMRI data, voxels consistently having intrinsic correlated temporal patterns across many clustering experiments and datasets are able to be identified. The framework starts from the raw fMRI data processing and correct extraction of BOLD time series corresponding to the stimuli in the fMRI experiment. The clustering part includes multiple single clustering algorithms and a consensus clustering paradigm tailored and adapted to suit the needs of analysing fMRI data. Note that the grid computing technique is embedded in the framework and employed on individual clustering experiments, attempting to address the emerging issues of heavy computational load and large amount of time needed when analysing large-scale datasets. The last part of the framework includes the automated selection of clusters and filtering on cluster topology to refine the clustering solution. In general, the proposed framework has been demonstrated to be very competent in terms of cluster completeness and robustness on real fMRI dataset, which is reported in (Liu et al., 2017).

b) **Methods for extracting neuroscientific information from clustering results.** To complement the main framework, two statistical analyses are designed to make neuroscientific explanations from clusters obtained from consensus clustering. These two analyses

consist of investigation into two different aspects of BOLD signal, which are the magnitude and temporal patterns. These two analyses can be incorporated into the main framework in a flexible manner, meaning to employ them or not depends on the research questions asked regarding the fMRI experimental paradigm. Meaningful information has been evidenced by these two analyses that are reported in Chapter 4 as well as in (Liu et al., 2017).

c) **Scalable clustering for the extension of the main framework.** A scalable clustering algorithm is developed to suit the further extension of the main framework. This algorithm aims to address the accurate identification of the number of clusters K in the context of large-scale neuroimaging data, with the support from the grid computing technique employed in the main framework. We believe with the help of high performance computing and smart design of algorithms and analysis strategy, the consensus clustering framework is able to can deal with large-scale neuroimaging data at a fast speed and, most importantly, can accurately and robustly retrieve the important information in a data-driven manner (Liu, et al. 2015).

1.2.2 Neuroscientific insights

a) **Exploring the functional connectivity during affective processing**. In Affect study, we aim to validate the proposed framework in the real fMRI experiment that scans the brain activities when listening to music with emotions (happy or sad) and preferences (liked or disliked). We have obtained clusters including functionally and anatomically related neural networks responding to emotional music, which have been spotted in model-based methods but rarely observed with data-driven approaches. With the statistical analysis of clusters generated by the consensus clustering framework, the difference between musicians and non-musicians in the temporal profiles of the BOLD response for the interconnected cortical areas of visual cortex is identified. In addition, we also attempt to disentangle the music emotion and preference that tend to elicit stronger BOLD responses in observed functional neural networks. These results are reported later in Chapter 4 in more details.

b) **Investigating the effects of attentional and intentional engagement on affective processing.** An interesting phenomenon of music listening in everyday life is that the

engagement with music is not always intentional, such as listening to some background music when people are shopping or doing other activities. Very little knowledge has been accumulated regarding the effects of intentionality during these listening experiences. In this study, we utilise the established consensus clustering framework on a real fMRI datasets including three levels of intentional engagement with the music listening. The results show that intentionality has different effects on auditory-limbic connectivity during affective processing of music and prove that intentionality in judging the hedonic value of a musical piece is important in shaping neural connectivity to music, and specifically in connecting brain regions related to attention and cognition. These results are reported later in Chapter 5 in details.

1.3 Thesis structure

Next section of this chapter lists the aim and contributions of the research work. **Chapter 2** firstly reviews the functional magnetic resonance imaging (fMRI) technique and standard pipeline of fMRI data preprocessing. Then the usage of clustering technique in analysing fMRI data is systematically reviewed, including the traditional single clustering methods and recently adopted consensus clustering concept. **Chapter 3** details the methods and techniques that are used in this thesis. In terms of clustering technique, three basic single clustering algorithms are included, followed by the consensus clustering paradigm employed to integrate clustering results from multiple methods and multiple datasets. In addition, statistical tests designed to analyse the cluster data are explained. **Chapter 4** and **Chapter 5** introduce the consensus clustering analysis on two real fMRI datasets (Affect and Affect2) using the methods detailed in chapter 3. The fMRI experiment design, preprocessing, clustering experiment, results, and conclusion are reported in detail in each chapter. More specifically, **Chapter 4** describes the study exploring the brain functional connectivity during listening to music with different emotions and preferences, i.e., affective processing. This chapter also serves as a validation of the consensus clustering paradigm in real fMRI data analysis by comparing the results between the traditional single clustering and tunable consensus clustering strategy. In addition, the robustness and reproducibility are also tested to demonstrate the effectiveness of the method. **Chapter 5** describes the study investigating the effects of intentionality on brain functional connectivity during listening to unfamiliar music. Based on the established consensus clustering analysis pipeline in Chapter 4, this chapter focuses more

on neuroscientific explanations of the clustering results, highlighting the contributions to the enrichment of the literature in the study of affect in the field of neuroscience of music. **Chapter 6** summarises and discusses the work in the thesis and provides insights into future work, including one developed clustering algorithms designed to obtain clusters with number of clusters K automatically detected from a large dataset at a faster speed.

1.4 List of academic activities

1.4.1 Academic visiting

I went to Finland as a visiting research student in University of Helsinki and Aalto University for two months during the first year of my PhD study. I worked together with the collaborators in University of Helsinki, and Aalto University for 2 months from 1st September 2013 to 31st October 2013. During this period of time, I attended the simultaneous EEG-MEG data collection experiment and the fMRI data collection.

In Aalto University, I also worked with the staff in biomedical engineering and computer science where I learned the pipeline of preprocessing the EEG/MEG data. Afterwards, I participated in the preprocessing work for the free-listening experiment and contributed to the work related with Electrooculography (EOG) and electrocardiogram (ECG) artefacts removal. I also took part in fMRI data preprocessing training from fMRI school organised by School of Science in Aalto University as well as the experienced SPM software user.

1.4.2 Journal publication

1. Liu, C., Abu-Jamous, B., Brattico, E., Nandi, A.K., Towards Tunable Consensus Clustering for Studying Functional Brain Connectivity During Affective Processing. International Journal of Neural Systems, 27(2), p.1650042, 2017.
2. Abu-jamous, B., Liu, C., Roberts, D., Brattico, E., Nandi, A.K., Data-Driven Analysis of Collections of Big Dataset by the Bi-CoPaM Method Yields Field-Specific Novel Insights. Chapter in Lecture Notes in Electrical Engineering, March 2017 (invited book chapter).

1.4.3 Submitted journal manuscript

(Under review) Liu, C., Brattico, E., Abu-Jamous, B., Pereira, C.S., Jacobsen, T., Nandi, A.K., “Effects of Intentionality on the Neural Connectivity During Enjoyment of Unfamiliar Music – An fMRI and behavioural study”.

1.4.4 Peer-reviewed and full-length conference papers

1. Wong, M.L.D., Liu, C., Nandi, A.K., Classification of Ball Bearing Faults using Entropic Measures. In Surveillance 7, Chartres, France, 2013.
2. Nandi, A.K., Liu, C., Wong, M.L.D., Intelligent Vibration Signal Processing for Condition Monitoring. Proceedings of the International Conference Surveillance 7, Chartres, France, pp.1–15, 2013.
3. Liu, C., Fa, R., Abu-Jamous, B., Brattico, E., Nandi, A.K., Scalable Clustering Based on Enhanced-SMART for Large-scale fMRI Datasets. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, pp.962–966, 2015.
4. Liu, C., Abu-Jamous, B., Brattico, E., Nandi, A.K., Clustering Consistency in Neuroimaging Data Analysis. In 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, pp. 1118–1122, 2015.
5. Liu, C., Fa, R., Li, M., Nandi, A.K., Network Community Degree Based Fast Community Detection Algorithm for fMRI Data. In 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, pp. 1739–1743, 2016.

1.4.5 Poster presentation

1. Liu, C., Nandi, A. K., “Read Your Mind – Brain Signal Processing”. Presented in School of Engineering and Design ResCon13, Brunel University London.
2. Liu, C., Nandi, A. K., “Brain Signal Processing – A Data Driven Approach”. Presented in Graduate School ResCon14, Brunel University London.
3. Liu, C., Abu-Jamous, B., Brattico, E., Nandi, A.K., “Consensus clustering reveals neural networks during affective music processing”. The Neurosciences and Music V, 2014, Dijon, France.

Chapter 2 Literature Review

This chapter will first introduce the functional magnetic resonance imaging (fMRI) technique, including its basic concepts, and the structure of the fMRI image. Secondly, the standard pipeline of preprocessing fMRI data is introduced. Next two sections will review the general clustering technique in the broader domain and its application in cognitive neuroscience. The last section will summarise the reviewed literature and list the issues and challenges in applying clustering analysis in modern neuroimaging studies. In general, this chapter will rationalise the motivation and aim of the following study in the thesis.

2.1 Introduction to fMRI technique

2.1.1 fMRI basics principles

Magnetic Resonance Imaging (MRI) is an imaging technique utilising the strong artificially generated magnetic field (in units of Tesla) to create the image of various biological tissues (Huettel et al., 2009). The strong magnetic field is used to align the magnetic polarisation of certain atoms within the scanning area. While this strong magnetic field itself will not generate any MR signal, the actual MR image is created by using a series of changing magnetic gradients and oscillating electromagnetic fields, also known as pulse sequence, to alter the alignment of the magnetisation. Depending on the frequency of the pulse sequence, energy from the electromagnetic fields may be absorbed by certain atomic nuclei. After the energy is absorbed, the electromagnetic energy is later emitted by the nuclei, and the amount of emitted energy depends on the number and type of nuclei present. The emitted energy will be received by the radiofrequency coils. This detected electromagnetic pulse defines the raw MR signal. The pulse sequence can be constructed to create different contrasts on the images that reflect the amount of sensitivity to different tissues based on their physical properties. For example, the most commonly used structural contrast for anatomical images of brain is T_1 weighting while the images showing maximal signal in fluid-filled regions are T_2 weighted. There is another commonly used contrast called T_2^* related to functional MRI, which will be introduced in the next paragraph.

Functional Magnetic Resonance Imaging (fMRI) uses MR imaging technique to measure the metabolic changes in blood. The functions of human or other animals' brain are formed by the activities of the neuron cells, which demand the energy supplied by chemical reactions of glucose and oxygen. The neurons do not reserve energy in the form of sugar and oxygen internally, so the active neurons demand these energy substances to be brought in at a greater rate than to the inactive neurons. Hemoglobin in the blood is responsible of carrying oxygen. The form of hemoglobin with oxygen is called oxyhemoglobin existing in oxygenated blood while the form of hemoglobin without oxygen is called deoxyhemoglobin existing in deoxygenated blood. The intense need of oxygen supply by active neurons cause a change of the relative levels of oxyhemoglobin and deoxyhemoglobin in the surrounding blood vessels. The work carried out by Pauling & Coryell [1936] found that the oxyhemoglobin and deoxyhemoglobin have different magnetic properties. In 1990, with the MRI imaging technique already invented, Ogawa and colleagues (Ogawa et al., 1990) found that the presence of deoxyhemoglobin can decrease the MR signal on images with T_2^* contrast that is also named blood-oxygen level dependent contrast (BOLD). Since regional blood flow is closely related to neuroactivity, large volume of oxygenated cerebral blood will flow towards activated brain areas, bringing more oxyhemoglobin to these areas. Thus the bulk effect of neuroactivity will cause a regional decrease of deoxyhemoglobin, and an increase of signal strength on T_2^* weighted images.

In summary, the fMRI technique is able to measure the neuroactivity when the brain is performing certain functions. By designing the appropriate experiment and applying corresponding analysis method, the brain regions responding to the experimental paradigm can be identified. The introduction of fMRI provides neuroscientists with a non-invasive approach with high spatial resolution and many significant findings on the mechanisms of both human and animal brain have been made.

2.2 Preprocessing of fMRI data

2.2.1 fMRI data structure

As introduced in section 2.1.1, MRI technique could image different brain tissues based on the contrast predefined. By using T_1 contrast, the structural image of the human brain can be obtained. Each volume of brain structural image can be treated as a three dimensional matrix whose

elements are voxels (e.g., $2mm \times 2mm \times 2mm$ cube). Each voxel has a position $v(x, y, z)$ and an intensity value $I(x, y, z)$. Figure 2-1 (a) and (b) are examples of T_1 and T_2 weighted structural brain MRI image.

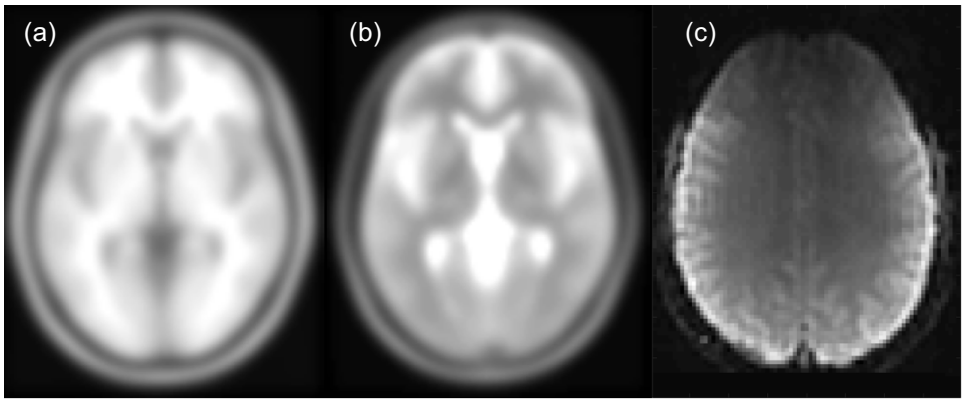


Figure 2-1. Examples of T_1, T_2, and T_2* images. T_1, T_2 images are from SPM8 software templates. Functional image is from the dataset used in the thesis.

Each volume of functional MRI images, as illustrated in Figure 2-1 (c), is also represented with a voxel matrix. The volumes are sampled repeatedly at a time interval T. Thus a complete fMRI dataset is a four dimensional matrix whose elements are voxels $v(x, y, z, t) = I$.

Ideally, among all the volumes during one session of fMRI scan, the voxel with the same coordinate (x, y, z) should represent the same location in the brain. However, due to various factors such as head motion, inhomogeneous magnetic field, and physiological noises (e.g., heart beats and respiration), distortion of raw MRI/fMRI data exists. To compensate, preprocessing is essential to reduce the variability of the data that is unrelated to the experiment.

2.2.2 Preprocessing of fMRI data

Slice-timing correction

Most fMRI data are acquired using two-dimensional pulse sequences that acquire images one slice at a time (Huettel et al., 2009). Depending on the capability of scanner, a typical pulse sequence might need 24 slices or more to cover the whole brain within 1.5 to 3 seconds. For example, one dataset (Affect) in this thesis uses 33 slices obtained in 3 seconds. All the slices are acquired at an equal spacing across the whole TR period. The slice-timing issue lies in the fact that all these slices covering the whole brain are not acquired at the same time but across a time period (TR). In addition, most pulse sequence uses interleaved slice acquisition, which means

the scanner first collects all the odd slices and then collects all the even slices, to avoid the interferences from the adjacent slices (cross-slice excitation). For instance, in Figure 2-2, there are three successive slices 1, 2, and 3. Assuming the TR is 2 seconds and slice 1 is sampled at time t, then the slice 2 is not sampled until $t + 1$ second. However, it is often needed that all the slices are sampled at a time point t without any delay to accurately model the BOLD signal change.

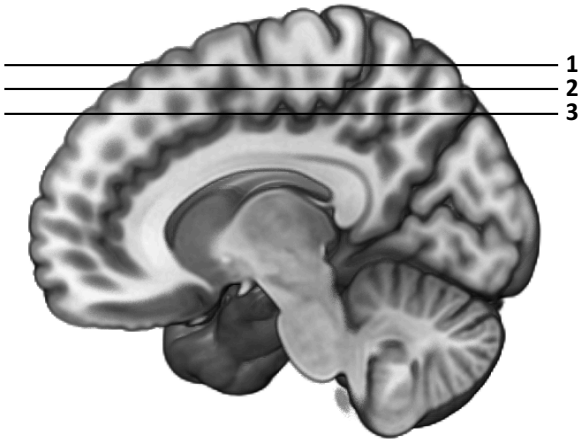


Figure 2-2. Slice-timing demonstration.

Thus, the correction for slice-timing is required.

The most commonly used slice-timing correction approach is temporal interpolation. Temporal interpolation utilises the timing information from adjacent slices to estimate the BOLD signal strength at the onset of a particular TR. Ideally, after the correction, the information in all the slices will be from the brain state at a time-point t. However, no matter what interpolation method is used, it is impossible to perfectly reconstruct the missing information from samples. In general, the interpolation correction is more effective for the data acquired with short TRs than that with long TRs (e.g. $\geq 3s$). Nevertheless, due to the large time intervals between slice acquisitions, data acquired with long TRs has a greater need for accurate slice-timing correction.

Head motion correction/realignment

In fMRI data analysis, each voxel is assumed to represent a fixed brain location across all the volumes during the whole scanning session. However, if there is head motion happening, the above assumption will not be valid anymore. If the MRI signal value of a certain voxel is concatenated to form the time-series, the MRI signal value of adjacent voxel might be included due to the motion effect, resulting in a mismatch of the location of subsequent images in the time-series.

Even a small amount of head motion is likely to cause great damage to the raw fMRI signal. To address the head motion problem, researchers normally use two procedures that are head motion restriction and mathematical correction of head motion in fMRI data.

Like many problems, head motion is easier to prevent than to correct. So before the actual correction of head motion is carried out, laboratories normally use some facilities to prevent the subjects' head from moving. Such facilities include biting bars, face masks, paddings, etc. It should be noted that although some head restraints approaches, such as biting bars, are very effective, they often raise the subject comfortability problems since the head is forced to be immobilised. When subjects are uncomfortable, they are likely to be distracted from performing the experimental tasks that they are asked to do, thus reduce the reliability of the data collected. Apart from head restraint approaches, head motion effects can also be greatly reduced by subject training. Before the real scanning session, subjects take part in a training session that will simulate the real experiment situations to let subjects be familiar with what will happen during the scan. By doing so, subjects will be more relaxed and comfortable in the real scanning session and can focus on the experimental tasks. Those who cannot accept the confinement of the scanner or avoid their head from moving too much can be excluded from real scanning session.

With the head restraints and subject training, head motion effects can be greatly reduced but small movement still exists. Thus the head motion correction is needed. The goal of head motion correction is to guarantee that brain is in the same position in every image across the whole scanning session. The process of spatially matching two images is called coregistration. For motion correction, all the image volumes are coregistered to a reference volume (e.g. the first volume of the whole volume set). Since the type of head motion problem belongs to bulk movement; the coregistration is performed by using rigid body transformation. Rigid body transformation uses translations (i.e., moving the entire image volume along the x, y, and z axes) and rotations (i.e., rotating the entire image volume around three axes) to impose one image over another. To determine these parameters, most coregistration algorithms use iterative approaches to identify a set of parameters that provide the best match between the processed image volume and the reference volume. Cost function is used to evaluate how well the two image volumes to be coregistered are matched. The ideal set of parameters should yield a minimum value of the cost function. But practically, the local minimum of cost function is often used to determine the parameters

due to the large amount of possibilities of the six parameters. After a set of translation and rotation parameters are computed, the original imaging data with head motion is resampled accordingly to estimate the values of voxels when there are no head motions.

Spatial normalisation

Almost all the fMRI studies collect data from more than one subjects. The implication is that to generalize one's inference to the population one must have a large number of subjects to reliably assess the between-subject variability (Friston et al., 1999). Some recent neuroscientific studies such as WU-Minn Human Connectome Project (Van Essen et al., 2012) even scanned more than 1000 subjects' brain activities during different experimental tasks. To draw the conclusion from a group of subjects, researchers often need to compare and summarise the activation within the same brain region. However, the human brain has big variabilities across individuals from size, shape, to the layout of different brain structures. Thus, the brain images of each subject must be transformed into a standard space where the brain images from all the subjects have identical size, shape, and structural layout. This process is called spatial normalisation that is important in voxel-based fMRI/MRI studies (Friston et al., 2003).

When the subjects' fMRI images are transformed into the standard space, each location of the brain can be uniquely indicated by a coordinate system also called atlas. The most commonly used normalised coordinate spaces are Talairach space (Talairach and Tournoux, 1988) and MNI space (Evans et al., 1994, 1992). In this thesis, MNI coordinates are used to indicate positions in brain.

Due to the fact that the spatial resolution of fMRI images is often lower than the structural images, a structural MRI brain image of the same subject is commonly obtained during each fMRI scanning. To match the individual functional brain images and the brain structures in a standard space more accurately, functional images are firstly registered to the structural images of the same subject using affine registration. The functional-structural coregistration can also correct the image distortion induced by pulse sequenced used to obtain functional images. Secondly, individual structural images are normalised into a standard space. Then the same normalisation process for normalising structural images into a standard space is applied to the functional images to make them match the standard space.

Spatial and temporal filtering

Spatial smoothing is another important step in preprocessing. There are several reasons to do it. Firstly, the functional signal to noise ratio (SNR) is very low in fMRI images, spatial smoothing can increase functional SNR by filtering out those high frequency background noises. Secondly, spatial normalisation cannot completely align brain structures in functional images due to the complexity of fMRI/MRI images as well as the limitation of normalisation algorithms. Spatial smoothing can blur the functional images for a better match with the brain structures in standard space. Thirdly, spatial smoothing can improve the statistical validity. In the statistical analysis of fMRI data, multiple comparison problems exist due to the large amount of tests carried out. Spatial correction can decrease the number of local maxima showing significant activity. Moreover, statistical tests commonly assume that error in measurement is normally distributed. Spatial smoothing increases the normality of data as the average of multiple observations tends toward a normal distribution (central limit theorem).

Temporal filtering is used to remove the artefacts lying in the time domain in fMRI data. The time related noises include high frequency ones such as heart beat (~1.0-1.5 Hz) as well as respiration (~0.2-0.3 Hz) and low frequency ones such as scanner drift (~0-0.015 Hz). For experiments with block design, which has a relatively slow task frequency, the low frequency noise can be really problematic when its frequency is close to the task frequency (Huettel et al., 2009). Depending on the task frequency, one normally use the high-pass filter with appropriate cut-off frequency to remove the low frequency components in the fMRI data. It is important to recognise that, like other preprocessing steps, temporal filtering cannot completely eliminate the noises in time domain but attenuate them. The inappropriate use of temporal filtering is likely to filter out the useful information, and as a result, temporal filtering should be carried out with caution. Thus, when filtering is needed in this thesis, commonly accepted procedures are adopted and they have been used in published works, indicating its validity.

2.3 Review of clustering analysis and its applications in fMRI studies

2.3.1 Introduction to clustering analysis

There are two categories of classification problems in machine learning field, namely supervised classification and unsupervised classification. Supervised classification predicts the class labels of new objects based on the rules extracted from the training objects with known class labels. Clustering analysis is an unsupervised classification technique, which organises a set of objects without class labels into classes where objects in the same class are similar to each other while are dissimilar to the objects in other classes, based on a given similarity or dissimilarity criterion. Each of these classes is called a cluster. In the context of clustering analysis of fMRI data, voxels exhibiting correlated BOLD responses are grouped into one cluster and voxels in different clusters exhibit different response patterns of their time series. For example, clustering algorithms are often used in analysing functional connectivity of human brain to parcellate brain into many homogeneous regions (van den Heuvel and Pol, 2010; Venkataraman et al., 2013).

2.3.1.1 Proximity measures

The proximity measures are fundamental elements of clustering, which define the criterion of grouping objects into a certain cluster. There are many different terms for proximity measurement; for instance, similarity, dissimilarity, distance and correlation. Despite many definitions, they are conceptually similar to each other in that higher similarity or higher correlation means that two patterns are more similar (with less dissimilarity) or geometrically closer (with less distance).

There are many distance metrics in the literature. Based on their definition of distance, they can be used in different applications. In general, the distance metrics can be grouped into two categories, which are distances of discrete feature objects and distances of continuous feature objects. In this paper, the fMRI data belongs to continuous feature objects as the BOLD signal value of one voxel is continuous. Xu and Wunsch (Xu and Wunsch 2005) summarised some common properties of these distance metrics, namely symmetry, positivity, triangle inequality, and reflectivity.

2.3.1.2 Measuring distances of discrete feature objects

Two well-known distance metrics for discrete data object are hamming distance and matching coefficient. Suppose there are two vector $\boldsymbol{u}$ and $\boldsymbol{v}$ with equal length l, whose elements can be binary $\{0,1\}$ or any finite number letters alphabet. The number of positions at which the elements need to be changed to turn $\boldsymbol{u}$ into $\boldsymbol{v}$ (or vice versa) is m. The hamming distance $HD(\boldsymbol{u},\boldsymbol{v})$ and matching coefficient $MC(\boldsymbol{u},\boldsymbol{v})$ are defined in Equation (2.1) and (2.2) respectively.

$$HD(\boldsymbol{u},\boldsymbol{v}) = m \tag{2.1}$$

$$MC(\boldsymbol{u},\boldsymbol{v}) = \frac{l-m}{l} \tag{2.2}$$

2.3.1.3 Measuring distances of continuous feature objects

Many distance metrics for continuous data were proposed such as Minkowski distance, Euclidean distance, Manhattan distance, Chebyshev distance, Mahalanobis distance, Pearson correlation, Jackknife correlation, Spearman's rank correlation, and Cosine similarity. Euclidean distance and Pearson correlation are commonly used in many fMRI studies (Ghosh et al., 2013; Liao et al., 2008; M. van den Heuvel et al., 2008; Yeo and Ou, 2004) to measure the similarity between the time series of two voxels. If one voxel's time series is denoted by a vector $\boldsymbol{p} = (p_1, p_2, p_3, \dots, p_n)$ and another voxel's time series is denoted by a vector $\boldsymbol{q} = (q_1, q_2, q_3, \dots, q_n)$, the Euclidean distance $d(\boldsymbol{p},\boldsymbol{q})$ and Pearson correlation $r(\boldsymbol{p},\boldsymbol{q})$ are defined in Equation (2.3) and (2.4) respectively.

$$d(\boldsymbol{p},\boldsymbol{q}) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{2.3}$$

$$r(\boldsymbol{p},\boldsymbol{q}) = \frac{\sum_{i=1}^{n}(p_i-\bar{p})(q_i-\bar{q})}{\sqrt{\sum_{i=1}^{n}(p_i-\bar{p})^2}\sqrt{\sum_{i=1}^{n}(q_i-\bar{q})^2}}, \bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i, \bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i \tag{2.4}$$

It should be remarked that different metrics are not always interchangeable. Choosing an appropriate metric depends on the different questions need to be solved. For example, the Euclidean distance pretty much sorts the time courses by their means (magnitudes) and ignore the shapes

of time courses. While Pearson correlation might be more suitable when the shape of time course is more important than the magnitude (Romesburg, 2004). In the context of fMRI data analysis, Pearson correlation consider time courses going up and down together as similar (correlated), ignoring the magnitude information of BOLD responses. However, when many clustering algorithms such as K-means are applied, the time courses of voxels are often normalised, thus the Euclidean distance can still be used to group those voxels exhibiting similar BOLD responses.

2.3.1.4 Categories of clustering algorithms

There are many ways of grouping clustering algorithms into different categories, depending on the chosen criteria to differentiate them. In addition, there are a large amount of publications on clustering algorithms and it is not easy to place each algorithm into a specific category. Thus, in this section, only those commonly used clustering algorithms in analysing neuroimaging data as well as the clustering strategy used in this thesis are included and they are grouped into seven categories that are partitional clustering, hierarchical clustering, fuzzy clustering, neural network based clustering, mixture model clustering, graph based clustering, and consensus clustering. The clustering methods used in this thesis will be introduced later in Chapter 3.

Partitional clustering

Partitional clustering algorithms aim to partition the dataset into a set of disjoint partitions, with each partition representing a cluster, according to a pre-specified optimisation criterion. Each cluster contains at least one object and each object belongs to only one cluster. The most representative example is K-means algorithm (MacQueen, 1967). There are some variants of K-means such as K-medoids (Kaufman and Rousseeuw, 1987) that using medians (data points) instead of means as centres of clusters during iterations.

Hierarchical clustering

Hierarchical clustering (Johnson, 1967) clusters the dataset into a set of nested clusters having a hierarchical structure that can be graphically represented as a tree structure. The tree structure is called dendrogram. In general, there are two types of strategies in hierarchical clustering, namely divisive approach and agglomerative approach. The divisive approach is a “top down” process that starts from a cluster containing all the data points, and splits are performed as one

moves from top to bottom of the hierarchy. The agglomerative approach treats each data point as a cluster at the beginning, then pairs of clusters are merged as one moves up the hierarchy.

Fuzzy clustering

Fuzzy clustering utilises fuzzy logic to determine which cluster each data point belongs to. Instead of assigning each data point to exactly one cluster, fuzzy clustering algorithms allow each data point to be associated to more than one cluster with a degree of membership, representing the different possibilities of belonging to corresponding clusters. A good example of fuzzy clustering category is fuzzy C-means (FCM) (Bezdek, 1983; Dunn, 1974) that has been used in many areas, including analysing fMRI data (Valente de Oliveira and Pedrycz, 2007).

Neural network based clustering

Neural network-based clustering starts with a set of nodes that are all the same except for some randomly initialised parameters that make each node behave slightly differently. Then these nodes learn from the data in a competitive fashion: active nodes reinforce their neighbourhood within certain regions, while suppressing the activities of other nodes. The nodes in output layer carry information of the membership of each data points. A well-known algorithm in this category is self-organising map (SOM) (Kohonen, 1982).

Mixture model clustering

Mixture model clustering is a model-based approach that assumes the data points within clusters follow the distribution of certain models (e.g. Gaussian or Poisson model) and attempts to optimize the fit between the data and the predefined models. The whole dataset is modelled by a mixture of these models. The distribution used to model a single cluster is often called component distribution. The most commonly used method of this category is based on Gaussian mixture model, which models the data points within a certain cluster using Gaussian distribution. By using this method, a scalable clustering algorithm has been proposed to cluster large-scale dataset with the number of clusters K automatically detected (Chapter 6).

Graph based clustering

Graph based clustering utilise the graph theories to partition the data modelled as a network represented graphically by a collection of nodes and edges. After the theory that human brain is a complex network with large amount of intra- and inter-regional connectivity is well accepted, the graph theory has been increasingly used to study the properties such as topological organisations of human brain network. One important task of investigating network is to identify communities (clusters) that are groups of nodes which are densely connected within each community while sparsely connected between communities (Fortunato, 2009). For example, the classical normalised cuts algorithm has been used to investigate the different functional areas in resting-state (i.e., without performing any active cognitive tasks) human brain (M. van den Heuvel et al., 2008).

Consensus clustering

Consensus clustering, also named ensemble clustering, refers to a situation that multiple clusterings results have been obtained for a particular dataset or a number of datasets and it is desired to find a single (consensus) clustering solution which is a better fit in some sense than the existing clustering result (Strehl and Ghosh, 2002), which is considered a solution to the problem of inconsistency of stochastic clustering algorithms or clusterings with different parameters. Normally consensus clustering does not operate on data directly, rather it incorporates the results, in terms of partition matrix or vector of labels, from multiple algorithms into a single representative or consensus, emphasising the common organisation in the different clustering results. Consensus clustering is expected to possess a set of properties such as providing more robust, more novel and more stable clustering results than single clustering algorithms (Ghaemi et al., 2009). The consensus of many clustering algorithms may compensate for possible errors by individual algorithms, generating more statistically reliable final clustering results than any single one can.

In general, there are four classes of consensus clustering methods:

1) Partition-partition comparison. The objective of this approach is to maximise the similarity between the final consensus partition and each individual partition.
2) Cluster-cluster comparison. In contrast to 1) that compares the partition with partition, methods in this class compare single cluster from different partitions.

3) Member-in-cluster voting. The first step is to match most-similar clusters from different partitions to each other. Then partition vote for the membership/belongingness of each object to the corresponding cluster.
4) Member-member co-occurrence. A co-association matrix is constructed based on the frequency of co-occurrence of pairs of objects, i.e., they are clustered in the same cluster in different partitions.

2.3.2 Clustering in fMRI data analysis

In general, there are two big families of methods of analysing fMRI data, namely hypothesis-driven approach and data-driven approach. A representative example in hypothesis-driven family is the well-known general linear model (GLM) (Friston et al., 1995). The GLM method assumes that the observed fMRI data can be modelled as the sum of separate factors along with additive Gaussian noise. This assumption often limits the performance and application of the GLM method. For example, it requires the accurate estimate of the BOLD signals corresponding to the experiment tasks, which is often very difficult to achieve due to the factors such as the subject variations. Besides, some experiment paradigms are very complicated and the BOLD changes corresponding to the tasks are hard or even impossible to model precisely such as the studies of long-term memory formation with complex time series data (Hasson et al., 2008) and decoding of intensions (Haynes et al., 2007). Recently, the design of some fMRI experiments does not include the periodic tasks such as the naturalistic listening to music (Alluri et al., 2012; Burunat et al., 2016), which is even harder to model using GLM approach. Under the circumstances where it is not suitable to use GLM, data-driven and machine learning approaches provide a powerful complementary tool to extract interesting information from data.

Clustering technique is among the most popular techniques in data-driven family and has been applied in analysing fMRI data for more than two decades. However, it still draws a lot of attention nowadays, with new algorithms often proposed and discussions on their applications in analysing fMRI data. Clustering has been used in detecting brain activation, generating brain parcellations, and analysing functional connectivity, with the latter two being the most popular applications of clustering analysis nowadays.

Functional connectivity is defined as the temporal dependency of neuronal activation patterns of anatomically separated brain regions (van den Heuvel and Pol, 2010). Functional connectivity study using fMRI (fcMRI) examines regional interactions in the brain and has drawn increasing attention in the past years. Brain parcellations divide the brain's spatial domain into a set of non-overlapping regions or modules that show some homogeneity with respect to one or more metrics that define homogeneity, such as the anatomical connectivity and functional connectivity. The rationale behind brain parcellation is the structures of interest are often not at the level of a single voxel but at a level of a region consisting of many voxels (Thirion et al., 2014). Using the information from multiple voxels can reduce the computational pressure (i.e., parcellation vs individual voxel) as well as reduce the noise level from single voxel, which is similar to Gaussian smoothing in fMRI data preprocessing. In the following part, only the brain parcellations based on functional connectivity are discussed as other homogeneity measures such as anatomical connectivity are based on other neurological data rather than fMRI data. In addition to classical model-driven seed-based approach, many data-driven methods, such as principle component analysis (PCA), independent component analysis (ICA), have been applied to analyse functional connectivity. Clustering technique, which aims to maximise the similarity (i.e., homogeneity in the context of brain region activity) between data points within the same cluster and the dissimilarity between the data points from different clusters, is therefore also widely applied to analyse functional connectivity or generate brain parcellations. Besides, clustering analysis has several advantages over other approaches used in functional connectivity analysis. For example, ICA-based methods search for a number of maximally independent components representing the underlying sources of the observed BOLD signal. But these independent components are often perceived as very difficult to understand (Thirion et al., 2014). Other decomposition based methods such as PCA also suffers from this issue. On the other hand, clustering analysis groups the voxels, which represent the physical brain regions, into clusters and reflect functional connections among different brain regions more directly and thus easier to interpret and more comparable to the classical seed-based functional connectivity map (fcMAP).

Unlike the brain activation detection that is done on task-related data, most functional connectivity analysis and brain parcellations are performed on resting-state fMRI (rsfMRI) data that measures

the spontaneous brain activity when subjects are instructed to relax and not to perform any intentional tasks (Biswal et al., 1997, 1995). The BOLD signals from resting-state brain are mainly dominated by low frequency (<0.1 Hz) components. It is believed that these resting-state BOLD fluctuations of cortical and sub-cortical areas originate, at least partly, from spontaneous neuronal activity and the temporal correlation between the BOLD time-series of different brain regions reflects the functional connectivity between different brain regions during rest. Examining resting-state brain has many applications. For example it has been shown that several brain regions display a high level of correlated BOLD signal activity across heathy subjects (van den Heuvel and Pol, 2010). Thus studying the rsfMRI data is useful to diagnose possible neurological or psychiatric diseases by comparing the brain functional organisations between healthy people and patients. Apart from the majority of studies using rsfMRI data, there are also research using paradigm fMRI data for studying the functional connectivity during certain cognitive tasks (Lashkari et al., 2012, 2010; Michel et al., 2012).

2.3.2.1 Activation detection

As introduced in the last two paragraphs, model-based methods are most widely used to detect activations in brain despite some limitations. Researchers have explored alternative ways to analyse the activation patterns of voxels. Fuzzy clustering such as fuzzy C-means is one of the most used clustering methods and has been applied to fMRI data analysis from a very early time. Two studies (Baumgartner et al., 1998; Moser et al., 1997) used fuzzy C-means to detect the activation within the visual cortex. These two studies considered the time course of each voxel as a vector T with a dimension of N that is the number of the time points. Fuzzy C-means was applied directly on the time domain to acquire the clusters. The signal changes within each cluster were represented by the centroid of the corresponding cluster. The cluster whose centroid exhibiting an activation pattern is considered the activated regions. They also compared the performance of fuzzy C-means with other approaches such as correlation analysis (CA) and concluded that fuzzy C-means performs well on fMRI data with typical (low) functional contrast to noise (CNR) level and does not require prior knowledge about the stimulation protocol. Despite this claim, there are several problems with traditional fuzzy C-means. Firstly, fuzzy C-means is sensitive to noise and requires the number of clusters as a user input, which greatly influences the clustering results.

The results also depend on the initialisation used. Later another study (Chuang et al., 1999) designed a cascade classification scheme where Kohonen clustering neural network was used in the first stage to obtain many small clusters at a faster speed than fuzzy C-means. Then fuzzy C-means was used to reorganize and merge the identified features into primary ones in the second stage. However, in this study, the same questions such as the optimal number of clusters are still open to discussions. Fadili et al. discussed the number of clusters and the fuzziness index for fuzzy C-means when applied on fMRI data (Fadili et al., 2001). However, the decision on optimal number of cluster K was based on repeatedly testing K value using validation method. This is applicable when the size of data is relatively small. But when the size of fMRI data is huge, this approach will be very computationally expensive and is almost impossible for average lab computers to execute.

K-means and hierarchical clustering are two other algorithms that were applied at an early stage to analyse fMRI data since the BOLD contrast was discovered. Goutte et al. used K-means and hierarchical clustering together with a cross correlation metric to analyse the activation in human brain visual area (Goutte et al., 1999). In this study, both clustering algorithms were carried out on cross correlation functions between voxel time series. The authors explored the choice of number of clusters K and its influence on the homogeneity of the clusters and called it a number of cluster/homogeneity dilemma. Then the author stated the combination of the two methods is probably a potential way to exploit the attractive features of each algorithm. This is probably the first call for a consensus clustering analysis of fMRI data. Later Stanberry et al. designed a dendrogram sharpening technique combined with a hierarchical clustering algorithm, which does not need the number of clusters as user input, to identify brain activation (Stanberry et al., 2003). The dendrogram sharpening technique started from removing data from the low-density regions in order to obtain a clearer representation of the data structure. Then after the centres of clusters were identified, the classification algorithm was run on voxels to reassign them to the detected centres of clusters. The disadvantage of this method, which was also mentioned in the discussion of the original paper, is that it is highly sensitive to the other parameters needed to carry out the dendrogram sharpening. Following the data reduction scheme used in previous two studies, Chen et al. presented a neighbourhood correlation (NC) approach to keep the candidates of the actual active voxels and then used the hierarchical clustering together with a newly proposed spatial-

temporal measure, which incorporates the NC and the L1 norm spatial distance, to detect brain activations (Chen et al., 2006). This approach has several limitations such as the high sensitivity to signal patterns and time delays of the BOLD responses. The incorporation of spatial constraint might weaken its ability of detecting separate regions showing similar BOLD responses considering the wildly accepted theory that the brain is a complex network (Bassett and Bullmore, 2006; Bressler and Menon, 2010; van den Heuvel and Pol, 2010). In addition, to reduce the amount of data, the threshold value for the NC map needs careful consideration, whose problem is similar to the choice of parameters of clustering algorithms.

For some experimental paradigms, the timing and location of the activation are completely unknown. Thus the GLM model is impossible to use. Liu et al. (Liu et al., 2000) studied the temporal response of the brain after eating. Since it is not known when and where the BOLD signal start to change after food digestion, clustering the whole time series (the continuous functional scan lasted for 48 mins) would be computationally expensive and is very likely to suffer from the curse of dimensionality in clustering problems. To address this issue, the authors developed temporal clustering analysis (TCA), in which the three-dimensional brain was collapsed into a one-dimensional space, to search for the maximal response in a combination of signal intensity and spatial extent. TCA method can only detect the largest peak of the activation time windows well when applied only once, if multiple response peaks at the same location of the brain occur. Gao and Yee developed the iterative temporal clustering analysis for detecting multiple response peaks in fMRI (Gao and Yee, 2003). Although these methods proved to be effective, the application is limited to the aforementioned experiment scenarios.

Some studies applied clustering techniques as an auxiliary tool to detect activation in the brain. Meyer and Chinrungrueng used clustering to distinguish the activated voxels and non-activated voxels by dividing the voxels covering a local brain into two clusters in Fourier domain (Meyer and Chinrungrueng, 2005), based on the previous finding that the BOLD response to a periodic stimulus can be well characterised by Fourier coefficients (Lange and Zeger, 1997). Similarly, support vector clustering (SVC) was used to classify the Fourier coefficients of fMRI time series (Wang et al., 2005), utilising SVC's features that no shape or number of clusters are required as user input. Albeit, in this study, there are two crucial free parameters that influence the effectiveness of this approach, i.e., Gaussian kernel width σ, determining how fine the samples are clustered, and

regularization constant C that affects the amount of outlier points. No systematic solution of choosing the appropriate values of these two parameters was given. Furthermore, the Fourier transformation based approaches are limited to periodic experimental paradigms.

Another interesting work related to activation detection is estimating hemodynamic response function (HRF). Badillo et al. used consensus clustering based on random parcellation of fMRI data to robustify the HRF estimation by combining hemodynamics results provided by different parcellations. The consensus clustering strategy used here consisted of running a clustering algorithm multiple times, with different number of clusters K, on different perturbations of the original data and combining the resulting clusters. The perturbations were generated by randomly undersampling the data along the temporal axis. To combine multiple clustering results, the connectivity matrices were averaged to yield a consensus matrix whose entries represent the number of times two particular voxels were assigned to the same cluster. Based on the parcellations generated by clustering the consensus matrix, the HRF estimation was carried out. Similar to (Bellec et al., 2010), this work used a $N \times N$ matrix, with N be the number of voxels in the dataset, to indicate the membership of the voxels, which limits its application to ROIs based parcellations or clustering a specific area of the whole brain.

2.3.2.2 Functional connectivity and functional brain parcellations

Fuzzy clustering

Golay et al. developed a correlation based fuzzy clustering algorithm to analyse the functional connectivity of human brain (Golay et al., 1998). The authors proposed two distance metric D_1 and D_2 based on Pearson correlation coefficient $CC_{x,y}$. D_1 is defined as $\left(\frac{1-CC_{x,y}}{1+CC_{x,y}}\right)^{\beta}$ and D_2 is defined as $2(1-CC_{x,y})$. The distance between two voxels below a predefined threshold is considered functionally connected. The above two distance metrics were compared by using fuzzy clustering and the distance D_1 was claimed to yield the best results.

Hierarchical clustering

Golay et al. computed the Pearson correlation coefficient based on the whole time-series containing all frequency information in addition to the very low frequency (≤ 0.1 Hz) component that

reflects the underlying neural activities during rest. Cordes et al. firstly utilised the hierarchical clustering to study functional connectivity when brain is at resting-state using a new distance based on frequency analysis (Cordes et al., 2002). This study implemented spectral decomposition (Cordes et al., 2001, 2000) of correlation coefficient $cc_f(p,q)$ between two voxels p and q, which has a property that the sum over all frequencies will yield the correlation coefficient between voxels p and q. Then the author proposed a distance measure $d(p,q) = 1 - \sum_{f=0}^{0.1Hz} cc_f(p,q)$ to retain the correlation coefficient from the low frequency components in the BOLD signals. Only the voxel pairs with $d \leq 0.7$ (i.e., low frequency correlation coefficient ≥ 0.3) were kept for single link hierarchical clustering analysis. A linkage inconsistent coefficient $I = 0.9$ (the difference between the current link height and the mean, normalized by the standard deviation) was used as the threshold for cluster size identification. The author only chose four slices covering auditory cortex, motor cortex and visual cortex rather than ran the experiment on the whole brain fMRI data, partially due to the hardware limitation at that time, which also might be the reason the single linkage method was used as it was easy to implement. However, a recent study (Thirion et al., 2014) claimed that the Ward's method (Ward, 1963) in hierarchical clustering yield comparatively better clustering results.

Salvador et al. later also used hierarchical clustering to investigate the basic hierarchy of brain functional organisation in healthy volunteers (Salvador et al., 2005). Firstly, the brain images were parcellated into 90 regions according to a prior anatomical template. Mean time-series of BOLD signals within each region were then computed for each subject. An individual inter-regional partial correlation matrix was constructed from these regional mean time-series for each subject. Lastly the group mean inter-regional partial correlation matrix was formed by averaging the individual partial correlation matrix of each subject. The author did not explicitly specify the cluster number k but inspected the hierarchy of dendrogram at three general different levels, namely lowest level, intermediate levels, and highest level. It was found that the symmetrical links between bilaterally homologous cortical and subcortical regions were represented at the lowest level of the hierarchy; The symmetrical pairs of homologous regions were clustered together in configurations corresponding approximately to sub-lobar or gyral domains at intermediate levels; At highest level, six large clusters were formed corresponding to four neocortical/paralimbic systems,

one paralimbic/limbic system and one subcortical/limbic system. This study illustrated the advantage of hierarchical clustering that one can view the data from multiple perspectives, yielding different organisations of the data at different hierarchies. The way that using the mean time-series to represent the activity of the anatomically separated brain regions was debatable as the anatomically homogenous region does not necessarily activate in a similar way. In other word, the functional and structural organisation of brain network is not a one-to-one relationship and need more investigations (E. Bullmore and Sporns, 2009). Eickhoff et al. applied hierarchical clustering using Euclidean distance and Ward merging criterion to combine the individual seed voxels, obtained by meta-analytic connectivity mapping, into larger regions (Eickhoff et al., 2011). Compared with other studies using K-means, the authors claimed the advantage of hierarchical clustering, like in (Salvador et al., 2005), that it allows the multiple-layered views of various possible clustering results.

The classical unsupervised hierarchical clustering can also be used in a supervised manner. Michel et al. designed a supervised hierarchical clustering that addressed some limitations of the unsupervised feature agglomeration approaches (Michel et al., 2012). Firstly, a hierarchical subdivision of the search domain using Ward merging criterion was constructed. In this step, a connectivity constraint was added so that only the adjacent clusters can be merged together (parcels), which was also used later in (Blumensath et al., 2013). Then instead of cutting the dendrogram horizontally to yield corresponding parcellations, the dendrogram, which can be viewed as a tree, was pruned in a supervised way. The cut was initialised at the highest level of the hierarchy, followed by successively finding the new sub-tree that maximise a prediction score. The prediction score was generated by a prediction function, such as support vector machine (SVM), instantiated with the parcel-wise signal averages at the current step. one parcel was split at each step, where the chosen split yielding the highest score was performed. Finally, model selection was used to select the sub-tree that generates the optimal value for prediction score, with the corresponding parcellation be the final result. Such a supervised cut (SC) scheme requires computational power heavily especially for the large scale data, in which case the prediction function needs lower complexity to make the scheme applicable.

K-means

Flandin et al. utilised K-means with geodesic distance to partition the brain into a user defined number of regions (Flandin et al., 2002). Then the functional images were oversampled at this resolution using spline interpolation, followed by averaging all the voxels' time-series within each cell/region as the mean activity. The aim of this strategy is to increase the sensitivity of detection in GLM analysis. By using a cognitive paradigm studying motor task related brain areas, the authors found a large increase in detection sensitivity under 340 and 1700 parcels compared to voxel-based GML analysis. A more recent study carried out by Yeo et al. used K-means to parcellate the cerebral cortex into networks of functionally coupled regions (Yeo et al., 2011). This study used a stability analysis of clustering algorithm to explore the appropriate number of clusters. In brief, the region of interests (ROIs) were randomly and repeatedly divided into two groups and clustering was run separately on the two groups, followed by measuring the reproducibility of clustering algorithm's results for a certain number of cluster K. Then for the same K, all the vertices (18715) were also randomly and repeatedly divided into two groups. The model parameters learned from clustering one group of vertices were then used to predict the clustering results of the second group of vertices. By comparing the results of prediction and clustering on the second group, the generalization power of the clustering with corresponding parameters were measured. The stability analysis suggested a relatively coarse solution with 7 clusters and a finer one with 17 clusters. Nonetheless, none of the conclusions depended on a strong assumption that these two number were the correct solution to parcellate the cortex. Later Kahnt et al. used standard K-means with correlation distance metric to parcellate human orbital frontal cortex (OFC) based on functional connectivity (Kahnt et al., 2012). The number of clusters K was chosen from 2 to 10 and various approaches were applied to ensure the accuracy and stability of clustering results. For each K value, the best result from 100 repetitions with different initialisation methods was retained. Squared Euclidean distance was used to rule out the bias caused by the distance metric based on the fact that the results generated by two metrics were very similar. When averaging the connectivity matrices across subjects, the leave-one-out strategy was used N (the number of subjects) times to obtain the average stability map. To evaluate the cluster solutions, the variation of information (IV) (Meila, 2007) was used and, similar to Yeo et al.'s work, a split half strategy

was employed to measure the stability of cluster solutions. The authors examined the OFC parcellation with different K values (similar to a hierarchical approach) and then tried to identify the optimal K value ($K = 6$ in this study). On choosing the most appropriate parameters, both of the above two work relied on repeatedly running the algorithm or dividing the datasets into halves and compute the relative scores indicating the quality of clustering results. It is applicable for the relatively small range of K values or small scale datasets (e.g. part of the brain rather than the whole brain) and it becomes increasingly difficult for large scale data (Lichtman et al., 2014).

Mixture model

Golland et al. proposed an unsupervised segmentation of fMRI volumes using mixture model together with Expectation–maximization (EM) algorithm (Golland et al., 2007). The authors formulated the connectivity analysis as partition of voxels into classes that are well characterized by N_s representative hypotheses, or time courses, m_1,… m_{N_s} based on the similarity of their time courses to each hypothesis. At the beginning, random N_s voxels were selected and their time-courses were used as an initial guess for the cluster means. Then a set of update rules were applied at each step before convergence. Various number of classes were used to generate a hierarchical view of parcellation. To make sure that the results of segmentation of each subject comparable, an approximate algorithm was employed to match the label assignment across subjects so that the same label number indicates the same class. Results of mixture model highly depend on the chosen model where this study used a normal distribution to model class-conditional densities. Later several studies also used mixture model to parcellate the fMRI data (Lashkari et al., 2012, 2010; Tucholka et al., 2008) where other models such as von Mises-Fisher distribution were used.

Spectral clustering

Thirion et al. combined the parameters in GLM analysis of an event-related paradigm together with spectral clustering and K-means (Thirion et al., 2006). K-means clustering algorithm was used to derive parcel prototypes on GLM parameters under a spatially constrained setting. Then seed voxels were found that correspond to the parcel prototypes defined. Spectral clustering was used to assign the voxels in each individual brain to these seed voxels. The author discussed a potential alternative to spectral clustering, which is agglomerative clustering, and claimed spectral

clustering had advantages over agglomerative approach for irregularly sampled data. In addition, agglomerative clustering algorithms are known to be sensitive to noise and outliers. Later Van den Heuvel et al. reported a voxel based model-free normalized cut (Ncut) graph clustering approach with whole brain coverage for group analysis of rsfMRI data (M. van den Heuvel et al., 2008). Each individual's rsfMRI dataset was modelled as a weighted and undirected graph with nodes representing the voxels and edges connecting two voxels were computed as the correlation between their filtered BOLD time-series. In order to reduce the graph complexity, a cut-off threshold rc was applied to make the weights of the edges that did not reach rc to zero. rc was set to 0.4 based on the facts that the group clustering results did not vary significantly from individual clustering results compared with other rc values (0.3 and 0.5). Regarding the number of clusters, the authors choose 20, double the number of 6 to 10 networks reported in previous fMRI resting-state studies, to avoid the underclustering that is the number of clusters used is less than the actual classes in the data. In addition, it was demonstrated that the overclustering at the individual level would not affect the nature (e.g. shape or outline) of group clustering results. The group clustering was done on the group graph that reflects the RSN consistency across the group of subjects. One important aspect in group clustering was parameter setting. The authors firstly performed Ncut with respect to different combination of cut-off threshold and number of clusters, yielding a $P \times Q$ matrix, with P the range of used numbers of clusters and Q the range of used graph complexity thresholds. For each combination, the Ncut cost, defined as the summation of the weights of the edges that have to be removed to divide the group in multiple sets, was computed. The combination with the smallest Ncut cost was chosen as the optimal parameters for generating the final group clustering results. Spatial constraints can also be introduced in spectral clustering. Craddock et al. proposed a spatially constrained spectral clustering of whole-brain rsfMRI data by only computing the correlations between a certain voxel and its neighbours (26 voxels, face and edge touching) when forming the similarity matrices (Craddock et al., 2012). This constrained the clusters to contain the contiguous voxels and the spatial constraints resulted in sparse similarity matrices, which reduce the computational loads. In terms of parameters selection, the authors firstly obtained the clustering results from all the combination of parameters, similar to (M. van den Heuvel et al., 2008), and evaluated them using leave-one-out cross-validation (LOOCV) and silhouette width (SI).

In addition to the classical spectral clustering, Chen et al. used a multi-view spectral clustering approach (Kumar and Daumé, 2011) to inspect the group-wise consistent multimodal (diffusion tensor imaging (DTI) and rsfMRI) brain networks (Chen et al., 2013) based on 358 ROIs developed in an earlier work (Zhu et al., 2013). A co-training approach based on spectral clustering to maximize the agreement between the structural network and functional network, and then the agreement between subjects, to find the group-wise consistent multimodal connectomes of the human brain. The agreement between DTI and rsfMRI views was achieved by projecting the data affinity matrix of one view to the eigenspace of the other view. The final fused connection matrix can be obtained by calculating the average normalized matrix between different subjects and views. Base on fused connection matrix, the final multi-modal connectomes of human brain will be obtained directly by applying classical spectral clustering algorithm. During the co-training stage, the affinity matrix was projected to the first K eigenvectors of the graph Laplacian. Ideally, K should be set equal to or larger than the number of clusters in the data. The authors tested different K values and set K to 25, considering the number of nodes in the constructed network was 358.

Network-based community detection

Note that spectral clustering deals with a network constructed from the datasets and in the field of network theory, the clustering problem is commonly called community detection. Human brain has been identified as a complex network based on structural connectivity or functional interdependence (Bressler and Menon, 2010; Sporns, 2014). Graph theory has become the main analytical tool in the quantitative analysis of complex networks, and has naturally been employed in the studies of brain networks topology and organisation (E. T. Bullmore and Sporns, 2009). Identifying network communities is one of the important contributions of graph theoretic network analysis. Community detection techniques, which are data driven methods, have gained great popularity in the study of identifying RSNs and functionally connected regions (Crossley et al., 2013; Power et al., 2011; Schwarz et al., 2009; Sporns, 2014; M. P. van den Heuvel et al., 2008). In addition to the Ncut algorithm introduced in the spectral clustering section, a leading eigenvector community detection algorithm by Newman (Newman, 2006) was used in a meta-analysis of a large primary literature that used fMRI or PET to measure task-related activation (Crossley et al., 2013). The major issue of Newman ’s algorithm is that matrix decomposition (same as Ncut) is

required and this limits its use in the large-scale voxel-based network analysis. Schwarz and colleagues employed the 'Fast Modularity' algorithm (Clauset et al., 2004) in their study of the response of the rat brain to acute pharmacological challenge (Crossley et al., 2013). Although claimed to be "fast", in practice, the 'Fast Modularity' algorithm takes a relatively long processing time in large-scale networks analysis. Power and colleagues (Power et al., 2011) utilised the InfoMap algorithm, which was developed by Rosvall and Bergstrom (Rosvall and Bergstrom, 2008) to test their hypothesis that specific patterns of high correlation within functional systems would be reflected as sub-graphs within a brain-wide resting state fMRI network. The InfoMap algorithm is faster than the 'Fast Modularity' algorithm; however, because of its stochastic nature of random walk strategy, like other traditional clustering algorithms, each run of the algorithm will provide different community results. Obtaining meaningful and stable networks require multiple trials and evaluations. Nevertheless, InfoMap is very competitive among the community detection algorithms.

Consensus clustering

Consensus clustering has not been often used in fMRI data analysis. To date, only a few studies adopted this concept including published works from this thesis (Liu et al., 2017, 2015a). The rationale behind consensus clustering is to extract the stable features from the clustering results and thus reduce the random effects from clustering algorithms. Bellec et al. proposed a generic statistical framework called bootstrap analysis of stable clusters (BASC), implemented with K-means, to quantify the stability of RSNs. Because it is often impossible or impractical to repeat fMRI data acquisition for an experiment multiple times, BASC utilized bootstrapping to generate a large number of samples from the existing data and applied a clustering algorithm (K-means in this case) on them, followed by averaging all the adjacency matrices from one subject into an individual stability matrix. At the group level, bootstrapping was applied to the individual stability matrices. To mimic random variations in subject recruitment, B sets of subjects were generated by drawing subjects (with replacement) from the existing (original) group of subjects, with each set having the same number of subjects as the original. Then within each new set of subjects, individual stability matrices from the subjects in this set were averaged to create an average stability matrix. Next a clustering algorithm was applied to each average stability matrix to generate a group-level adjacency matrix. All group-level adjacency matrices were then averaged to create

a group stability matrix. Finally, a clustering algorithm was run on the group stability matrix to derive the stable clusters. In general, BASC is a generic method to investigate the stability of any random clustering process. A core part of the method is resampling new simulated data based on the existing data. Note that in Bellec's study, the time-series were mean time-series from ROIs rather than from voxels and the Monte-Carlo simulation is known to be computationally expensive, especially when the scale of simulated data is large (e.g. voxel-wise time-series). Nevertheless, the BASC is a good attempt to address the stability of clustering results. Later, based on the results from aforementioned BASC, Bellec proposed a method to select a limited number of scales (i.e. the number of clusters K) that are representative of the full hierarchy of the clustering results with different choices of K. The estimation of the result quality with a certain scale was done through interpolating the results with existing scales. The author stated that the mathematical criterion used to select the number of scales is heavily influenced by the size of the clusters and whether the multiscale clustering is actually hierarchical or not. And clustering algorithms are likely to affect the selection more than the data itself. Elucidating these questions is an important future research direction.

The BASC together with the scale selection were used further to investigate whether the rich variety of responses observed, associated with task performance, across the brain is functionally meaningful and consistent across individuals (Orban et al., 2015). Thus the task-evoked fMRI data was used in this study. The task-evoked hemodynamic response at each voxel was estimated by interpolating the time points within the task window and averaging them across all the aligned events. The clustering analysis was done for 957 regions, formed by region growing algorithm (Bellec et al., 2006), rather than voxel-wise to reduce the computational burden. Then the BASC was applied to these region-wise averaged hemodynamic responses to obtain the task-evoked brain network. From the methodological aspect, this is one of the few studies that incorporate consensus clustering strategy into the analysis of paradigm fMRI data.

In addition to the work based on BASC, Kelly et al. studied the functional architecture of insula using a different consensus clustering strategy (Kelly et al., 2012). A distance metric called eta^2 was used to quantify the similarity between all possible pairs of intrinsic functional connectivity (iFC) maps associated with voxels within insula. The eta^2 was defined as follows:

$$eta^2 = 1 - \frac{\sum_{i=1}^{n}[(a_i - m_i)^2 - (b_i - m_i)^2]}{\sum_{i=1}^{n}[(a_i - \bar{M})^2 - (b_i - \bar{M})^2]} \quad (2.5)$$

where a_i and b_i are the iFC values at voxel i in two iFC maps a and b, m_i is the mean value of the two images at voxel i and $\bar{M}$ is the grand mean iFC value across all voxels in both iFC maps. It is claimed that eta^2 takes into account differences in scaling and offset between images, which does not affect spatial correlation value. Spectral clustering together with K-means were applied to eta^2 matrix to generate parcellations. Then adjacency matrix was constructed at individual level and consensus matrix was formed by averaging multiple adjacency matrix from multiple number of clusters K, and multiple data collection sites. Lastly, spectral clustering was applied to generate final cluster solutions. In general, the consensus strategy used here follows the same idea as BASC, which is to combine multiple adjacency matrix to yield a consensus matrix reflecting the agreement that certain two voxels/ROIs belong to the same cluster.

Another recent work about consensus clustering analysis of fMRI data was proposed by Ryali et al. (Ryali et al., 2015). In this work, a consensus clustering evidence accumulation (CC-EAC) framework was developed to combine multiple clustering methods for segmenting brain regions using rsfMRI. Correlation between voxel time-series was used as the distance metric. CC-EAC consists of base clustering algorithm (e.g. K-means, spectral clustering, hierarchical clustering) and consensus clustering. Several combinations of base clustering algorithm and algorithm in consensus clustering were used such as K-means with hierarchical (Bellec et al., 2010), K-means with spectral clustering, and spectral clustering with spectral clustering (Kelly et al., 2012). To determine the optimal number of clusters K, five object criteria, namely normalised mutual information (NMI), variation of information (VI), Rand index (RI), probabilistic Rand index (PRI), and a modified silhouette used in (Bellec et al., 2010), were used. Depending on the definition of these criteria, the K that yielding the optimal average criterion between all pairs of subjects was considered the optimal. Based on the results, a combination of K-means and consensus hierarchical clustering, together with the probabilistic Rand index and modified silhouette, is effective in uncovering the optimal number of stable clusters from simulated and experimental rsfMRI datasets.

Note that all the aforementioned works have only used one clustering algorithm and combined the results from multiple samplings, data modalities, parameters, or collection sites in their consensus strategy. In fact, repeating a single clustering algorithm and combining the results will reduce the random effects from factors such as different parameters. However, it will not change the underlying assumption of a certain clustering algorithm, which is likely to introduce an algorithm bias towards the clustering results. Nevertheless, in Ryali's work, the authors realised the potential and flexibility of consensus clustering and further proposed an idea in the future work section, which is combining complementary clustering methods as opposed to using just one clustering method.

2.4 Summary of current issues and challenges in clustering analysis of fMRI data

1. Individual clustering algorithm often requires some pre-defined parameters to execute, such as the popular K-means, fuzzy C-means, and SOM. The accuracy and efficiency of the clustering results largely depend on the choice of these parameters. From the methodological side, users often use some heuristic approaches to estimate parameters like the number of clusters K. Alternatively, clustering can be run with a range of parameters first and then each solution will be validated and compared with each other, followed by choosing the best solution based on the validation results. This is possible when the size of data is relatively small, but in the context of neuroimaging data whose size keeps increasing due to the higher spatial and temporal resolution as well as more and more participants recruited in a single fMRI experiment, the brutal exhaustive search becomes more and more impractical.
2. A large number of clustering algorithms have been proposed to address various problems. Each of these algorithms has some underlying assumptions such as the data structure, the way objects connect with each other and so on. Traditionally, one can use the discriminative approach such as model selection procedure to rank the competing algorithms, like the parameters selection in single clustering algorithm, from a series of intensive computational simulations. Or one can employ the generative approach to model the data generation process and create a model that suits the data best. However, due to the

lack of ground truth in clustering problems and the large scale of datasets, the discriminative approach become increasingly difficult. The generative approach provides a feasible way of modelling the data structure when the fMRI experiment design is relatively simple but when the data generation process is complicated such as the two fMRI experiments used in this thesis, it is also difficult to employ.

3. The aforementioned two problems and challenges are more or less related to the large scale of datasets to be analysed. In this era filled with big data in many areas such as the neuroimaging field, the development of computational methods needs to consider not only the accuracy and validity but also the capability of dealing with large-scale datasets regarding the available computational resources. To achieve this goal, both the hardware platform and algorithms themselves need to cope with each other when a particular analysis strategy is developed. In other words, in addition to the smart design of the analysis paradigm or framework, the included methods also need to be fit in the high performance computing technique rather than indiscriminately feeding everything into a high performance computer and let the hardware-side deal with the rest.

Chapter 3 Methodology

This chapter details the methods and techniques in the consensus clustering analysis framework that have been adopted to produce the results presented and discussed in the subsequent chapters. Firstly, the details of fMRI data collection are described, followed by the introduction of the preprocessing of the raw data and the preparations of the data for clustering analysis. Then the clustering paradigm is described, including single clustering algorithm and integration of multiple clustering results for consensus analysis. The last part of this chapter details the usage of grid computing technique for reducing the total execution time as well as two statistical tests for analysing strong BOLD responses elicited by stimuli and response shapes respectively.

3.1 fMRI data collection and preprocessing

3.1.1 Data collection

The data collection was carried out by the research staff and students at Aalto University and University of Helsinki in Finland. There are two fMRI studies inspecting the different aspects of brain responses to music. In this section, only the common data collection procedure is introduced and the experiment paradigm for each fMRI experiment will be introduced later in the chapters that report the procedure of the fMRI data analysis and the results.

There are two MRI scanners used to collect the data. One was conducted with a 3 Tesla scanner (3.0T Signa VH/I General Electric) in the advanced magnetic imaging (AMI) Centre at Aalto. Another one was performed using a 3T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) and a standard 20-channel head-neck coil, also at the AMI Centre. Different scanning parameter settings were used on these two scanners based on their performances to get high quality fMRI images. The detailed parameters (e.g. field of view, oblique slice thickness, and spacing) will be specified in the chapters later. Despite of the different settings, all the data collection procedures strictly conform to the neuroscientific standard and were carried out by experienced researchers and technicians to ensure the quality of the data. For example, in order to eliminate the static magnetic field interruption from the magnetic coil located

in the normal earphones, the sound signal (music) was sent to the participants via an fMRI compatible earphone. In addition, the fMRI studies were approved by the ethical committee of the Helsinki University Central Hospital and complied with the Helsinki Declaration.

3.1.2 Preprocessing

Functional MRI scans were preprocessed on a MATLAB platform using SPM8 (Statistical Parametric Mapping), VBM for SPM (Voxel Based Morphometry; Wellcome Department of Imaging Neuroscience, London, UK), and customized scripts developed by the researchers and technicians in Finland. 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 grey matter, white matter, and cerebrospinal fluid by using VBM, and registered to the corresponding segmented high-resolution T1-weighted structural images. These were in turn normalized to the MNI 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 8 mm full-width-at-half-maximum Gaussian filter.

Note that the raw fMRI data preprocessing steps follow a commonly accepted procedure. Although one cannot guarantee the noise and misalignment are completely removed by these procedures, no perfect preprocessing pipeline exists. And the work of preprocessing performance

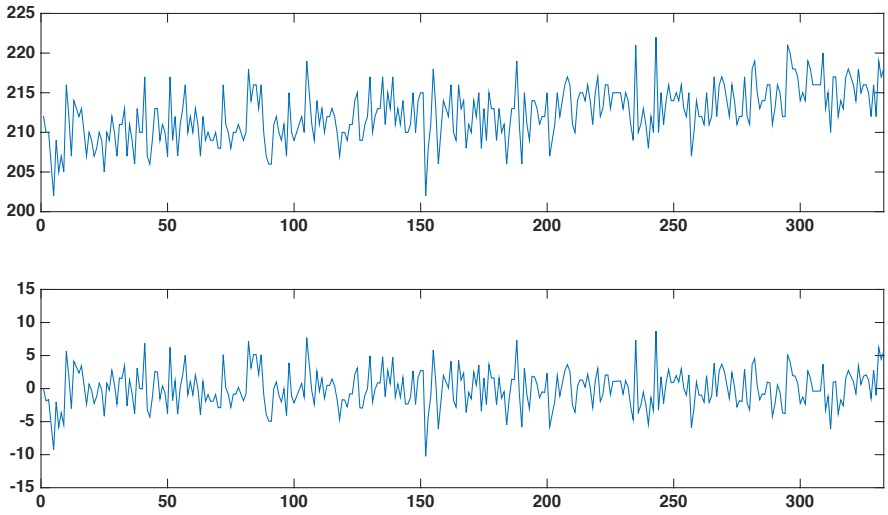


Figure 3-1. Demonstration of high-pass filtering of the BOLD signal. The top one is the original time-series and the bottom one is the filtered time-series.

improvement is out of the scope of the studies in this thesis. Nevertheless, the processed data have been used in other published works using classical hypothesis-driven approaches.

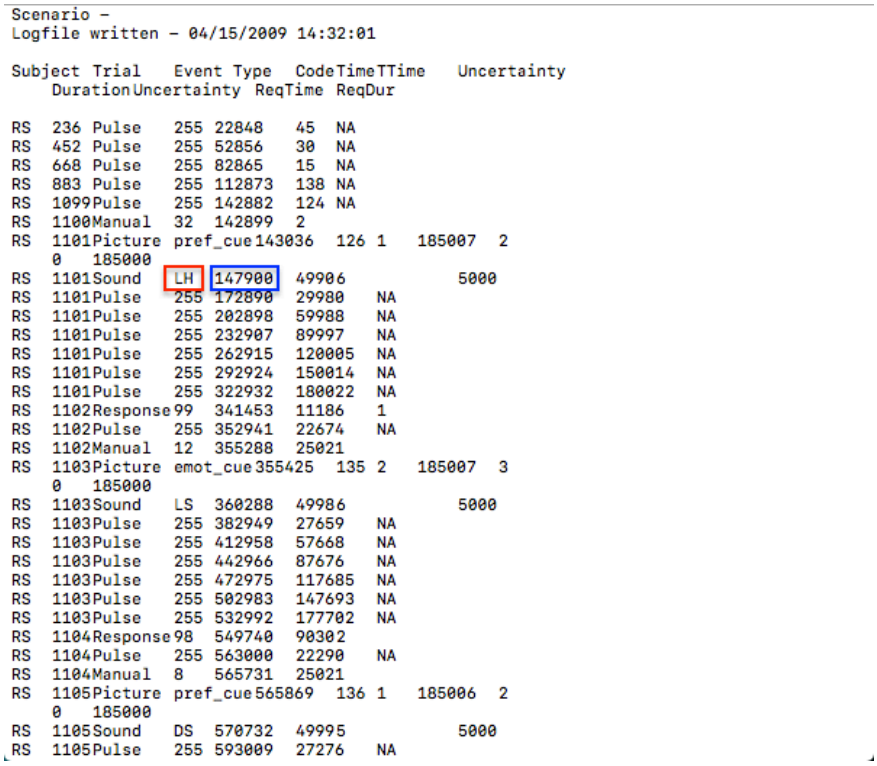
3.1.3 Preparation of data for clustering analysis

3.1.3.1 Scanner drift removal

The 3D volume data was converted to a vector (228453×1) by using a standard brain mask. The above step was applied to every 3D volume scan from each subject and all the scans (along the time domain) of each subject were combined sequentially, forming the fMRI time series of each subject. During the scanning, the baseline signal may change over time, i.e. the general BOLD magnitude measured at a certain time is different from the general magnitude at another time. This phenomenon is called scanner drift. Many sources can lead to the scanner drift, including the common cause of the change in the resonant frequency of hydrogen protons associated with subtle changes in the strength of the static field (system noise). Even though it is powered by superconducting currents, the main magnet still experiences minute drifts in stability over time. In addition, the physiological noise is also thought to contribute to the low frequency drift in the voxel time-series. Since the signal changes from low frequency drift is not caused by the experiment stimulus, it would add non-relevant signal to the real BOLD signal changes caused by the experiment stimulus, thus the accuracy of either hypothesis-driven or data-driven analysis would be undermined. Figure 3-1 demonstrates the effect of the low frequency drift removal. Note in GLM approach, low frequency drift can be modelled as confounding effects during statistical data analysis. But in data-driven analysis such as clustering, the drift has to be removed before further data processing. In this thesis, a high-pass filter with a cut-off frequency 1/120 Hz is used to remove the low-frequency BOLD signal drift (Brattico et al., 2011).

3.1.3.2 Extracting paradigm time-series

During the whole fMRI scanning session, not every image is related to the experiment paradigm. For example, the participant could have a short rest between two tasks. The images sampled at these non-task time windows should be excluded for clustering analysis as they do not contain the information of brain activities from the tasks the subjects perform.



```
Scenario -
Logfile written - 04/15/2009 14:32:01

Subject Trial   Event Type  CodeTimeTTime   Uncertainty
    DurationUncertainty ReqTime ReqDur

RS  236 Pulse   255 22848   45  NA
RS  452 Pulse   255 52856   30  NA
RS  668 Pulse   255 82865   15  NA
RS  883 Pulse   255 112873  138 NA
RS  1099Pulse   255 142882  124 NA
RS  1100Manual  32  142899  2
RS  1101Picture pref_cue143036  126 1   185007  2
    0   185000
RS  1101Sound   LH  147900  49906           5000
RS  1101Pulse   255 172890  29980   NA
RS  1101Pulse   255 202898  59988   NA
RS  1101Pulse   255 232907  89997   NA
RS  1101Pulse   255 262915  120005  NA
RS  1101Pulse   255 292924  150014  NA
RS  1101Pulse   255 322932  180022  NA
RS  1102Response99  341453  11186   1
RS  1102Pulse   255 352941  22674   NA
RS  1102Manual  12  355288  25021
RS  1103Picture emot_cue355425  135 2   185007  3
    0   185000
RS  1103Sound   LS  360288  49986           5000
RS  1103Pulse   255 382949  27659   NA
RS  1103Pulse   255 412958  57668   NA
RS  1103Pulse   255 442966  87676   NA
RS  1103Pulse   255 472975  117685  NA
RS  1103Pulse   255 502983  147693  NA
RS  1103Pulse   255 532992  177702  NA
RS  1104Response98  549740  90302
RS  1104Pulse   255 563000  22290   NA
RS  1104Manual  8   565731  25021
RS  1105Picture pref_cue565869  136 1   185006  2
    0   185000
RS  1105Sound   DS  570732  49995           5000
RS  1105Pulse   255 593009  27276   NA
```

Figure 3-2. An example of information recorded in the logfile during fMRI experiment.

When the MRI machine is scanning the brain states of the participants, all the events are recorded in a logfile, including stimulus onset time, timing of each pulse, and the responses of subjects (e.g. code indicating left or right button pressed). Through the information in logfile together with the paradigm design parameters, the fMRI images related to the tasks can be extracted out. Figure 3-2 is an example of a logfile of a random subject whose data is used in this thesis. There are a lot of information in the file but two types of them are particular important for extracting the paradigm related fMRI images. The first one is the stimulus type, which is marked with red rectangle. The second one is the onset time of each stimulus, which is marked with blue rectangle. In this example, 147900 indicates 14.79 seconds after the scanning starts. Another information needed is the duration of each stimulus. Then the set of fMRI images covering a particular stimulus is calculated according to Formula (3.1).

$$image\_range = \left(ceil\left(\frac{onset}{TR}\right) + 1\right) : \left(floor\left(\frac{onset + stimulus\_duration}{TR}\right) + 1\right) \quad (3.1)$$

The $ceil$ and $floor$ functions are used to ensure only the fMRI images within the period of stimulus are included when the onset time is between two adjacent images, which is often the case.

Note that all the TRs are not exactly the same as the logfile shows, however, this is not an issue considering the small differences and the slow hemodynamic response changes.

Then for each subject and each stimulus, the above process is applied to extract the corresponding BOLD time-series. These time-series are grouped according to their categories and used later for clustering analysis.

3.2 Individual clustering algorithms

3.2.1 K-means

The K-means algorithm is arguably the most commonly used partitional clustering algorithm for vector data in the clustering family (Thirion et al., 2014). The goal of K-means is to obtain the partition that minimises the within-cluster dissimilarities for a given number of clusters. Given a set of voxel's time-series $(\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \dots, \boldsymbol{t}_n)$ with each voxel's time-series having d BOLD response values, K-means partitions these time-series into K clusters ($K \leq n$) $\boldsymbol{C} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \dots, \boldsymbol{C}_K\}$ so as to minimize the within-cluster sum of squared-error. The objective function is defined in Equation (3.2).

$$\arg\min_{\boldsymbol{C}} \sum_{i=1}^{K} \sum_{\boldsymbol{t} \in \boldsymbol{C}_i} \|\boldsymbol{t} - \boldsymbol{\mu}_i\|^2 \tag{3.2}$$

where $\boldsymbol{\mu_i}$ is the mean time-series of all the members in cluster $\boldsymbol{C}_i$. The procedure of K-means is described as follows:

Algorithm K-means:

1. Initialise K clusters either randomly or deterministically;

2. Do:

{

a. Assign each time-series to its nearest cluster $\boldsymbol{C}_i$, where $i = \arg\min\|\boldsymbol{t} - \boldsymbol{\mu}_i\|$;

b. Update the centroid of each cluster $\boldsymbol{\mu}_i = \frac{\mathbf{1}}{N_i}\sum_{\boldsymbol{t} \in \boldsymbol{C}_i} \boldsymbol{t}$ after membership reassignment (N_i is the number of members in cluster $\boldsymbol{C}_i$);

}

Until (convergence or the iteration number is larger than the predefined value);

The optimisation problem to find the optimal partition minimising the total squared-error is a non-deterministic polynomial-time hard (NP-hard) problem. According to the standard algorithm, the global optimum is not guaranteed to achieve. Although the standard algorithm is guaranteed to converge at a local optimum. In practice, the number of iterations is often defined manually in advance to stop the iterations.

As described above, the result of iterations depends on the initialisation of K clusters at the first step. Kaufman approach (KA) is a deterministic initialisation method (Kaufman and Rousseeuw, 1990). A work carried out later showed that KA and random partition initialisation perform better than the other initialisation methods (Peña et al., 1999). When using random initialisation, multiple runs with different initialisations and the same other parameters (e.g., number of clusters) is often employed to remove the random effects. Considering the size of data in this thesis, the KA approach is chosen as the initialisation method for K-means. The implementation of K-means is from the $kmeans$ function in Statistics and Machine Learning Toolbox in MATLAB.

3.2.2 Hierarchical clustering

Hierarchical clustering is also one of most popular clustering methods in the literature. In contrast to partitional clustering such as K-means, which attempts to directly divide the dataset into a set of disjoint clusters, a hierarchical clustering method is a procedure of forming a nested partition from a proximity matrix, which can be graphically represented by a tree, called a dendrogram. By cutting the dendrogram at a certain level, the number of clusters K and the corresponding partition is obtained. Cutting the dendrogram at different levels will lead to different clustering results. There are two strategies for hierarchical clustering, namely agglomerative and divisive strategies. In general, the complexity of hierarchical clustering makes it not as fast as K-means for clustering large dataset. In addition, the divisive strategy is even slower than agglomerative approach. Here, the agglomerative strategy is chosen as the base hierarchical clustering algorithm. The procedure of agglomerative hierarchical clustering is described as follows:

Algorithm agglomerative hierarchical clustering:

1. Each individual object/voxel is considered as a cluster. Calculate the initial distance matrix for all clusters;

2. **Do**:

{

a. Merge the two clusters with the minimal distance in the current distance matrix;

b. Calculate the updated distance matrix for the newly merged cluster and the other clusters;

}

Until (all objects/voxels are in one cluster);

In step 2-a, the minimal distance between certain two clusters are used as the criterion to merge these two clusters. There are many different definitions of the distance between clusters, namely single linkage, complete linkage, average linkage and Ward linkage. Ward linkage is generally very efficient compared to other linkage methods and recently its efficiency and accuracy are further discussed (Thirion et al., 2014). In this thesis, the agglomerative hierarchical clustering with Ward linkage is chosen.

Ward linkage is also known as Ward's minimum variance method, originally presented by Ward (Ward, 1963). Similar to K-means, the objective of Ward linkage is to minimise the total within-cluster variance. To achieve this, the minimum distance in step 2-a is defined as the minimum between-cluster variance and the two clusters that lead to the minimum increase in total within-cluster variance after merging will be merged together. Therefore, the distance function between two clusters in Ward linkage is defined as the within-cluster variance by considering them as one cluster as shown in Equation (3.3). The within-cluster variance is also called error sum of squares (ESS).

$$ESS = \sum_{\boldsymbol{t}_i \in \boldsymbol{C}} \|\boldsymbol{t}_i - \boldsymbol{\mu}\|^2 \tag{3.3}$$

After the hierarchy of the dataset is constructed, represented by dendrogram, the number of clusters K is given to cut the dendrogram and form the corresponding partition of the data. The implementation of hierarchical clustering is from the function $linkage$ and $cluster$ in Statistics and Machine Learning Toolbox in MATLAB.

3.2.3 SOM

SOM is one of the most popular clustering algorithms, ideally suited to exploratory data analysis, allowing one to facilitate easy visualisation and interpretation of the clustering results. SOM belongs to the neural network based clustering family, which is generally based on competitive learning (CL) model. The competitive learning is a process that different neurons or processing elements compete on who is allowed to learn to represent the current input. In the case of SOM, neighbouring cells in a neural network compete in their activities by means of mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns. The objective of SOM is to represent high-dimensional input patterns with prototype vectors that can be visualised in a usually two-dimensional lattice structure.

The SOM clustering algorithm starts with a pre-defined neuron grid whose shapes can be rectangular, hexagonal, circular, and so on. Assume that are K neurons in the grid, each neuron has a weight vector denoted by $\boldsymbol{w}_i$, $i = 1{:}K$ that is initialised randomly. The input data is denoted by $\boldsymbol{T} = \{\boldsymbol{t}_i | i = 1,2,\dots,N\}$. In the case of clustering fMRI data, the vector $\boldsymbol{t}_i$ can be voxel time-series or other computed features of the voxels. The Euclidean distances between input data $\boldsymbol{X}$ and weight vectors $\boldsymbol{w}$ are calculated, and then, for each input data vector $\boldsymbol{t}_i$, the best match unit (BMU) that minimises the Euclidean distance between $\boldsymbol{t}_i$ and weight vectors is found. BMU is expressed in Equation (3.4).

$$BMU = \arg\min_{\boldsymbol{j}} \left(\boldsymbol{t}_i - \boldsymbol{w}_j\right) \tag{3.4}$$

After the BMU for each input vector is found, the weight vector of corresponding BMU is updated. The winner-take-all (WTA) paradigm only allow BMU to update its weight vector. Virtually, SOM employs the WTM paradigm, where at each learning step all neurons within a neighbourhood set (a set of neurons) around BMU can also be updated. The width or radius of this neighbourhood set is time-varying: it shrinks monotonically with time and ends with only BMU in the set. Let N_s denote the neighbourhood set. The basic weight vector updating process in time t is expressed in Equation (3.5).

$$\boldsymbol{w}_i(t+1) = \begin{cases} \boldsymbol{w}_i(t) + h(t)\left[\boldsymbol{t}_j - \boldsymbol{w}_i(t)\right] & if\ i \in N_s(t) \\ \boldsymbol{w}_i(t) & if\ i \in N_s(t) \end{cases} \tag{3.5}$$

where $h(t)$ is the neighbourhood function defined in Equation (3.6).

$$h(t) = \alpha(t)\exp\left(-\frac{\boldsymbol{r}_{BMU} - \boldsymbol{r}_i^2}{2\sigma^2(t)}\right) \tag{3.6}$$

where $\alpha(t)$ is the monotonically decreasing learning rate, $\boldsymbol{r}$ represents the position of corresponding neuron, and $\sigma(t)$ is the monotonically decreasing kernel width function. The procedure of SOM is summarised as follows:

Algorithm SOM:

1. Initialise the topology of SOM and each node's weights $\boldsymbol{w}_i$, $i = 1{:}K$, randomly;

2. **Do:**

{

a. Find the BMU by calculating the distance between the input vector and the weights of each node, as shown in Equation (3.4);

b. The radius of the neighbourhood around the BMU is calculated. The size of the neighbourhood decreases with each iteration according to Equation (3.6);

c. Each node in the BMU's neighbourhood has its weights updated, according to Equation (3.5), to become more like the BMU. Nodes closest to the BMU are altered more than the nodes furthest away in the neighbourhood;

}

Until (convergence or iteration number is larger than the predefined value);

The implementation of SOM is done by combining several functions in neural network toolbox in MATLAB R 2013b, namely $newsom$, which initialises a SOM network, $train$, which trains the SOM network, and sim, which performs the clustering using the trained network. The input data to the $train$ function are the same as those to the sim function for clustering purposes.

3.3 Consensus clustering for robustness and stability

3.3.1 Individual partition generation

The individual partitions are from the clustering results of the excerpt fMRI time-series recorded when certain type of stimuli are presented. For example, if M functional images are acquired when a certain music excerpt is presented and the number of voxels covering the whole brain is N, then

the excerpt fMRI time-series is a $N \times M$ matrix with its rows representing voxels and columns representing the time profiles of each voxel. Figure 3-3 is the demonstration of 50 random voxels' time-series during a random stimulus. These time-series are used to define the activities of voxels, based on which the voxels sharing a similar activity pattern are grouped together by clustering algorithms.

For each aforementioned excerpt fMRI time-series, a clustering algorithm with a parameter setting is applied. The clustering result is denoted by a $K \times N$ matrix $\boldsymbol{U}$ called partition matrix, where K is the number of clusters and N is the number of voxels in total. The elements of partition matrix represent the membership of each voxel. Each combination of clustering algorithm and parameter setting will yield a partition matrix for one excerpt fMRI time-series. In total, there are $\sum_i Para_i$ partitions generated for each excerpt time-series, where $Para_i$ is one combination of clustering algorithm and parameter used.

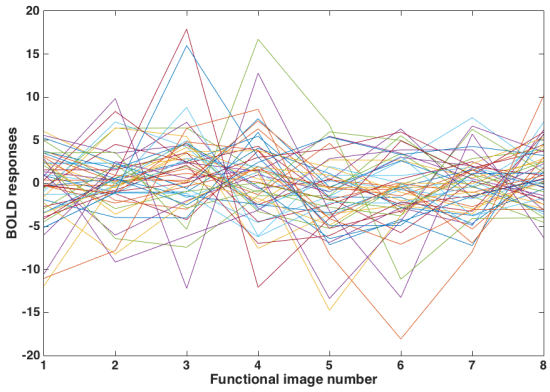


Figure 3-3. Demonstration of 50 voxel' time-series during a randomly selected stimulus. The horizontal axis indicates the number of functional images that are sampled and the vertical axis indicates the response strength at different time.

In the case of crisp clustering where each object belongs exclusively to one cluster, the partition matrix $\boldsymbol{U}$ is a binary matrix with the value 1 at position (k, n) representing the n-th voxel belongs to the k-th cluster. In the general case of fuzzy clustering, the elements can have any value between 0 and 1 inclusively. In general, a partition matrix $\boldsymbol{U}_{K \times N}$ fulfils the following three conditions.

$$\boldsymbol{U}_{k \times n} \in [0,1], \qquad \forall k, \forall n \tag{3.7}$$

$$\sum_{k=1}^{K} \boldsymbol{U}_{k\times n} = 1, \qquad \forall n \tag{3.8}$$

$$0 < \sum_{n=1}^{N} \boldsymbol{U}_{k\times n} < N, \qquad \forall k \tag{3.9}$$

Equation (3.8) necessitates that the total membership of any given voxel in all of the clusters should be 1 because the fuzzy membership values represent the probability of the voxel' belongingness to different clusters.

3.3.2 Partition matrices relabelling

Clustering algorithm is unsupervised, thus the order of the cluster label is not guaranteed to be aligned. For example, the k-th cluster in a given partition may not correspond to the k-th cluster in other partitions. The summation of unaligned partitions will mix the irrelevant clusters together, making the whole following procedure meaningless. Therefore, before the next stage, it is essential to reorder the clusters in all of the partitions such that they are aligned. After the alignment, the k-th cluster in a given relabelled partition corresponds to the k-th cluster in each one of the other partitions.

Depending on the specific applications, the priorities of relabelling may differ. In some applications, all of the resulting clusters are of interest for investigation, and the priority in this case is to optimise the overall relabelling accuracy. On the other hand, many applications only need the few focused high-quality clusters while ignoring the rest of the clusters. Two relabelling technique called min-min and min-max were proposed by Abu-jamous et al., which facilitate the pipeline used later. Min-max approach is for overall relabelling accuracy (Abu-Jamous et al., 2013) and min-min approach is for obtaining focused high-quality clusters (Abu-Jamous et al., 2015a). When carrying out the connectivity analysis in this thesis, the main objective is to identify the brain structures that consistently having highly correlated BOLD responses across different stimuli types rather than partitioning the whole brain into many small regions (brain parcellation). So the focused high-quality clusters are more interesting to investigate than the overall parcellation of the brain. Consider these research objectives, the min-min approach is chosen as the relabelling method for aligning the large amount of individual excerpt partition matrices.

Relabelling a partition matrix $\boldsymbol{U}$ to be aligned with a reference partition matrix $\boldsymbol{U}^{ref}$ aims at finding a matrix $\widehat{\boldsymbol{U}}$ which represents one of the permutations of the rows of $\boldsymbol{U}$ such that its similarity to $\boldsymbol{U}^{ref}$ is maximised. Finding the optimal labelling correspondence is an NP-complete combinatorial problem (Ayad and Kamel, 2010), which has a search space of $K!$, making exhaustive search impractical for not so large value of K. Thus the heuristic algorithms, e.g., min-min method, are often used to approximate the solution. The min-min relabelling method consists of four steps:

Step 1. A dissimilarity matrix $\boldsymbol{S}_{K\times K}$ is constructed with pairwise Euclidian distances between the rows (clusters) of the matrix $\boldsymbol{U}$ and the rows of the reference matrix $\boldsymbol{U}^{ref}$.

Step 2. The minimum value in each column of $\boldsymbol{S}$ is found, followed by identifying the minimum value of these minima. Then the rows (clusters) from $\boldsymbol{U}$ and $\boldsymbol{U}^{ref}$ which correspond to this dissimilarity value are mapped to each other, i.e., these two clusters are considered to have the same cluster label.

Step 3. The row and the column that intersect at the aforementioned dissimilarity value are deleted from the similarity matrix $\boldsymbol{S}$. Then the minimum of the column's minima in the reduced matrix is further identified leading to mapping next pair of clusters from $\boldsymbol{U}$ and $\boldsymbol{U}^{ref}$.

Step 4. Repeat step 2 and 3 until all rows (clusters) from $\boldsymbol{U}$ are mapped to their corresponding rows (clusters) in $\boldsymbol{U}^{ref}$.

In summary, after the relabelling of any two partition matrices, the rows of the two partition matrices indicate the same cluster label. So they can be further manipulated (e.g., summation or subtraction) to yield more information from the clustering results. Next section will introduce how the relabelling method is used to generate the consensus partition matrix from integrating multiple individual clustering partition matrices.

3.3.3 Fuzzy consensus partition matrix generation

3.3.3.1 Merging multiple partition matrices into consensus partition matrix

A fuzzy consensus partition matrix (CoPaM) is a matrix from combining many clustering partition matrices (clustering results). The membership of each voxel in the CoPaM is derived from its memberships in different partition matrices. In the last section, the relabelling method is introduced. To combine R partition matrices $\{\boldsymbol{U}^1, \boldsymbol{U}^2, \dots, \boldsymbol{U}^R\}$ into a single CoPaM, all these R partition

matrices need to be aligned to a reference partition matrix first. There are several strategies of relabelling and combining. For example, one can randomly choose one partition matrix as the reference and relabel the others to this reference matrix, following by generating the CoPaM from the R partition matrices $\{\boldsymbol{U}^1, \boldsymbol{U}^2, \dots, \boldsymbol{U}^R\}$. This approach has a risk of choosing a poor partition matrix as the reference at the first step, undermining the effects of relabelling as all the other partition matrices are relabelled to a poor clustering result. One can also use an intermediate fuzzy CoPaM $\boldsymbol{U}^{inter(i)}$ that is initialised with the first partition $\boldsymbol{U}^1$, and then the rest partitions are relabelled and fused with this intermediate matrix one by one while considering $\boldsymbol{U}^{inter(i)}$ as the reference at each step. Let the function $Relabel(\boldsymbol{U}, \boldsymbol{U}^{ref})$ denote relabelling the partition matrix $\boldsymbol{U}$ by considering $\boldsymbol{U}^{ref}$ as the reference partition, $\widehat{\boldsymbol{U}}^r$ be the relabelled partition matrix of the partition $\boldsymbol{U}^r$, and $\boldsymbol{U}^{inter(k)}$ be the intermediate CoPaM, which is the reference partition matrix for the next partition to be relabelled, after relabelling and fusing the first k partitions $\{\boldsymbol{U}^1, \boldsymbol{U}^2, \dots, \boldsymbol{U}^k\}$. The procedure of generating fuzzy CoPaM is described below:

1. $\boldsymbol{U}^{inter(i)} = \boldsymbol{U}^1$
2. for $r = 2 : R$
   a. $\widehat{\boldsymbol{U}}^r = Relabel(\boldsymbol{U}^r, \widehat{\boldsymbol{U}}^{inter(r-1)})$
   b. $\widehat{\boldsymbol{U}}^{inter(r)} = \frac{1}{r}\widehat{\boldsymbol{U}}^r + \frac{r-1}{r}\boldsymbol{U}^{inter(r-1)}$
3. $\boldsymbol{U}^{final} = \boldsymbol{U}^{inter(R)}$

In summary, the final fuzzy CoPaM $\boldsymbol{U}^{final}$ is generated through the accumulated evolution of intermediate CoPaM that serves as the reference partition matrix for the relabelling of the next partition matrix to be merged. The iteration keeps running until all the partition matrices are merged into a final single fuzzy CoPaM.

3.3.3.2 The order of merging multiple partition matrices

When choosing the first partition matrix $\boldsymbol{U}^1$, it can be the case that $\boldsymbol{U}^1$ is selected randomly from all the partition matrices to be merged. However, it is not guaranteed that the chosen initial reference partition matrix is a good clustering result. Also, the order of the rest partition matrices to be merged into the intermediate CoPaM influences the overall quality of final CoPaM when some of

the clustering results are very noisy. Thus clustering results evaluation is needed to yield a reasonable partition matrices merging procedure. It is generally considered that a good clustering result should consist of clusters having a small intra-cluster variance and large inter-cluster dissimilarity. To achieve this goal, an appropriate number of clusters K needs to be identified to avoid under or over clustering of the data. But intense evaluation of optimal K in the context of large-scale datasets, such as the fMRI data used in this thesis, is often impractical. So a fast evaluation method, incorporating the neuroscientific fact that brain activations often cover relatively large and continuous area, is designed.

The mean squared error (MSE) metric has been used in many studies to evaluate the quality of clusters by quantifying the dispersion within the cluster. Here a modified version of MSE is proposed named cluster-wise MSE (clustMSE). Two factors, cluster size and clustMSE, are considered to evaluate the quality of a certain clustering result. For a partition consisting K clusters $\{C_1, C_2, \dots, C_K\}$, the cluster size, i.e. the number of voxels included, and cluster-wise mean square error are calculated for each cluster C_k. The cluster-wise MSE is defined in Equation (3.10).

$$clustMSE = \frac{1}{M \cdot N_k} \sum_{i}^{N_k} \|\boldsymbol{t}_i - \bar{\boldsymbol{t}}\|^2 \tag{3.10}$$

where M is the number of time point in the time series, N_k is the number of voxels in the cluster C_k, $\boldsymbol{t}_i$ is the normalised BOLD time-series of i-th voxel, and $\bar{\boldsymbol{t}}$ is the mean time-series of all the voxels. Since each cluster has a pair of values (clustMSE, N) that can be seen as a point in a two dimensional space, a coordinate system whose horizontal axis is clustMSE and vertical axis is cluster size is defined on which all the clusters are plotted. In order to make these values comparable among partitions, clustMSE and cluster size are normalised to the range [0, 1] by dividing the biggest clustMSE value and total number of voxels respectively. Figure 3-4 shows a naive example of this coordinate system.

The black square denotes the cluster having a normalised clustMSE value of 0.5 and normalised cluster size 0.4. The blue dot at the top left corner indicates an ideal scenario that the cluster has a very big size (in this case, all of the voxels are included) and zero clustMSE value. The length of the arrow, which can also be considered as the distance, connecting the black square and blue

dot indicates “how far a cluster is from a perfect case”. The shorter the distance, the better the quality of this cluster is.

The general quality of a partition, denoted as meanDist, is obtained by firstly computing all the distances between each cluster and the top left corner, followed by averaging these distances. The mathematical expression is shown in Equation (3.11).

$$meanDist = \frac{1}{K}\sum_{i=1}^{K}\sqrt{(clustMSE_i - 0)^2 + (N_i - 1)^2} \tag{3.11}$$

where $clustMSE_i$ is the normalised cluster-wise mean square error of the i-th cluster and N_i is the normalised size of the i-th cluster. The smaller the meanDist value, the better the general quality of this partition. Suppose there are R partitions to be merged to form the final fuzzy CoPaM, these partitions are firstly ranked according to their meanDist values, then the partition having the smallest meanDist value will be taken as the first reference partition matrix and the rest partitions, which are already ranked in an order, will be merged into the intermediate CoPaM sequentially. This evaluation approach will also be adopted later when addressing the cluster selection.

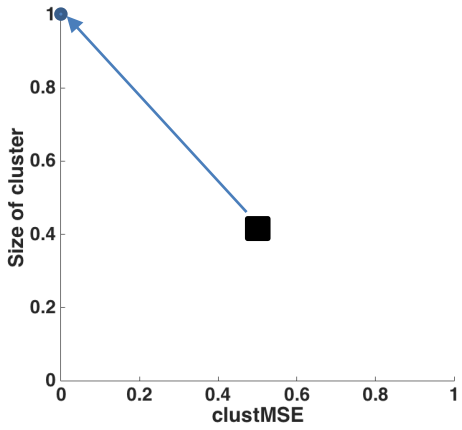


Figure 3-4. Example of scattering clusters on a coordinate system (clustMSE and size of cluster).

3.3.4 Binarisation of the consensus partition matrix (Bi-CoPaM)

The fuzzy consensus partition matrix includes the membership values ranging from 0 to 1 for all of the voxels in each of the K clusters. A membership value of 1 indicates the full belongingness of the given voxel to the corresponding cluster while a membership of 0 means the given voxel is never considered a member in the corresponding cluster. The membership value between 0 and

1 indicates the respective possibilities that the given voxel belongs to different clusters. If all of the individual partitions have always assigned a given voxel to the same cluster, the membership of this voxel in that cluster in the final fuzzy CoPaM will be 1 while being nil in all of the other clusters. However, if some individual partitions have disagreement in assignment of a certain voxel, its membership value is distributed over all of the clusters to which some partitions assign this voxel. In summary, the membership value of the voxel to different clusters is set to be proportional with the number of individual partitions which assigned this voxel to those clusters respectively.

It is easy to determine the memberships of voxels having a value 1 in final CoPaM as all the individual partitions consensually assign them to the same clusters. However, when there are disagreements on the membership in some of the individual partitions, the decision of final memberships of voxels is made based on the fuzzy values in the final CoPaM. In brief, a fuzzy membership value of a voxel will be binarised to 1 or 0 to indicate the belongingness to one or more clusters. To achieve this, the easiest way is to apply a manually defined threshold and then the fuzzy values above this threshold become 1 and the other values below it become 0. Or one can further simply this procedure by assigning a voxel to the cluster in which this voxel has the largest fuzzy membership value. Here, a tunable binarisation approach is adopted, allowing the different membership assignment with respect to different cluster tightness level.

Let the final binary CoPaM be $\boldsymbol{B}$ that has K rows (clusters) and N columns (voxels). The element in $\boldsymbol{B}$ is denoted by $b_{k,i}$, representing the binary membership of i-th voxel in the k-th cluster. The fuzzy CoPaM before binarisation is $\boldsymbol{U}$ with the elements $u_{k,i}$ ranging from 0 to 1 inclusively. The tunable binarisation approach assigns a voxel to the cluster in which it has its maximum membership value only if this value is far from the closest competitive cluster at least by the value of the tuning parameter δ that has a continuous value from 0 to 1 inclusively; it is not assigned to any of the clusters otherwise. This approach is named difference threshold binarisation (DTB) and the assignment criteria is expressed in Equation (3.12).

$$b_{k,i} = \begin{cases} 1, & u_{k,i} \geq \max u_{j,i} + \delta \,, 1 \leq j \leq K, j \neq k \\ 0, & otherwise \end{cases} \tag{3.12}$$

When the δ is equal to 0, each voxel is assigned to the cluster in which this voxel has the largest fuzzy membership value. When δ increases, some voxels' largest fuzzy membership value cannot exceed the second largest value by δ and those voxels are not assigned to any cluster as a result. In another word, the DTB starts to tighten the clusters. When the maximum value of δ is reached, i.e. 1, only those voxels having a membership value of 1 will be assigned to the corresponding cluster, yielding the tightest consensus clustering result while abandoning many voxels unassigned. At $\delta = 1$, DTB may generate many empty clusters, however, this result would not be trivial if some of the clusters still preserved a considerable number of voxels that covering a relatively large and continuous brain region.

In summary, Bi-CoPaM has an important feature that is tunable. As clusters are tightened, many voxels are unassigned from clusters and are left without being assigned to any other cluster, and smaller and more focused clusters are generated. Some clusters might become completely empty at relatively low δ values while others would resist higher levels of tightening. In general, the role of many of the clusters that become empty at low δ values is to contain and then filter out the majority of the voxels that are irrelevant to the context. On the other hand, the voxels that resist higher δ values are those which have been assigned to the same cluster by higher numbers of individual partitions, and are therefore expected to be more consistently correlated and more relevant to the context.

3.3.5 Evaluation and selection of the final consensus clustering results

After using DTB approach to convert the final fuzzy CoPaM to multiple binary consensus partition matrices with respect to different δ values (e.g. from 0.0 to 1.0 with 0.1 steps) and number of cluster K, clusters with varying levels of tightness are generated. Not only there are a large amount of clusters generated but also they largely vary in size, which significantly affects the validity of known validation techniques rendering them unreliable in this particular context. So a customised cluster evaluation and validation technique is adopted to address this problem. Similar to the partition quality ranking introduced in section 3.3.3.2, the evaluation starts from calculating two values of each cluster generated, which are cluster-wise MSE over all of the datasets, and the number of voxels included in the cluster, or more specifically, the logarithm of this number

(Liu et al., 2017, 2015a). The average MSE per voxel for the k-th cluster can be calculated according to Equation (3.13).

$$MSE_{cluster(k)} = \frac{1}{L \cdot N_k} \sum_{l=1}^{L} \sum_{n \in C_k} \left\| \boldsymbol{t}_n^l - \boldsymbol{u}_n^l \right\|^2 \tag{3.13}$$

where L is the number of datasets, N_k is the number of voxels in the k-th cluster, C_k is the set of voxels in the k-th cluster, $\boldsymbol{t}_n^l$ is the normalised BOLD signal vector of the n-th voxel in the cluster from the l-th dataset. $\boldsymbol{u}_n^l$ is the average normalised BOLD signal vector of the voxels in the k-th cluster from the l-th dataset. The objective it to maximize the number of voxels included in the clusters while minimising the dispersion within the clusters measured by the modified MSE metric. All of the individual clusters that appear in the results are scattered on a 2D plot where the horizontal axis (M) represents the cluster-wise MSE values of the cluster over all of the datasets, and the vertical axis (N) represents the logarithm of the number of voxels in the cluster. Both axes are normalized to have unity length. Figure 3-5 shows an example of M- N scatter plot. The cluster closest to the top left corner (red dot) of the plot is selected as the best cluster (blue dot). This cluster is expected to be large with many voxels (high vertical axis value), yet tight with high correlation (low horizontal axis value). The selected cluster and all of the other clusters that have overlaps with it, even by a single voxel, are removed from the plot. Then, the closest remaining cluster to the top left corner of the plot is selected as the second best distinct cluster. The steps

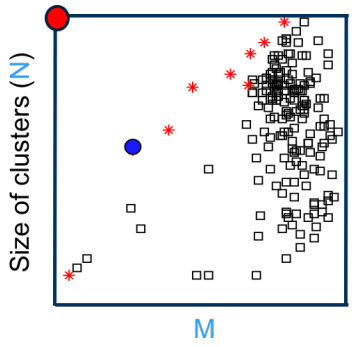


Figure 3-5. Example of M–N scatter plots technique. The horizontal axis is average MSE for a certain cluster across all the datasets. The vertical axis is the normalised size of the cluster.

of selecting clusters and removing those with overlaps with the selected ones are repeated iteratively up to a pre-set maximum number of clusters or earlier when the scatter plots are empty.

The final number of clusters is not pre-determined as it depends on when the plot becomes completely empty. Moreover, the produced clusters are ordered in a descending manner regarding their tightness and size measured by their closeness to the top left corner. Practically, the top selected clusters are of interest to the downstream analysis while most of the low ranked clusters may be considered as containers of irrelevant voxels and are thus discarded.

3.3.6 Topological refinement of raw consensus clustering results.

The topological refinement contains two steps. Firstly, the original clusters selected by M-N scatter plot are filtered by discarding those voxels with weak responses (voxels whose time series have a small variance), since the data used to be clustered are normalised and thus lost the signal magnitude information. In this analysis, the voxels whose variance corresponded to less than half of the mean of the variance for all the voxels from one subject are discarded. After repeating this process for N subjects, N thresholded partitions are obtained. Then if more than 70% percent of the subjects show a strong response at a certain voxel, this voxel is retained for the following analysis. Secondly, the resulting clusters from the previous step are filtered by using the hypergeometric distribution test that discards isolated voxels. Voxels covering large connected brain area would feature a very small p value (normally below 0.001 level) while those covering tiny isolated brain structures would result in a relatively high p value (normally above 0.1 level). We choose p equals to 0.001 to distinguish the large clusters from the isolated voxels. After this step, the clusters only contain voxels having strong BOLD responses as well as covering large and continuous brain regions. The brain regions within each cluster are defined by using Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

3.4 Other methods aiding clustering analysis

3.4.1 Grid computing for clustering large-scale datasets

Grid computing is a computing network consisting of a collection of computer resources from multiple locations. Each computer's resources (e.g., processing powers and storage spaces) are shared with other computers in the system. Different from the traditional high performance computing system where the computers are physically coupled to carry out the same job in parallel, grid computing system consists of loosely-coupled computers to perform large tasks, with non-

interactive workloads that involves a large number of files. Due to this non-interaction feature, grid computing can be more heterogeneous, meaning each worker (computer) can carry out different tasks from the others as they do not exchange task information with each other.

The clustering experiments in this thesis often consist of thousands of single clustering processes on a certain dataset, e.g., an excerpt BOLD response time-series. These large amount of clustering results are to be combined through the consensus clustering paradigm later but when the clustering results are generated, each clustering process does not share voxel's information with other clustering processes. For example, when K-means clustering is running on dataset D_m with number of clusters $K = 20$ and hierarchical clustering is running on dataset D_n with $K = 100$, they carry out the two clustering process independently. This independent clustering process means all the individual clustering can be performed in parallel and fit the grid computing platform perfectly where each worker in the grid computing system can carry out an individual clustering experiment.

To make the MATLAB clustering functions compatible with the grid computing environment, the MATLAB compiler mcc function under version 2013b is used to compile the MATLAB dependent code into standalone application. The machines in grid computing system is installed with MATLAB runtime that can execute compiled MATLAB application or components without installing MATLAB. After the standalone applications are generated, next step is to submit these applications together with the data to be clustered to the grid computing system. London grid computing system that Brunel University London contributes part of the computing resources is used for the large-scale clustering experiment in this thesis. A grid user certificate is applied to allow the jobs to be submitted and executed.

Pion2 workstation in tower D at Brunel University London is used as an interface machine for the London grid computing system. An interface machine is a computer running essential softwares that can allow grid computing users to submit the jobs to the grid system. Pion2 is a powerful workstation equipped with high-end hardware to facilitate the demanding jobs. The specification of Pion2 is as follows.

- A Supermicro 1U SC818G chassis with a X8DTG-DF motherboard
- Dual quad-core Intel Xeon X5550 processors @ 2.67GHz

- 24 GB of 1333 MHz registered DRAM
- 3x1 TB drives (one for system, two RAIDed into one volume for user data)
- 1x C2060 GPU and 1x GTX460 GPU
- Scientific Linux CERN 6 operating system
- glite middleware for grid operations

Pion2 and the worker nodes within Brunel are installed with MATLAB 2013b runtime, enabling the aforementioned MATLAB standalone application to be executed on these worker nodes. The specification of worker nodes in the grid system varies, with some be powerful computers and some be average.

Ideally, the total computation time can be reduced to the longest time a certain individual clustering process takes, where each worker node performs only one single clustering process. However, in practice, the number of the worker nodes installed with MATLAB runtime is limited and these worker nodes are shared among users. Together with the fact that the number of clustering experiments is huge (e.g., more than 20,000), it is almost impossible to have more than 20,000 available worker nodes for the jobs. Thus the job distribution in this thesis is done based on subjects, which means all the clustering experiments for the datasets from one subject will be run on one worker node and cannot be divided further. In this case, if there are around 30 available worker nodes with MATLAB runtime installed, the total computation time will be reduced to 1/30 of that needed on a single computer. Even when there are less than 30 available worker nodes the moment jobs are submitted or some nodes have relatively lower performance, the usage of grid computing still greatly accelerates the data processing speed.

3.4.2 Hypergeometric test

The hypergeometric test uses the hypergeometric distribution to calculate the statistical significance of having drawn specific k successes out of n total draws without replacement from the population whose size is N and has K successes. Equation (3.14) is the formula for calculating the probability.

$$P(X = k|N, K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{3.14}$$

where the bracket is the binomial coefficient (Rice, 2007). Hypergeometric test is often used to identify which sub-population is over or under represented in a sample. Here, the hypergeometric cumulative distribution function (CDF) is used to compute the p value. The hypergeometric CDF is shown in Equation (3.15).

$$P(X \geq k|N,K) = \sum_{i=k}^{n} P(X = i) \tag{3.15}$$

The P in Equation (3.15) is the probability of getting equal or more than k instances of class a if one randomly selects n instances from a population whose size is N and has K class a. For example, this test is used in later chapter to examine the distribution of two different classes of stimulus categories and one class of subject groups, i.e., liked music versus disliked music and happy music versus sad music as well as musician versus nonmusician, compared with their background frequency (N, K in the above formula). We take the null hypothesis to be that different categories or groups have equal effects on the BOLD signal. If a certain stimulus category or subject group is significantly over represented in terms of p value (e.g., $p < 0.005$), we drew the conclusion that this category or group has effect on the BOLD signal in the corresponding condition.

3.4.3 Excerpt BOLD pattern analysis

By utilizing the fact that the clustering experiment is based on the BOLD response shapes corresponding to stimuli, a method for inspecting the differences in response shapes is designed to test the hypothesis that different stimulus categories (e.g., listening to happy music or listening to sad music) or different group of participants (e.g., musician versus nonmusician) would elicit distinct shapes. Once the final clusters are obtained, the time series of the voxels within each final cluster were averaged for each stimulus, which represents the mean time profile for this stimulus within this cluster, based on which we carried out the hypergeometric test on the response shapes. Figure 3-6 illustrates the excerpt data averaging process.

The averaged excerpt data are further clustered into groups with each one having distinct response shapes. Hypergeometric tests are carried out between different groups of participants, different comparison of stimulus type (e.g., liked versus disliked stimuli), or other possible contrasts that would yield interesting information. If in a particular cluster one stimulus category or

participants' group is significantly represented, then this category or group would be declared to tend to have the corresponding response shape.

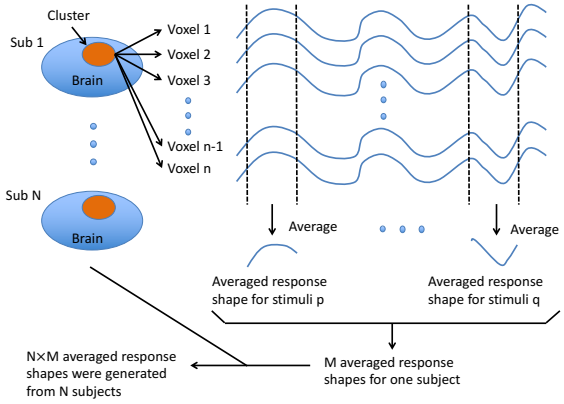


Figure 3-6. Illustration of excerpt BOLD responses shape averaging. For each subject and each excerpt, the time series of the voxels within each final cluster were averaged to obtain the mean time profile for this stimulus within this cluster. Repeating the averaging process on all the data for 29 subjects gives 1856 (= 29 subjects ×64 excerpts) averaged time profiles for each cluster.

3.4.4 "Strong" BOLD responses analysis

In analysing fMRI data, stimuli causing strong BOLD responses are an important and interesting aspect to be investigated. When the BOLD responses levels between different stimulus categories are compared, the categories that tend to elicit stronger BOLD signal response than others can be identified. To achieve this research goal, an analysis pipeline based on the time series of the clustering results is designed. In the case of a particular participant, for one cluster that contains V voxels and for each voxel having a time series of N time points (scans), it gives $V \times N$ time points in total. Each of these data values (amplitudes), corresponding to $V \times N$ time points, is assigned a number from 1 to 4 corresponding to its position in quantile 1 to quantile 4 of all the data values.

After obtaining the quantile data matrix ($V \times N$) for a particular subject, the modes of the quantile values within the time windows covering the duration of each stimulus (e.g. music clips) are examined. If in a certain time window, the mode would be 4, then this would be taken to mean that the "strong" response dominates during this stimulus. Hence, the excerpt categories eliciting the "strong" responses for that particular subject are extracted out, and this is repeated for every cluster for that subject, and then for every other subject and all clusters. To compensate for the

individual variability in the BOLD response strength, i.e. some participants might have generally stronger responses than others no matter what stimulus is given, subjects' response values are coded in the range from 1 to 4, irrespective of the range of the original responses, and without reference to anybody else's responses. Furthermore, for a particular subject, the score of each excerpt was also computed without reference to any other excerpts of any category for this subject. This ensures that the scores of the excerpts are independent subject-wise, category-wise, and excerpt-wise. Afterwards, the distribution differences among all the stimulus categories and the two participant groups are tested to inspect which type of stimuli or subject group tend to elicit stronger responses than others within the same cluster.

3.5 Summary of the consensus clustering analysis framework

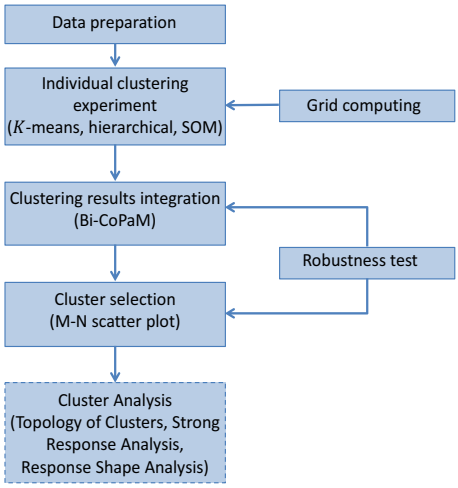


Figure 3-7. The structure of consensus clustering analysis framework for fMRI data.

This chapter systematically introduces the technical details of proposed consensus clustering analysis framework for fMRI data. The main framework includes:

a. Data preparation. This step is for extracting the time-series of voxel corresponding to different stimuli, which are used for clustering experiment.
b. Individual clustering experiment. This step contains the individual clustering experiment with support of high performance computing facility (grid computing).
c. Clustering results integration, selection, and validation. The partition relabelling allows multiple clustering results to be merged together, forming the fuzzy consensus partition matrix. The binarisation of consensus partition with respect to different threshold yields

consensus clustering results with different tightness level. The M-N scatter plot strikes the balance between the intra-cluster similarity and cluster size, which selects the non-trivial clusters. The validation gives the information such as how robust is the framework and whether the consensus clustering solutions are stable and reproducible.

d. Two statistical tests. The excerpt BOLD pattern analysis is for analysing the shape of the voxel's time-series while the "strong" BOLD response analysis is for the amplitude of the voxel's time-series. These two tests are optional in the framework. To carry it or not depends on the fMRI experiment paradigm design and research questions.

The structure of the whole framework is demonstrated in Figure 3-7. This framework will be employed on two real fMRI datasets in the next two chapters.

Chapter 4 Investigation of Functional Brain Connectivity during Affective Processing

This chapter will detail the comprehensive clustering analysis of an fMRI dataset, named Affect, that explores the brain activations and functional connectivity during the affective processing of music excerpts with different preferences and emotions. By obtaining stable and reproducible clusters from a model-free tunable consensus clustering approach and analysing of the spatial and temporal features of these clusters, the neuroscientific insights are found and discussed. This study serves as a validation of a novel consensus clustering analysis of fMRI data as well as an exploration of decoding the brain functional connectivity during affective processing when listening to music, which is an important area within cognitive neuroscience of music. The results show the consensus clustering paradigm has notable advantages over traditional single clustering algorithms in being able to evidence robust connectivity patterns even with complex neuroimaging data involving a variety of stimuli and affective evaluations of them. This work is published in International Journal of Neural System (IJNS) with an impact factor of 6.5 at the time of its acceptance.

4.1 Background of the analysis of fMRI data in affective processing

The 1990s is known as “the decade of the brain” (Jones and Mendell, 1999) during which the neuroimaging research gained great attention with mass funding initiatives, publications in prominent journals as Nature, Science, and lavish attention from the press. In the next decades neuroimaging of humans has gained a position of status within neuroscience, and data-driven approaches and functional connectivity analyses of fMRI data are increasingly favoured to depict the complex architecture of human brains. However, the reliability of these findings is jeopardized by too many analysis methods and sometimes too few samples used, which leads to discord

among researchers. There is currently a vein of scepticism towards neuroimaging findings. Papers making the headlines announced significant brain activity in dead salmon or evidenced magical high correlations between behavioural and brain data (Vul et al., 2009). Although the fascination of brain data can still blur critical thinking in front of a crudely-built mock brain scanner (Ali et al., 2014), scientists are questioning the reliability of neuroimaging and the danger of false positives and reverse inference, hence compromising the relevance of a whole field for the general scientific community (Abbott, 2009). A recent main criticism relies on the wide variety of analysis strategies, combined with small sample sizes, used to investigate regional brain activity measured with fMRI and leading to inconsistent findings (Button et al., 2013). To overcome the limitations of model-based methods, data-driven methods imposed by the researchers, such as clustering, independent component analysis (ICA), seed-based functional connectivity analysis (van den Heuvel and Pol, 2010) and inter-subject correlation analysis (Hasson et al., 2004) are increasingly adopted. However, new algorithms are often proposed, augmenting the discrepancies in the results and the difficulty of choosing the most appropriate data-driven method. Calls for consensus in analysis and meta-analysis methods for neuroimaging have been made (Kriegeskorte et al., 2010).

In research fields characterized by the sheer complexity of the stimulus parameters and the subjectivity of the individual mental state, such as the investigation of musical emotions, the aforementioned difficulties are even more reflected in the wide discrepancy of results. In a recent meta-analysis of 21 fMRI studies on musical emotions, the amygdala (Gosselin et al., 2007), the anterior cingulated cortex, the insula (Damasio et al., 2000), the orbitofrontal cortex (Blood and Zatorre, 2001; Khalfa et al., 2005) and the reward circuit (Mitterschiffthaler et al., 2007) were found to be associated with any musical emotion. While this meta-analysis shows a consistent view, the picture becomes more fragmented when looking at different types of emotions. Very recently, data-driven methods, such as graph theory (Karmonik et al., 2013), eigenvector centrality mapping (Koelsch and Skouras, 2014), network science (Ahmadlou et al., 2014, 2012; Wilkins et al., 2014) and ICA (Cong et al., 2013) have been only marginally adopted to investigate functional connectivity during listening to music. Clustering analysis (Liao, 2005; Shi et al., 2011) has instead not been applied to music neuroscience. In the broader domain of cognitive neuroscience, many methods have been used to address the clustering problem such as K-means (Bello-Orgaz et al.,

2012), hierarchical clustering (Goutte et al., 1999), artificial neural network (Ahmadlou and Hojjat, 2010) based self-organizing maps (SOMs) (Haykin, 1999), graph clustering (Ahmadlou and Adeli, 2011; Bello-Orgaz et al., 2012), and fuzzy clustering (Baumgartner et al., 1998). As introduced in Chapter 2, there is one critical procedure that determines the appropriate algorithm and related parameters such as the number of clusters K. Traditionally, one can either use the discriminative approach such as model selection method to rank the competing algorithms, which are K-means, hierarchical clustering and SOM in this study, based upon some given criteria or use the generative approach to model the data generation process. However, obtaining the most appropriate algorithm and parameters from discriminative approach becomes increasingly difficult due to the lack of ground truth in clustering problems and large dimension of data. The generative approach is also difficult when the data generation process is complicated such as the fMRI paradigms used in this study.

To achieve consensus results in a model-free context and provide a more feasible alternative to generative design, we employ the consensus clustering (Ghaemi et al., 2009) framework to analyse fMRI data. Rather than ranking the different clustering algorithms, Bi-CoPaM in the framework integrates the results generated by multiple clustering methods. Moreover, the results are able to be tuned in terms of the consensus level reflecting the quality of the clusters. In this study, three widely used clustering algorithms in neuroimaging are selected, namely the K-means (Kahnt et al., 2012), hierarchical clustering (Blumensath et al., 2013; Ferrarini et al., 2009) and SOM (Goutte et al., 1999; Liao et al., 2008; Peltier et al., 2003), to be fed into the framework to produce consensus results. The clusters are extracted from many (1856) datasets consisting of fMRI trials associated with each subject's listening to hundreds of emotional music clips and seeks the ones characterized by consistently synchronized fMRI signal changes in most of the datasets. By using the developed framework, we found that several brain structures related to visual, reward, and auditory processing have intrinsic temporal patterns of coherent neuroactivity during affective processing without defining any explicit model.

4.2 Affect fMRI dataset and consensus clustering analysis

4.2.1 Participants

Twenty-nine healthy subjects without any hearing, neurological or psychological problem participated in this study (15 females). Among these twenty-nine participants, thirteen are musicians who possessed formal musical training, on average, for 16.1 ± 6 (SD) years. The others were non-musicians who did not receive any formal musical training but nevertheless had an interest in listening to music consistently. At the time of experiment, musicians reported that they practice their instruments, on average, for 2.2 ± 1.3 (SD) hours per day and non-musicians declared that they listened to music 7.6 ± 5.6 (SD) hours per week. The study is approved by the ethical committee of the Helsinki University Central Hospital and complied with the Helsinki Declaration. The dataset is a subset of a larger data collection, parts of which have been published in (Brattico et al., 2016, 2011; Saarikallio et al., 2012).

4.2.2 Music (Stimuli)

Four stimulus categories were used in the fMRI experiment. These categories consist in music that was classified by subjects to be liked and happy (LH), liked and sad (LS), disliked and happy (DH), and disliked and sad (DS). Four music pieces were chosen for each category, giving 16 pieces in total. Two 18-second long music excerpts with 500ms fade-in and fade-out were selected from each music piece using Adobe Audition and on the basis of a listening test where participants were asked to rate the familiarity of music excerpts with a scale from 1 to 5 with 1 representing unfamiliar and 5 be familiar. This yielded 32 excerpts with 8 in each stimulus category for each participant.

4.2.3 fMRI experiment procedure

The fMRI measurements were conducted with a 3 Tesla scanner (3.0T Signa VH/I General Electric) in the advanced magnetic imaging (AMI) Center at Aalto. Participants rested on the scanner bed in a supine position. Music was presented via fMRI compatible earphone with about 30dB of gradient noise attenuation. Thirty-three oblique slices covering the whole brain (field of view 20mm; 64×64 matrix; thickness 4mm; spacing 0 mm) were acquired using an interleaved gradient

echo-planar imaging (EPI) sequence with TR equal to 3s, echo time 32ms and flip angle 90º, sensitive to BOLD contrast.

During the fMRI experiment, participants listened to the 32 18-s excerpts of music selected as described above. The music excerpts were delivered to the participants via high fidelity MR compatible earphone. Each participant was presented with 32 excerpts for two times in a random order, prompted by a visual cue on the screen (one time it shows like? Dislike?, and another time it shows sad? happy?) to keep the participants concentrating on the emotional aspects of the stimuli. Following the end of the stimuli was a 3-second interval without music stimuli during which another cue asked the participants to answer the questions showed on the screen when they listened to the previous music excerpt by pressing a MR compatible button pads with the second and the third fingers of the left or right hands; After the interval a sinusoidal tone indicated the start of next trial. The scanning session lasted for 23 minutes. After a short break, anatomical T1 weighted MR images (field of view 26 mm; 256×256 matrix; thickness 1 mm; spacing 0 mm) were acquired for about 9 minutes.

4.2.4 Data preprocessing and preparation

The fMRI data preprocessing follows the general procedure described in Chapter 2. In brief, the whole-brain images were preprocessed by statistic parametric mapping 8 (SPM8) and voxel-based morphometry (VBM) for SPM. Each participant's images were segmented, realigned, spatially normalized into the Montreal Neurological Institute (MNI) template and spatially smoothed by Gaussian filter with an FWHM of 6 mm. For preparing the data for the consensus clustering

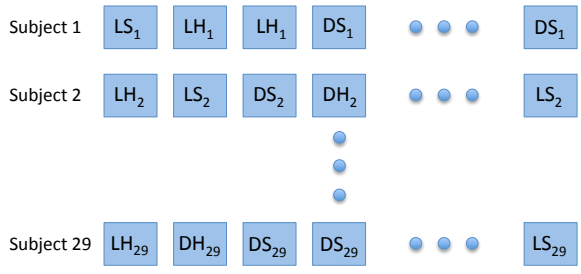


Figure 4-1. Structure of music excerpts where DH stands for disliked happy music, DS for disliked sad music, LH for liked happy music, and LS for liked sad music. Note that each category has different music excerpts for different subjects. For example, LS_1 listened by subject 1 is different from LS_2 listened by subject 2. Also, one certain category has different music excerpts for one particular subject. For example, in subject 1, the music representing the first DS_1 is not necessarily the same as the last DS_1.

analysis two steps are applied using the fMRItoolbox (implemented at the University of Jyväskylä in MATLAB environment): vectorization and segmentation. In vectorization step, the 3D volume data was converted to a vector (228453×1) by using a standard brain mask. The above step has been applied to every 3D volume scan from each subject and all the scans were combined sequentially, forming the fMRI time series of each subject. According to the order that musical excerpts were played, the whole fMRI time series were segmented into 64 EPI brain volumes, each containing 6 or 7 time points (covering 18 seconds at a sampling rate of 3 second), and each corresponding to instances when the participants were listening to music clips. Figure 4-1 illustrates the order each music excerpt was presented (partial).

For one excerpt we used a 228,453×6 (or 7) matrix with the row corresponding to the voxels and the column corresponding to the time profile for this excerpt. In total, there are 1856 excerpts for all subjects, resulting in 1856×228,453×6 (or 7) data points that were used in the following clustering analysis.

4.2.5 Clustering experiment

There are totally 1856 excerpt data for all the 29 participants. Each of these excerpt data (normalized to 0 mean and unit variance) is clustered by K-means, Hierarchical and Self Organizing Map (SOM) with the number of clusters K equals to 10, 25, 50, and 100. These clustering results generated by the three algorithms with four different cluster numbers are combined using the Bi-CoPaM paradigm and selected by M-N scatter plot, yielding the consensus clustering results. The raw clusters are filtered using the method introduced in section 3.3.6.

A sensitivity test is carried out with respect to the filtering parameters that are the percentage of the total participants and the percentage of the mean variance. The filtering process is run with different parameter combinations of percentage of total data and percentage of mean variance. The filtered cluster sizes are compared to investigate the filtering impact on the results.

4.2.6 Comparison among multiple clustering algorithms combination and single algorithm

To evaluate the advantage of consensus clustering against single clustering algorithm, the clustering results obtained by the following four experimental scenarios are compared:

1. K-means only

2. Hierarchical clustering only

3. SOM only

4. The combination of all the three methods in 1, 2, and 3.

Another question is that whether and how the number of clustering algorithm used in consensus clustering paradigm affect the final clustering results. To find out the answer, three more method combination scenarios are also used:

5. K-means and hierarchical clustering

6. K-means and SOM

7. Hierarchical clustering and SOM

To quantitatively assess the similarities or differences between any two clustering results, the Jaccard index is used. Jaccard index is defined as $|X \cap Y|/|X \cup Y|$, where X and Y are two clustering results and $\cap$ and $\cup$ are standard set operations (intersection and union). The range of Jaccard index is from 0 to 1, with 0 representing absolutely different and 1 be exactly matched.

4.2.7 Robustness test against individual functional data variability

Each participant is unique in a way that the brain activity to a certain stimulus vary among different individuals, i.e., different set of participants might yield different analysis results. This raises the problem whether the algorithm could give reasonable results regarding the functional data variation among individuals. To test the consensus paradigm's capability of dealing with the individual variability, the robustness test is carried out.

The design of the test is as follows. Two groups of subsets are generated for test purpose. One is created by randomly selecting 75% of the musicians (10 out of 13) and non-musicians (12 out of 16) as well as 75% of the excerpts for each participant, which yields a subset consisting approximately 56% of all the data from the fMRI experiment. The above random selection is repeated 10 times and these 10 subsets form group A. Similarly, a different ratio of 90% of the musicians (12 out of 13) and non-musicians (14 out of 16) as well as 90% of the excerpts for each participant is chosen, when the data are randomly selected. Repeating the selection with new ratio for ten times form group B consisting approximately 80% of all the data. Then Bi-CoPaM paradigm is applied on these subsets and the clustering results are recorded. Dice's coefficient

is then used to quantitatively compare the similarity between any two clustering results. Dice's coefficient is used. Dice's coefficient is defined as $2|X \cap Y|/(|X| + |Y|)$, where $|X|$ and $|Y|$ are the number of voxels included in the two indices, and $|X \cap Y|$ is the number of voxels shared by the two indices. The range of Dice's coefficient is from 0 to 1, with 0 representing absolutely different and 1 be exactly matched.

To evaluate the robustness, two key aspects need to be looked at to determine the performance. One is whether all the subsets of participants can yield the clusters covering the similar interesting brain regions and the other is the indices of these clusters should be as similar as possible. In another word, the more similar among the results from different subsets of participants, the more robust the method is.

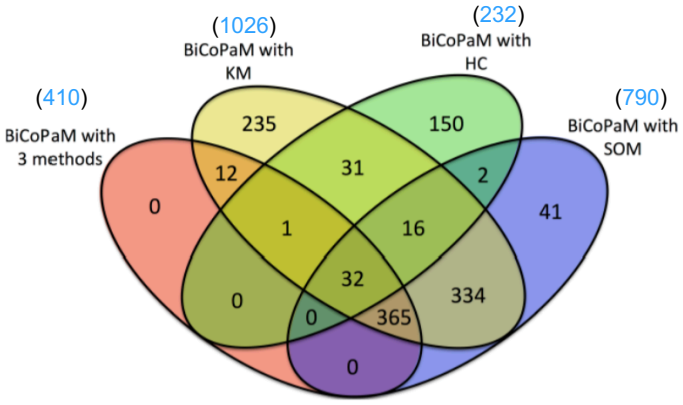
4.2.8 Statistical test of clustering results

The differences in response shapes are investigated to test the hypothesis that different music categories (DH, DS, LH, LS) or different group of participants (musician vs. non-musician) would elicit distinct BOLD responses shapes. Following the procedure introduced in Chapter 2, the 1856 averaged excerpt data are further clustered into groups with each one having distinct response shapes. Hypergeometric tests are then carried out between groups of musicians and non-musicians, liked vs. disliked stimuli, as well as sad vs. happy stimuli. Another test applied is the strong response analysis that searches the musical categories that tend to elicit strong BOLD signal response.

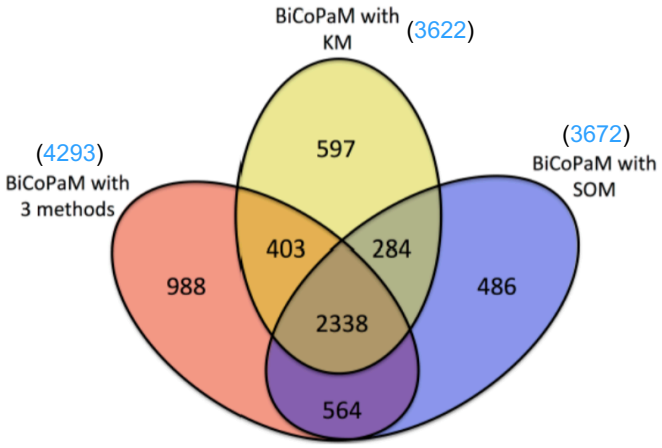
4.3 Results

4.3.1 Advantages of combining multiple clustering results from single algorithm

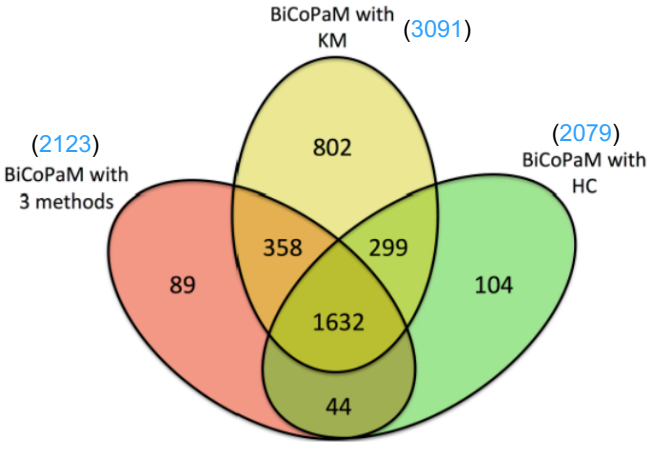
As described in section 4.2.6, seven scenarios with respect to different combinations of clustering algorithms and only with single algorithm are designed to demonstrate the advantages of combining multiple clustering results. The results are reported in two parts with one comparing the clustering results between consensus paradigm and single clustering algorithm and the other comparing the different combinations within consensus paradigm.



(a) Cluster A Visual



(b) Cluster B Reward



(c) Cluster C Auditory

Figure 4-2. The Venn diagram for comparing the set relationship among different clustering experiment settings. Each ellipse represents the clustering results with the value indicating the number of voxels in each set. KM represents K-means, HC represents hierarchical.

4.3.1.1 Multiple versus single

Figure 4-2 shows the relationships among the results obtained by different clustering algorithm settings (KM for K-means, HC for hierarchical clustering) for three clusters covering visual area (cluster A Visual), reward system (cluster B Reward), and auditory system (cluster C Auditory). From the Venn diagrams, all the four experimental scenarios could detect cluster A Visual, although the size and accurate position of the voxels contained in the cluster are different depending on the paradigm used. For cluster B Reward, the hierarchical clustering fails to detect the cluster within subcortical regions related to brain reward system. Bi-CoPaM with all the three methods not only contains the common parts of the results from the other two experiment paradigms, but also included 988 voxels that are not included in the results obtained with either KM or SOM method. Cluster C Auditory is not found in the results by the single SOM algorithm and similarly, the Bi-CoPaM with all the 3 methods successfully find clusters around the bilateral auditory cortex with 89 voxels not included in the results with either KM or HC method. Thus the current findings show that Bi-CoPaM with multiple clustering methods outperforms the results from the single clustering algorithm scenarios, with the fact that it never misses the three important clusters (Visual, Reward, Auditory).

4.3.1.2 Different method combinations

This part reports the results from different method combinations in consensus clustering framework. To make the comparison more complete, the results from the single methods are also included. Table 4-1 shows the Jaccard index between different method combinations for cluster covering visual area. Table 4-2 shows the Jaccard index between different method combinations for cluster covering reward system, and Table 4-3 shows the Jaccard index between different method combinations for cluster covering auditory system. From the three Tables, it is easy to spot the differences between any combinations of two or three clustering algorithms on all three important clusters. Note that for those methods or method combinations having all zeros Jaccard index, it means this method or method combination was not able to detect the corresponding cluster. For example, in Table III, the method combination of K-means and hierarchical clustering (KM & HC) did not group the voxels within the auditory system into a cluster. So, all the members

for clusters covering the visual areas end up producing a value of zero in the Jaccard index between KM & HC and any other method combinations.

Table 4-1. Jaccard index for cluster covering visual area.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.48	0.70	0.70	0.78	0.73	0.40
HC	0.48		0.38	0.57	0.53	0.56	0.26
SOM	0.70	0.38		0.58	0.66	0.62	0.50
KM+HC	0.70	0.57	0.58		0.76	0.84	0.32
KM+SOM	0.78	0.53	0.66	0.76		0.82	0.37
HC+SOM	0.73	0.56	0.62	0.84	0.82		0.35
KM+HC+SOM	0.40	0.26	0.50	0.32	0.37	0.35	

Table 4-2. Jaccard index for cluster covering reward system.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.00	0.56	0.41	0.53	0.52	0.53
HC	0.00		0.00	0.00	0.00	0.00	0.00
SOM	0.56	0.00		0.43	0.57	0.55	0.57
KM+HC	0.41	0.00	0.43		0.33	0.33	0.37
KM+SOM	0.53	0.00	0.57	0.33		0.67	0.61
HC+SOM	0.52	0.00	0.55	0.33	0.67		0.60
KM+HC+SOM	0.53	0.00	0.57	0.37	0.61	0.60	

Table 4-3. Jaccard index for cluster covering auditory system.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.49	0.00	0.00	0.41	0.71	0.62
HC	0.49		0.00	0.00	0.35	0.51	0.56
SOM	0.00	0.00			0.00	0.00	0.00
KM+HC	0.00	0.00			0.00	0.00	0.00
KM+SOM	0.41	0.35	0.00	0.00		0.48	0.40
HC+SOM	0.71	0.51	0.00	0.00	0.48		0.65
KM+HC+SOM	0.62	0.56	0.00	0.00	0.40	0.65	

There are in total 18 final clusters with 7 representing visual areas, 6 representing reward system, and 5 representing auditory system by different single methods and method combinations. To make the differences of these results more straightforward, the 3D rendering of these clusters are produced to reflect the positions of these clusters in the actual brain. To make the visualisation easy to inspect, the 3D transparent glass brain is used as the brain model and the clusters are rendered in red color. Thus the topology of the clusters can be viewed directly. Figure 4-3, Figure 4-4, and Figure 4-5 are the visualisation of the cluster topologies for the clusters covering visual, reward, and auditory system. In summary, the consensus clustering framework with all three methods never fails to detect any of the three important clusters, providing the most complete set of solutions.

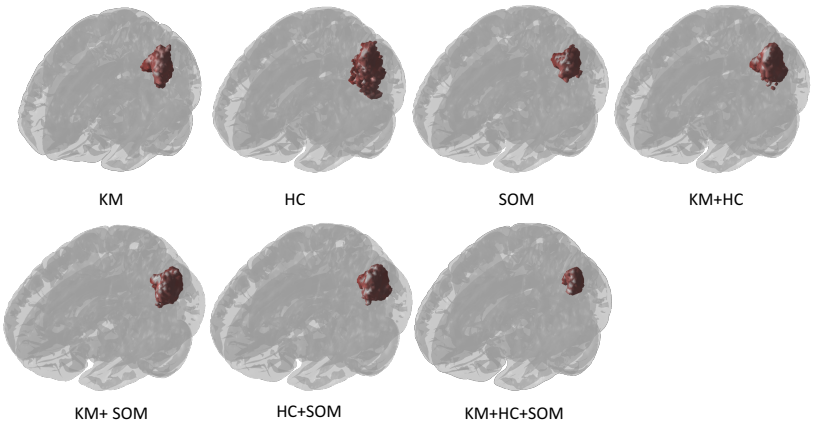


Figure 4-3. 3D illustrations of clusters covering visual area detected by seven different method combinations.

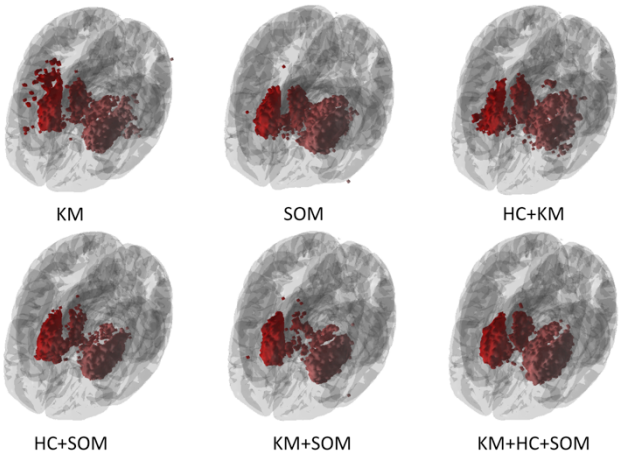


Figure 4-4. 3D illustrations of clusters covering reward system detected by six different method combinations.

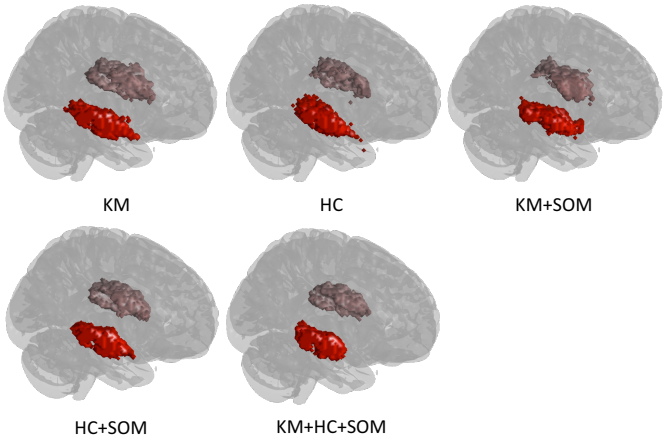


Figure 4-5. 3D illustrations of clusters covering auditory system detected by five different method combinations.

4.3.2 Sensitivity of filtering

As shown in section 4.3.1, the consensus paradigm with all the three clustering algorithms (K-means, hierarchical clustering, and SOM) yields the most complete set of solutions. Therefore, the filtering test is done by incorporating all the three algorithms in the framework. From Table 4-4 to Table 4-6, with the number in each cell representing the size of the cluster, we could verify that as long as the filtering parameter combinations were not extremely strict (bottom right corner of each table), the performance remained very stable. The rectangles in all the tables indicate the parameter combination used in this study. Furthermore, we do not claim that these choices are absolutely optimal, they are rather explorations of a new analysis strategy.

Table 4-4. Filtering results for cluster Visual.

	Percentage of Total Data for Cluster Visual									
	%	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0
Percentage of mean variance	12.5	429	429	429	429	429	429	429	429	429
	25.0	429	429	429	429	429	429	429	429	428
	37.5	429	429	429	429	429	429	429	421	398
	50.0	429	429	429	429	429	421	410	379	277
	62.5	429	429	427	425	415	389	339	249	121
	75.0	429	427	424	410	363	295	200	116	43
	87.5	429	423	396	340	251	153	84	36	7

Table 4-5. Filtering results for cluster Reward.

	Percentage of Total Data for Cluster Reward									
	%	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0
Percentage of mean variance	12.5	4777	4777	4777	4777	4777	4777	4777	4777	4777
	25.0	4777	4777	4777	4777	4777	4777	4777	4777	4776
	37.5	4777	4777	4777	4777	4777	4773	4748	4678	4407
	50.0	4777	4774	4747	4696	4606	4483	4293	4026	3384
	62.5	4763	4639	4491	4300	4132	3865	3510	2973	1894
	75.0	4622	4307	4058	3725	3341	2850	2216	1492	732
	87.5	4330	3870	3342	2747	2219	1550	1049	599	317

Table 4-6. Filtering results for cluster Auditory.

	Percentage of Total Data for Cluster Auditory									
	%	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0
Percentage of mean variance	12.5	3145	3145	3145	3145	3145	3145	3145	3141	3112
	25.0	3145	3145	3141	3137	3125	3096	3053	2962	2644
	37.5	3140	3133	3111	3076	3016	2938	2774	2372	1614
	50.0	3133	3110	3045	2952	2790	2522	2123	1513	767
	62.5	3122	3063	2911	2675	2349	1948	1422	847	308
	75.0	3101	2952	2652	2321	1868	1339	869	438	79
	87.5	3055	2781	2356	1864	1318	877	518	155	5

4.3.3 Robustness of consensus paradigm

Table 4-7 and Table 4-8 illustrate that in most of the trials, the three important clusters (Visual, Reward and Auditory) were always identified despite the different subsets being used. Meanwhile,

when the proportion of data used increased, the results became more stable. For example, the Reward cluster was missed three times in the first test reported in Table 4-7 but was never missed in the second test reported in Table 4-8. On one hand, this proved the framework is robust to variability of participants' data and thus generated reproducible results; on the other hand, it also showed the benefits of using a large number of subjects for more reliable results in data-driven analysis of functional brain imaging data.

Table 4-7. Results of test group A (56% of the full data). Final size is the obtained cluster size by using the whole participants' data. Trial size is the cluster size obtained in different trials. Intersection is the size of the part that the clusters in each trial intersect with the clusters obtained by using the whole participants' data.

Clusters		Trial										
		1	2	3	4	5	6	7	8	9	10	Mean
Visual	Final Size	410	410	410	410	410	410	410	410	410	410	
	Trial Size	551	1031	961	900	924	1838	713	877	2027	651	
	Intersection	357	410	410	410	405	410	396	409	410	393	
	Dice Coeff.	0.74	0.57	0.6	0.63	0.61	0.36	0.71	0.64	0.34	0.74	0.594
Reward	Final Size	4293	4293	4293	4293	4293	4293	4293	4293	4293	4293	
	Trial Size	2857	2591	4646	0	3263	1028	0	1620	0	4168	
	Intersection	1479	2046	3218	0	2278	915	0	1403	0	3303	
	Dice Coeff.	0.41	0.59	0.72	0	0.6	0.34	0	0.47	0	0.78	0.391
Auditory	Final Size	2123	2123	2123	2123	2123	2123	2123	2123	2123	2123	
	Trial Size	2381	3477	1452	2980	1576	2191	1801	2161	856	2006	
	Intersection	1738	2021	16	1995	439	1581	1459	1343	797	1673	
	Dice Coeff.	0.77	0.72	0	0.78	0.24	0.73	0.74	0.63	0.54	0.81	0.596

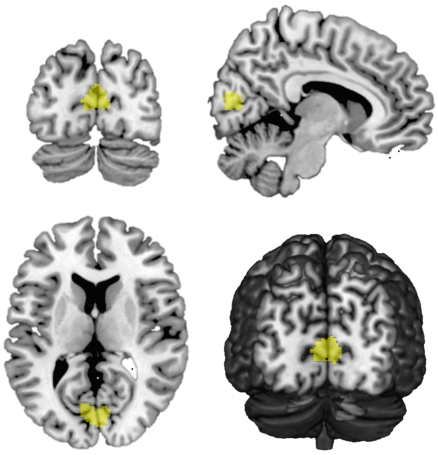
Table 4-8. Results of test group B (80% of the full data).

Clusters		Trial										
		1	2	3	4	5	6	7	8	9	10	Mean
Visual	Final Size	410	410	410	410	410	410	410	410	410	410	
	Trial Size	591	991	385	821	803	1119	784	1216	821	803	
	Intersection	402	410	326	406	406	410	407	410	406	406	
	Dice Coeff.	0.8	0.59	0.82	0.66	0.67	0.54	0.68	0.5	0.66	0.67	0.66
Reward	Final Size	4293	4293	4293	4293	4293	4293	4293	4293	4293	4293	
	Trial Size	3028	2867	3118	4381	1813	4390	4426	2362	4381	1813	
	Intersection	1273	1027	2467	3308	1519	3157	3461	1891	3308	1519	
	Dice Coeff.	0.35	0.29	0.67	0.76	0.5	0.73	0.79	0.57	0.76	0.5	0.592
Auditory	Final Size	2123	2123	2123	2123	2123	2123	2123	2123	2123	2123	
	Trial Size	476	1948	2910	1348	2496	3158	3036	771	1348	2496	
	Intersection	458	1548	1997	338	1855	2051	1704	578	338	1855	
	Dice Coeff.	0.35	0.76	0.79	0.19	0.8	0.78	0.66	0.4	0.19	0.8	0.572

4.3.4 Topology of final clusters

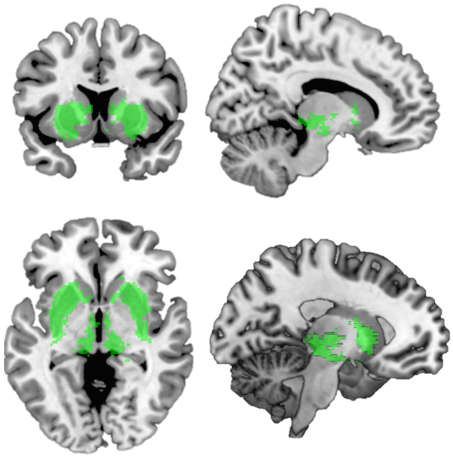
The first 20 clusters (ranked by M-N plots algorithm) are inspected, as these clusters show very strong similarity in the response shapes as well as covered large continuous regions, thus complying with expectations based on knowledge of brain physiology (Huettel et al., 2009). Among these clusters, we find that emotion and sensory-related brain areas are grouped into 3 clusters

Cluster A Visual



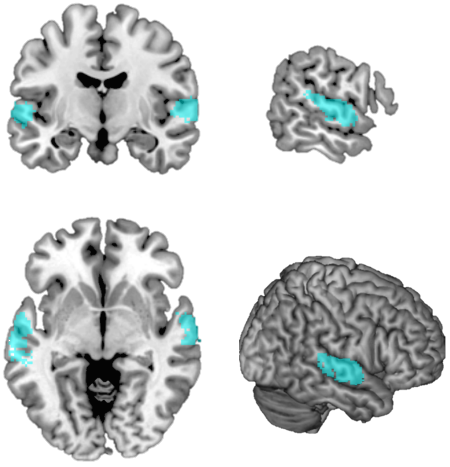
Area	Size (in voxel 2×2×2mm²)	MNI
Calcarine Fissure and surrounding cortex (L)	209	(-3 -77 11)
Calcarine Fissure and surrounding cortex (R)	95	(7 -78 11)
Cuneus (L)	70	(0 -80 19)
Cuneus (R)	36	(6 -81 18)

Cluster B Reward



Area	Size (in voxel)	MNI
Putamen (L)	830	(-25 3 0)
Putamen (R)	881	(28 4 0)
Globus pallidus (L)	102	(-20 0 0)
Globus pallidus (R)	111	(22 0 1)
Thalamus (L)	168	(-11 -19 1)
Thalamus (R)	176	(14 -21 1)
Caudate nucleus (L)	116	(-12 13 3)
Caudate nucleus (R)	116	(16 16 3)
Amygdala (L)	50	(-25 -2 -15)
Amygdala (R)	38	(28 -2 -14)
Insula (L)	121	(-33 3 -2)
Olfactory cortex (L)	34	(-20 7 -15)
Olfactory cortex (R)	17	(23 10 -14)

Cluster C Auditory



Area	Size (in voxel)	MNI
Middle temporal gyrus (L)	517	(-61 -28 0)
Superior temporal gyrus (L)	572	(-59 -16 4)
Superior temporal gyrus (R)	947	(61 -17 5)
Heschl's gyrus (L)	10	(-58 -12 8)
Heschl's gyrus (R)	42	(56 -9 6)
Rolandic operculum (R)	35	(60 -9 9)

Figure 4-6. The 3D illustrations of clusters and the size of each sub-cluster with voxels falling within a known anatomical brain structure, identified with the AAL atlas.

separately (Figure 4-6) which correspond well with the literature studying music emotions with the model-based approach (Brattico et al., 2011). Cluster A comprises bilateral visual areas, namely the calcarine fissure and the cuneus. Cluster B comprises bilateral neural structures of the reward system, namely the ventral striatum - extending to the globus pallidus - the thalamus, the amygdala, the orbitofrontal cortex, and the left insula. Cluster C comprises the auditory areas, namely the bilateral superior temporal gyrus, Heschl's gyrus, the left middle temporal gyrus, as well as one region of the somatosensory cortex, namely the right rolandic operculum.

4.3.5 Distinct temporal features elicited by non-musicians

The response shape analysis is carried out in the 3 final clusters Visual, Reward and Auditory with respect to the contrasts between experimental conditions and groups: liked vs. disliked, happy vs. sad, musician vs. non-musician. In cluster A Visual, we found the response difference in musician vs. non-musician group as shown in Figure 4-7. For the response shape (initially reduces and then steadily rises till the end of the stimuli) shown in the figure, non-musician is significantly over represented with p value equal to 0.00053. Additionally, no significant difference

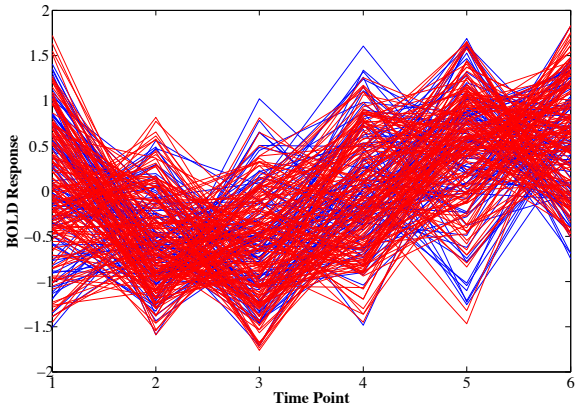


Figure 4-7. Response pattern in cluster A Visual. There are 248 excerpt responses (248 six-points temporal profiles) that were grouped together. Of these similar BOLD response shapes, 87 come from musicians (blue lines) and 161 come from non-musicians (red lines). The p value for the over representation of non-musician in this distribution is 0.00053, indicating non-musicians are more likely to elicit the above response shape than musicians.

in the shape of the BOLD responses was found on the contrasts liked vs. disliked and happy vs. sad in this cluster A. For cluster B Reward and cluster C Auditory, we did not find any significant difference between the response shapes among any of the contrasts.

4.3.6 Music categories and participant groups that cause "strong" BOLD response level

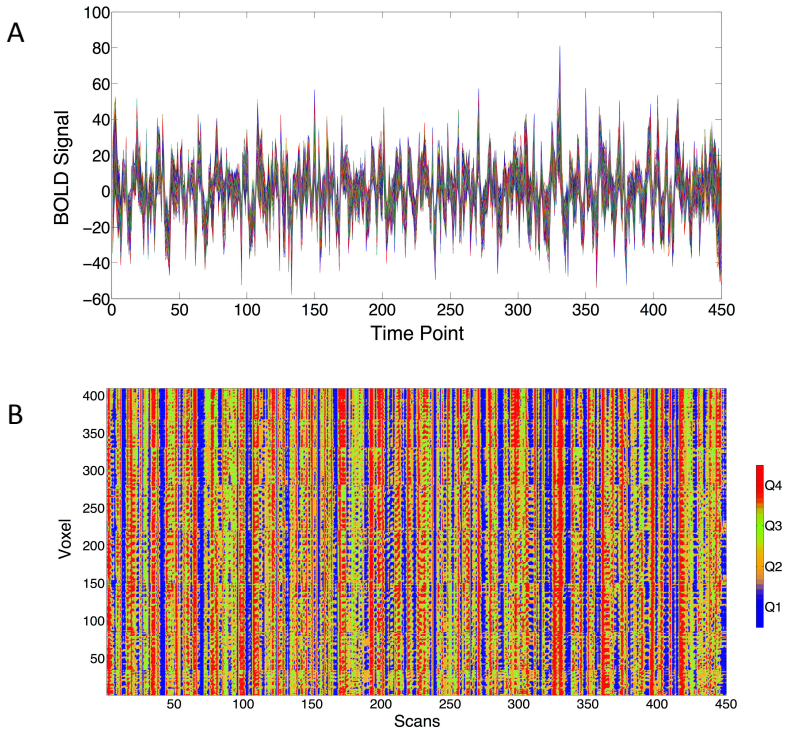


Figure 4-8. Visualization of cluster data from one random subject. (A) is the raw time series of the voxels within this cluster and (B) is the heat map. Values in first quantile (Q1) are plotted as blue, values in fourth quantile are plotted as red (Q4), and values in between are plotted with two other colors (Q2 and Q3). The visible vertical patterns through the whole scan session indicate the synchronized responses among the voxels within this cluster.

Following the procedure introduced in Chapter 2. Each BOLD value from one participant is coded with 1 to 4 (quantiles) indicating its response strength. Figure 4-8 is the visualization of the quantile data of a random cluster.

To demonstrate the individual variability of the BOLD response strength. The threshold values defining the "strong" response for each participant are scattered as shown in Figure 4-9. As can be seen from the figure, these thresholds values vary among different participant. If all the participant data use a global threshold, it is very likely that certain participants always have "strong" responses while some never exhibit "strong" responses. This is why the threshold value is defined participant-wise rather than group-wise.

We extracted out the stimuli that caused the "strong" BOLD response (predominantly higher amplitudes) for the clusters A Visual, cluster B Reward and cluster C Auditory. HyperGeo tests were

then carried out based on these distributions (Table 4-9) with respect to contrasts liked vs. disliked, happy vs. sad and musician vs. non-musician.

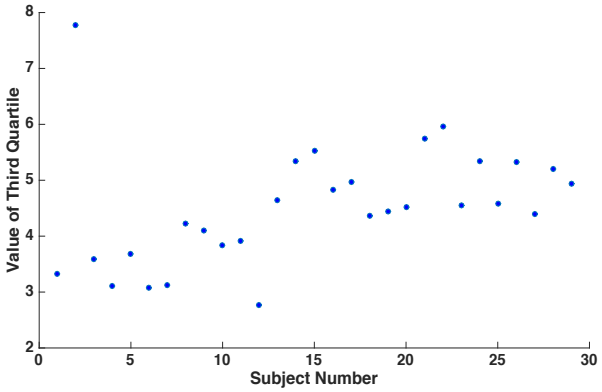


Figure 4-9. The threshold values/third quartiles for strong response in Reward cluster of all the 29 subjects.

Table 4-9. Distribution of the number of excerpts with strong BOLD responses in corresponding categories. Mus stands for group musician and NMus stands for group non-musician.

		DH	DS	LH	LS	Total
Cluster A	Mus	58	53	57	50	218
	NMus	84	73	87	58	302
	Total	142	126	144	108	
Cluster B	Mus	57	49	63	72	241
	NMus	75	69	113	78	335
	Total	132	118	176	150	
Cluster C	Mus	105	86	69	43	303
	NMus	129	94	97	60	380
	Total	234	180	166	103	
All Clusters	Mus	18	16	19	10	63
	NMus	33	18	32	22	105
	Total	51	34	51	32	

We identified the stimulus categories that are more likely to elicit “strong” responses (higher BOLD response levels) with the Bonferroni-corrected p value 0.017 (original p = 0.05). In the cluster A Visual (calcarine fissure and cuneus), a larger number of excerpts of happy music than sad music elicited stronger responses (p = 0.0042). In the cluster B Reward (striatum, thalamus, amygdala, globus pallidus, and olfactory cortex), a larger number of excerpts of liked music than disliked music elicited stronger responses (p = 0.000083). In the cluster C Auditory (superior and middle temporal gyrus, Heschl’s gyrus and Rolandic operculum), a larger number of excerpts of happy

music than sad music elicited stronger responses (p = 1.1e-8). It also showed that a larger number of excerpts of disliked music than liked music elicited stronger responses in the brain regions encompassed by cluster C Auditory (p = 1.8e-12). We also tested the stimuli that simultaneously elicited the strong and most similar responses in the brain areas within all the three clusters. Results showed that a larger number of excerpts of happy music than sad music elicited a stronger large-scale brain response network (p = 0.0023).

4.4 Neuroscientific insights and discussion

In this study, we have tested a novel data-driven consensus clustering framework with an aim to integrate results from several clustering algorithms (rather than applying a single one) to a complex research question related to brain processing of musical emotions. By doing this we obtained several brain regions having consistent and robust pattern of functional connectivity in response to different musical emotions. Based on the clusters obtained from the framework, we found the music categories that elicited strong responses in visual, reward and auditory brain regions. In addition, we also obtained different BOLD responses between musicians and non-musicians.

We used the following criteria to ensure the quality of the clustering results. Firstly, we chose three commonly used clustering methods that have been used in the study involving the analysis of fMRI data. Secondly, we did not give any spatial information but clustered the fMRI data purely based on its time series. In other words, voxels are clustered based on their temporal profiles and not on their topology in the brain. In addition, the fact that the overall size of the clusters was much larger than the Gaussian spatial smoothing kernel size (about 30 voxels) means it is not likely that this similarity comes from the preprocessing step, which made the results more reliable. Finally, the Bi-CoPaM paradigm and M-N scatter plot within the framework generated and selected the non-trivial clusters with a large number of members showing consistently synchronized activities during most of the experimental conditions. In neuroimaging studies, the signal to noise ratio is often very low (Bennett and Miller, 2010; Welvaert and Rosseel, 2013), making it hard to draw any conclusion based on a single or even several experiments. The paradigm in this study has overcome this issue, since integrating the results from 1856 independent clustering processes means that the results are reproducible: if certain clusters would only appear in few clustering trials, then they would be most likely due to random error or other factors and would not be

included in the final results. Remarkably, we obtained these results with a stimulation paradigm that included stimulus sets that varied across participants, and a subjective task related to affective ratings of the stimuli. Hence, the robust findings contrasted the variability of the experimental paradigm.

We further explored how the tunable consensus clustering results from individual clustering methods differ compared with the same from the framework utilizing all three clustering methods. The differences of the results mainly come from the integration process Bi-CoPaM, so in the following discussion, Bi-CoPaM will replace the term “framework” temporally. To compare the differences, for each of the three clusters described above and obtained from all three methods, we obtained the clusters generated by K-means and Bi-CoPaM, hierarchical clustering and Bi-CoPaM, as well as SOM and Bi-CoPaM separately. We compared the coherences and differences and found that the three individual methods give different clustering results. For example, hierarchical clustering did not identify the cluster B Reward and SOM did not identify the cluster C Auditory. Yet, by using the same experimental data, the clusters generated by Bi-CoPaM with all three methods not only include the intersections among the results from three individual methods but also areas that were not identified by any individual method on its own. Therefore, the fact that Bi-CoPaM with multiple algorithms detected clusters that could not be identified by a single algorithm, demonstrates the advantages of integrating multiple methods over using a single specific method.

The consensus clustering framework allows us to find clusters including functionally and anatomically related neural networks responding to emotional music. After the cluster generation and selection, emotion-related brain structures responsible for rewarding and pleasurable sensations such as the basal ganglia, thalamus, insula (Salimpoor and Zatorre, 2013), and other areas involved with processing of auditory features such as the Heschl’s gyrus, the Rolandic operculum and the superior temporal gyrus (Alluri et al., 2012) were grouped into corresponding clusters separately. One of the most important findings of this study is that, without any predetermined model assigning a value to each stimulus, the framework was able to obtain a single cluster including the anatomically connected subcortical and cortical structures of the reward circuit, responding selectively to liked, enjoyed music. This is one of the few studies obtaining such finding with a data-driven method. In a recent study using a data-driven network science method to study affective music processing (Wilkins et al., 2014), no reward circuit activity was found. Our study

confirms findings on the neural structures related to musical emotions obtained with model-based approaches (Barrett and Wager, 2006; Brattico et al., 2013, 2011, Koelsch, 2014, 2010; Koelsch et al., 2006; Salimpoor and Zatorre, 2013). Neural structures of the reward circuit have also been found to be more or less connected only in other functional connectivity analysis studies such as one studying attention-deficit/hyperactivity disorder (ADHD) (Tomasi and Volkow, 2012).

In our statistical tests, we also investigated the response shapes elicited by the different stimulus categories (liked, disliked, happy, and sad music) and experimental groups (musicians and non-musicians). Unlike the traditional statistical tests for fMRI data that compare the response strength differences using a general linear model (Friston et al., 1995), which often have been questioned (Poline and Brett, 2012), the clustering analysis of the mean excerpt response shape would distinguish the response shape difference, providing a finer temporal information than comparing the response magnitude level alone. The neurodynamics of functional connections related to affective responses to music have been previously studied only with model-based approaches and only within selected regions of interests, such as the caudate and the nucleus accumbens (Alluri et al., 2015; Salimpoor et al., 2013). With the current results, we replicated those findings without falling into the risks of circular analysis (Kriegeskorte et al., 2010). Moreover, we evidenced a difference between musicians and non-musicians in the temporal course of the BOLD response for the interconnected cortical areas of cluster A including the calcarine fissure and the cuneus. This finding suggests a larger involvement of visual processes that might be related to imagery or even to a relaxation state in non-musicians accumulated and achieved as a consequence of listening to 18 seconds of emotionally-loaded music. A similar result of tightened connectivity in visual clusters was found by Luo et al. (Luo et al., 2014) with participants lying in the MR scanner at rest with eyes closed, confirming the coupling of these areas during relaxation. Remarkably, while the authors focused the analysis of the differences between musicians and non-musicians on other regions, a tendency for a larger recruitment of visual clusters seems to be present in non-musicians, similarly to the current study.

On the other hand, when the same dataset was analysed using a model-driven approach for evidencing regional activations during music listening, musicians showed larger regional activations in somatomotor areas, such as the precentral and post-central cerebral gyri and the cerebellar declive, in musicians over non-musicians, whereas the latter group of participants did not

show any larger brain activity as compared with musicians. The apparent discrepancy with previous findings obtained with the same dataset relates to the divergent approaches used. In the current study functional coupling among areas within the same cluster was computed and the temporal dynamics of the BOLD response within each cluster was then compared between musicians and non-musicians for all the stimulus conditions, whereas in the study by Brattico et al. (Brattico et al., 2016) the overall magnitudes of the BOLD regional responses in the whole brain were compared between groups with the general linear model and post-hoc t-tests. Moreover, since in the current study our main goal was to validate a new clustering approach rather than testing the neural adaptations to affective music listening as a consequence of musical training, we did not proceed in studying the differences in the response shape patterns between musicians and non-musicians for each of the stimulus categories. In our previous study with the general linear model approach, we obtained new evidence for larger activations of reward-related areas in musicians than non-musicians, in the line of previous findings by Chapin et al. (Chapin et al., 2010) or James et al. (James et al., 2008). As discussed already in Ref. (Reybrouck and Brattico, 2015) future model-based and data-driven studies should solve the issue on the role of expertise in shaping emotional responses to music in the brain.

In summary, we first clustered the data not with one clustering algorithm but with three clustering algorithms independently. Then consensus clustering framework generates consensus among the three sets of clusters. This takes out the risks of capturing artefacts of an individual clustering algorithm. Furthermore, we analyse the data not with one set of parameters but with many sets of parameters. For example, different binarisation thresholds explore consensus clusters with different degrees of tightness, which naturally avoids the pitfalls of a single set of clusters found using a single set of parameters. Thus, we believe the proposed consensus clustering framework provides a robust solution for obtaining the consistently strong activation patterns in neuroimaging studies of the affect dataset.

Chapter 5 Investigation on Effects of Intentionality on The Functional Connectivity during Enjoyment of Unfamiliar Music

The consensus clustering framework has been applied on read fMRI data and it has been demonstrated to be effective and powerful to detect the brain areas consistently exhibiting highly correlated BOLD activities during certain experimental conditions. In this chapter, the consensus clustering paradigm is used to differentiate the functional connectivities under different experimental conditions, namely intentional listening and naturalistic listening. Distinct stable functional networks are obtained, subserving musical emotion processing during three levels of attentional and intentional engagement with the music, from naturalistic listening to descriptive and non-evaluative listening up to intentional evaluative listening. Results indicate the intentionality has different effects on auditory-limbic connectivity during musical emotion perceiving.

5.1 Background on intentionality in affective processing

Although enjoyment of music is a very common phenomenon, listening to music is not always intentional, since it often accompanies daily activities such as shopping or TV watching. According to a study by (Sloboda and O'Neill, 2001) using the experience sampling method, about 44% of the events recorded involved music but in only 2% of them music was listened to intentionally and attentively. In these instances of casual and unfocused listening, we do not necessarily carry out a conscious evaluation of the music heard in terms of aesthetic properties, such as beauty, structure or mastery. According to music psychologist (Sloboda, 2010), in everyday life the expression and induction of basic emotions such as joy or sadness by music are prioritized over "aesthetically tinged" emotions such as deep enjoyment, awe or frissons (Sloboda, 2010, p. 503). According to a recent account (Brattico, 2015; Brattico et al., 2013; Brattico and Pearce, 2013; Nieminen et al.,

2011; Reybrouck and Brattico, 2015), a full musical experience includes final outcomes such as aesthetic emotions or reward (often accompanied by bodily changes such as goose bumps on the skin, accelerated heartbeat, or tears in the eyes), aesthetic judgments ("this music is so beautiful") and the formation of specific preferences and musical taste ("I love chamber music"). In a broader framework encompassing all experiences of an art object, Chatterjee and Vartanian (2016) propose that all art phenomena emerge from the interaction between three main mental and neural system, a sensory-motor one (sensation, perception, motor system), a knowledge-meaning one (expertise, context, culture) and an emotion-evaluation one (reward, emotion, wanting/liking). Also in the framework by (Juslin, 2013) aesthetic judgment was viewed as the final outcome of several different emotion-inductive mechanisms. In several accounts, though, additional factors are listed that enable to reach a full aesthetic experience. Among them, one that is considered especially crucial is a dedicated, conscious intention toward the art object, also described as a special attitude ("aesthetic stance"), focus (Hodges, 2016) or intentionality (Brattico et al., 2013; Brattico and Pearce, 2013). Even if this intentionality factor is feasible to study, very little research has been dedicated to determine its role in an aesthetic response during music listening, such as pleasure or enjoyment.

In an analytic study of aesthetic processes in the visual modality, (Höfel and Jacobsen, 2007a) instructed participants to passively view abstract black and white patterns or to contemplate them aesthetically, i.e., to reflect upon the beauty of those shapes, although without giving an overt aesthetic judgment. The electric brain potentials elicited during the two experimental conditions evidenced that aesthetic evaluation occurred during contemplation only and not during mere viewing, as indexed by a late positive potential visible only in the contemplation condition. Furthermore, an early frontocentral negativity to "not beautiful" shapes reflecting impression formation and previously observed during tasks involving overt aesthetic judgments (Jacobsen and Höfel, 2003) was not found in that subsequent study (Höfel and Jacobsen, 2007a). The authors hence postulated a distinction between aesthetic mode or "central processes of thinking about aesthetic value" and "deciding upon an aesthetic judgment".

In terms of brain structures distinguishing intentional from involuntary pleasure, a rare meta-analysis by Kühn and Gallinat (2012) has combined results from 39 neuroimaging studies related to pleasure as induced by odour, taste, music or visual stimuli. Overall, positive correlates of conscious, subjective pleasure were selectively obtained in medial orbitofrontal cortex, left nucleus accumbens (ventral striatum), pregenual cortex, left thalamus, and mid cingulate cortex. Several of those structures are consistently found in relation to motivational stimuli, as well as expected or reward anticipation (Mueller et al., 2015). Particularly, the nucleus accumbens is described as the “hot spot” of the brain (Peciña et al., 2006). These results replicated a previous meta-analysis by (Brown et al., 2011). The latter study additionally identified the anterior insula as a hub common to all sensory modalities in association with pleasurable stimuli (whether consciously evaluated for their affective qualities or not). In the meta-analysis by (Kühn and Gallinat, 2012) the clusters of activation found in the selected studies were further subdivided into the ones in which participants judged pleasantness during scanning (18 studies) from those in which they judged the stimuli outside the scanner (11 studies). The main interest by the authors of the study was in testing the hypothesis of a medial orbitofrontal function for self-referential processes involving conscious hedonic decisions. However, no difference was found although a relation between left amygdala activation and conscious pleasure judgments done during scanning was noticed. This meta-analysis, while commendable in trying to discern neural correlates of distinct psychological processes, puts forward the need for further empirical work within each sensory modality.

In the music domain, a growing number of studies (including few meta-analyses) has looked at the brain structures and neural connections associated with perceived or felt musical emotions. Nevertheless, very little knowledge has been accumulated when it comes to intentionality during these emotional experiences. A rare attempt to study intentionality during affective processing with neuroimaging method has been done by (Bogert et al., 2016). In their study 30 music excerpts from blockbuster film soundtracks lasting 4 seconds were presented twice to subjects in two separate (counterbalanced) conditions. In one condition subjects were asked to pay attention to the numbers of instruments playing in the clip (implicit condition); in the other condition they were instructed to classify the emotions conveyed by the music (explicit condition). In the implicit condition (contrasted with the explicit one) the music stimuli activated bilaterally the inferior parietal lobule, premotor cortex, as well as reward-related areas such as the caudate (dorsal striatum)

and ventromedial frontal cortex. In contrast, dorsomedial prefrontal and occipital areas, previously associated to emotion recognition and cognitive processing of music, were active during explicit classification of musical emotions. Indeed, according to the conceptual-act model of emotions by (Barrett and Wager, 2006; Lindquist et al., 2012), discrete emotions occur only after the neurophysiological states of valence and arousal (which form what is called 'core affect') meet with an act of categorization and labeling happening in dorsolateral prefrontal and parietal cortices. This conceptual act occurs "in the moment" and uses pre-existing knowledge of emotions and language systems in the brain to attribute a lexical category (Barrett, 2006).

In a recent experiment we studied the chronometry of the neural responses during categorisation of musical emotions by using neurophysiological methods (Ellison and Brattico, 2015). We chose a very simplified paradigm in order to measure the phase-locked event-related responses allowing very fine temporal resolution in the order of milliseconds. Stimuli were chord cadences ending with a major or minor chord that could be tuned or mistuned in the middle note. Results showed that negatively rated (sad or incorrect) cadence endings in both tasks elicited early neural responses whereas only later responses peaking at around 500 ms, differed between sad and incorrect stimuli, suggesting a neural chronometry of music listening in which feature encoding and sensory memory processes are followed at a medium latency by affective classification, after which an evaluative stage takes place (similar findings have been obtained also in the visual domain: Höfel and Jacobsen, 2007b; Jacobsen, 2014; Jacobsen and Höfel, 2003).

To complement the scarcely available literature, in the present study, we set out to depict whether intention, i.e., attention focused on carrying out an explicit liking judgments of music (as opposed to attention to the making descriptive judgment of the music or to listening per se), is necessary to co-activate limbic, paralimbic and reward system regions of the brain. In other words, we wanted to determine whether liking and enjoyment are spontaneous affective processes during listening, even when attention is diverted towards some specific aspects of the music, or the mind is concentrated on the listening itself. To this aim, we used fMRI to scan the brain activities of healthy adult volunteers while they passively listened to 15 seconds music excerpts selected by the experimenters, varying in musical genre, acoustic features, and emotional connotations, or else they classified the excerpts based on the gender of the singer or based on whether they enjoyed the excerpts or not. Human brain has been considered to be a very efficient network and

examining the connectivity between different brain regions have drawn great attention these days (M. P. van den Heuvel et al., 2008; Wilkins et al., 2014). For the analysis of the interconnected neural networks, we adopted the developed consensus clustering framework to identify the subset of voxels that are consistently correlated under different circumstances (Abu-Jamous et al., 2015b, 2013). In the work reported in Chapter 4 (Liu et al., 2017), the consensus clustering framework successfully identified the brain structures functionally connected to evaluative liking judgments of music. In this study, we predicted co-activation in a network of mesiotemporal limbic structures, including the nucleus accumbens, in response to the liked musical stimuli, irrespectively of the experimental task performed by the subjects, namely irrespectively of whether they were focusing on making a liking evaluation or not. In turn, we anticipated functional connectivity within prefrontal and parietooccipital regions specifically in association with the conscious decision processes of liking judgment. Moreover, we predicted that the decision process might downregulate the activity and connectivity of the limbic and reward networks during listening to the liked music excerpts. For the disliked musical excerpts, we expected the involvement of the amygdala, insula and auditory cortices, similarly to the findings in Chapter 4 including liked and dislike music across all the experimental conditions, but particularly for the conditions not requiring the conscious liking decision (Burunat et al., 2015).

5.2 Affect 2 fMRI dataset

5.2.1 Participant

A total of 25 healthy volunteers (16 females and 9 males) without any hearing, neurological or psychological problem are included in this study. Before the experiment, participants were contacted by e-mail and asked to name three or four genres (or sub-genres) of music that they prefer and other three or four genres that they strongly dislike, along with examples. This information was used to select songs within three different sub-genres of the pop-rock repertoire that could accommodate the preferences of all the participants. The experiment is approved by the ethical committee of the coordinating Uusimaa and Helsinki Hospital and complied with the Helsinki Declaration.

5.2.2 Stimulation

According to the information provided by the participants, the music excerpts are selected by researchers. There are 36 audio excerpts from commercially available pop/rock songs. These excerpts varied in musical genre, acoustic features, and emotional connotations to minimise the possible effects of these factors on people's perception of the sound and emotions. Each music excerpt lasts 15 seconds. Table 5-1 lists the information of the stimuli used in this study.

Table 5-1. Stimulus information.

Stimulus	Name	Artist
1	No I in Threesome	Interpol
2	The Greatest	Cat Power
3	Lover, You Should've Come Over	Jeff Buckley
4	Once Upon A Time	Air
5	Karma Police	Radiohead
6	Glory Box	Portishead
7	Transmission	Joy Division
8	Son of Sam	Elliott Smith
9	No Brakes	The Bravery
10	Shadow Valley	Castanets
11	Temptation	Jeremy Warmsley
12	Souvenirs	Architecture in Helsinki
13	Closer	Ne-Yo
14	Ride It	Jay Sean
15	No One	Alicia Keys
16	Better in Time	Leona Lewis
17	One Step at A Time	Jordin Sparks
18	Run The Show	Kat DeLuna Feat. Busta Rhymes
19	Beautiful Liar	Beyoncé & Shakira
20	Ayo Technology	50 Cent Feat. Justin Timberlake
21	Whenever, Wherever	Shakira
22	Ass Like That	Eminem
23	Disturbia	Rihanna
24	Take A Bow	Rihanna
25	Thunderstruck	AC/DC
26	Ace of Spades	Motörhead
27	School's Out	Alice Cooper
28	Child in Time	Deep Purple
29	Bridges in The Sky	Dream Theater
30	The Wicker Man	Iron Maiden
31	Iron Man	Black Sabbath
32	Locomotive Breath	Jethro Tull
33	Death Row	Judas Priest
34	Immigrant Song	Led Zeppelin
35	The Beautiful People	Marilyn Manson
36	Nothing Else Matters	Metallica

5.2.3 fMRI experiment procedure

The experiment was conducted at the Advanced Magnetic Imaging (AMI) Centre at Aalto University, Espoo, Finland. Upon arrival at the laboratory, and before entering the scanner, participants were asked to listen to the stimuli to allow them to be equally familiar with them and thereby to minimise the possible bewilderment from unfamiliarity factor, which was previously shown to have a strong influence in the pleasure response of listeners to music (Pereira et al. 2013). Subsequently they changed their clothes and were prepared to enter the scanner room. Participants' fMRI responses were acquired while they listened to each of the musical stimuli in some random order. For each participant the stimuli loudness was adjusted to a comfortable but audible level inside the scanner room (around 75 dB (SPL)) and the sound was delivered via MRI compatible earphone that did not cause interferences with the magnetic field generated by MRI machine.

For each participant, there are in total three scanning sessions, where different experimental conditions are set for each session. The following describe what participants were asked to do in corresponding session.

- **Naturalistic listening session**: participants were asked to listen to the music naturalistically without making any explicit judgments.
- **Gender classification session**: participants were asked to determine the gender of the singer in each music excerpts presented.
- **Liking classification session**: participants were asked to decide whether they like or dislike each excerpt.

The first scanning session is always naturalistic listening for everyone. The orders of gender classification session and liking classification session were counterbalanced across subjects, i.e., some had gender classification session as the second one while others had liking classification session as the second. This design is to neutralise the possible effects on the decisions from the order of the sessions that participants took. When each music excerpt was played, participants were asked to look at the fixed cross symbol at the centre of screen in the scanner. After each 15 seconds long music excerpt, there was a 3.5 seconds interval for the participants to response to the question shown on the screen by pressing the corresponding button. The order of the answer each response button represents (e.g.: left for like, right for dislike) was counterbalanced between

subjects. Then there was a ten seconds long silence period before delivering the next music excerpt. Figure 5-1 illustrates the experimental paradigm in each block.

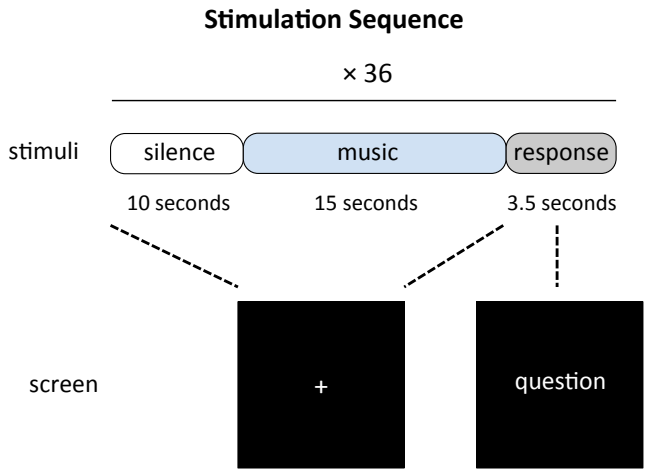


Figure 5-1. Schematic representation of the experimental trials used for each of the three sessions constituting the complete experiment (naturalistic listening, gender classification and liking classification).

5.2.4 data acquisition 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 AMI Centre. Using a single-shot gradient echo planar imaging (EPI) sequence, 33 oblique slices (field of view = 192×192 mm; 64×64 matrix; slice thickness = 4 mm, interslice skip = 0 mm; echo time = 32 ms; flip angle = 75°) were acquired every 2 seconds, covering the whole-brain for each participant. T1-weighted structural images (176 slices; field of view = 256×256 mm; matrix = 256×256; 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), VBM for SPM (Voxel Based Morphometry; Wellcome Department of Imaging Neuroscience, London, UK), and customized scripts developed by the technicians and researchers who took part in the data collection and preprocessing. 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 grey matter, white matter, and cerebrospinal fluid by VBM, 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 ratio by means of spatial smoothing using 8 mm full-width-at-half-maximum (FWHM) Gaussian filter.

The preparation of the data for the consensus clustering analysis follows the same procedure as used in Chapter 4. In brief, fMRItoolbox (implemented at the University of Jyväskylä in MATLAB environment, e.g., used in (Alluri et al., 2012; Burunat et al., 2016, 2015) is the main tool for preparing the data. Firstly, the 3D volume data was converted to a vector (228453×1) by using a standard brain mask. The above step had been applied to every 3D volume scan from each subject and all the scans were combined sequentially, forming the fMRI time series of each subject. The time series were high pass filtered with a cut-off frequency of 1/120Hz to remove the linear trend and scanner drift. Then, for each participant, according to the order that musical excerpts were played and responded by the participant, the whole fMRI time series were segmented into 36 segments, each containing 7 or 8 time points (covering 15 seconds at a sampling rate of 2 second per whole-brain image), and each corresponding to instances when the participants were listening to music excerpts.

After obtaining all the music excerpts time-series data, next step is to categorise them, that is which session each excerpt is played and what response is given to the corresponding excerpt. For each participant, all the 36 music excerpts were labelled as liked or disliked according to the responses the participants gave to each music excerpt in liking classification session. In gender classification block and naturalistic listening block, although the participants did not perform the liking judgment, the excerpts were still labelled as liked or disliked according to the response each excerpt received in liking classification block to study the effect of listening to liked or disliked music when participants were not actively performing liking judgments. In summary, there are six categories of music excerpts from each participant:

- **LL**: liked music excerpts in liking classification session (19 excerpts avg.)
- **LD**: disliked music excerpts in liking classification session (17 excerpts avg.)

- **GL**: liked music excerpts in gender classification session (19 excerpts avg.)
- **GD**: disliked music excerpts in gender classification session (17 excerpts avg.)
- **NL**: liked music excerpts in naturalistic listening session (19 excerpts avg.)
- **ND**: disliked music excerpts in naturalistic listening session (17 excerpts avg.)

For each category, different participant might have different number of music excerpts. In general, the average number of liked and disliked music per participant are roughly the same (liked: 19 excerpts on average; disliked: 17 excerpts on average).

5.2.5 Clustering experiment

5.2.5.1 Cluster generation

Following the same procedure used in generating individual clustering results from single clustering algorithms, each excerpts data (normalized to 0 mean and unit variance) was clustered by K-means, hierarchical clustering and SOM with the number of clusters K equals to 10, 25, 50, and 100 in separate partitions. The clustering index for each excerpt data had a label that was the same as the label of the corresponding music excerpt category (LL, LD, GL, GD, NL, ND). These labels are used later to form the dataset corresponding to different experimental conditions (e.g., liking judgment session) in the consensus clustering analysis.

The six music excerpt categories are divided into various combinations to reflect the corresponding experimental conditions. For example, by combining the clustering results from datasets LL and LD, the brain structures that consistently showing synchronised BOLD responses during liking judgment tasks were identified. This consensus clustering analysis procedure has been demonstrated effective against traditional single clustering algorithm analysis in Chapter 4. By obtaining robust clusters in different experimental conditions, the differences and common properties of the clusters can be compared, indicating the distinct brain functional connectivity under different level of affective processing of music emotions and preferences, i.e., intentional on liking judgment, naturalistic, and intentional but on gender judgment. The following list shows the music excerpt combination scheme:

- Liking judgment block (LL, LD)
- Gender judgment block (GL, GD)

- Naturalistic listening block (NL, ND)

Note the above three dataset combinations reflect the real fMRI experiment design, i.e., three sessions. Each of the dataset combination is fed into the consensus clustering framework to generate the consensus clustering results, following the methods introduced in Chapter 3 and 4.

5.2.5.2 Filtering

After the clusters are generated by Bi-CoPaM, the topological refinement introduced in section 3.3.6 is applied, which is also used in the study in Chapter 4. Note a slightly adapted procedure for removing the voxels with weak responses (voxels whose time series have a very small variance) is used. Since this fMRI experiment has three sessions, the decision of which original voxel time-series to use to determine the strong BOLD activities depends on the dataset combination that yields the clusters to be filtered. For example, the clusters from combination of LL and LD should be filtered using the voxel time-series in the liking judgment session. However, for the clusters from combination of LL, GL, and NL, all the voxel time-series from three fMRI sessions should be used as the generation of these clusters included all the data from these three sessions. Following the established parameters used in previous chapter, the criteria of keeping a voxel is that its variance has to be greater than half of the mean of the variance for all the voxels in a particular cluster. Then if more than seventy percent of the participants show a strong fluctuation in the whole session at a certain voxel, this voxel is retained. When data from multiple sessions are used in the filtering, then another criterion is added, which is the kept voxel has to show a strong fluctuation in all the sessions involved. Finally, we used the fMRItoolbox developed by University of Jyväskylä to remove the scattered tiny or even single voxel clusters.

5.3 Results

In each experimental condition, we inspected the first 20 clusters (ranked by M-N plots algorithm), as these clusters showed very strong similarity in the response shapes as well as covered large continuous regions, which complies with expectations based on knowledge of brain physiology.

5.3.1 Topology of clusters

The topology of clusters in each experimental condition is rendered on a standard structural 3D brain. Several slices covering the important brain structures are plotted separately to give a better view of the position and shape of the clusters. Each cluster has a number (e.g., C6), indicating its order selected by MN-plot technique, and marked with a colour to be distinguished from the other clusters within the same 3D brain.

5.3.1.1 Liking judgment session (LL+LD)

The anatomical labels, size, and MNI coordinates are shown in Table 5-2. Figure 5-2 illustrates the topology of clusters from liking judgment. Cluster C3 comprises areas such as the supramarginal and postcentral gyri, previously related to language and somatosensory processing as well as the middle temporal gyrus, Rolandic operculum, and inferior frontal gyrus, previously associated with the cognitive processing of sounds. In addition, brain regions related to action observation and motor preparation, such as the supplementary motor area (SMA), the precentral gyrus, are also included, as well as the bilateral angular gyri. Cluster C5 mainly includes higher-order structures involved with visual information processing, namely the cuneus, lingual gyrus, middle, inferior and superior occipital gyri, and fusiform gyrus.

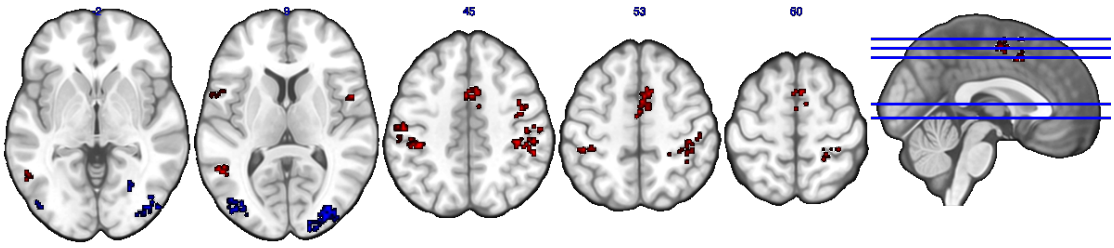


Figure 5-2. Topology of clusters from liking judgment session. There are two clusters with C3 coloured as red and C5 as blue.

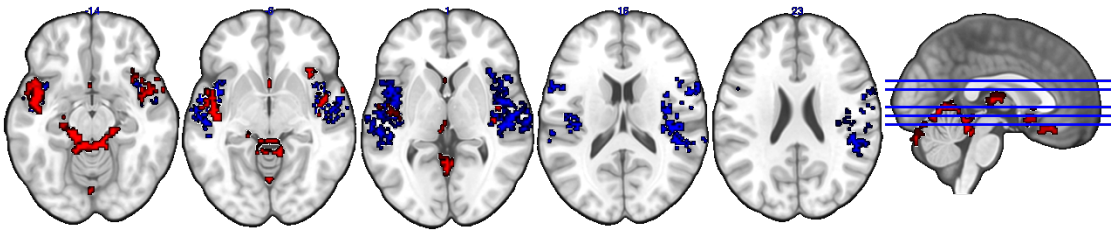


Figure 5-3. Topology of clusters from gender judgment session. There are two clusters with C3 coloured as red and C5 as blue.

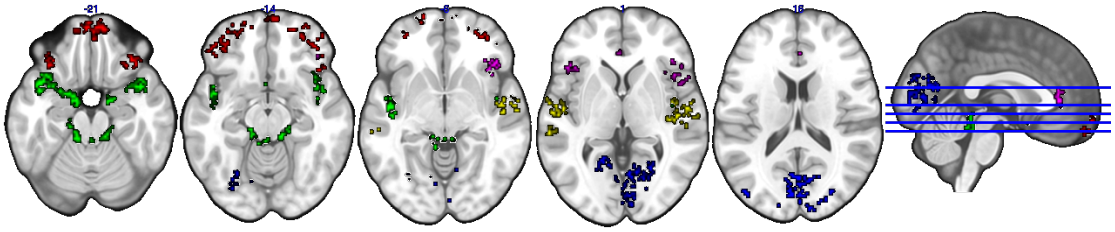


Figure 5-4. Topology of clusters from naturalistic listening session. There are six clusters with C4 coloured as red, C5 as blue, C6 as green, C8 as violet, C9 as yellow, and C11 as cyan.

5.3.1.2 Gender judgment session (GL+GD)

The anatomical labels, size, and MNI coordinates are shown in

Table 5-3. Figure 5-3 illustrates the topology of clusters from gender judgment. Cluster C3 includes three major systems, namely the auditory processing system (middle and superior temporal gyri) limbic system (thalamus, amygdala, parahippocampal gyrus, orbitofrontal gyrus, insula, putamen) and the cerebellum. Cluster C5 comprises a broad area of the auditory cortex and he bilateral insula. Other structures interconnected in this cluster are the bilateral inferior frontal gyrus, the SMA, the right supramarginal gyrus, the precentral and postcentral gyri, plus a small bit of the right putamen.

5.3.1.3 Naturalistic listening session (NL+ND)

The anatomical labels, size, and MNI coordinates are shown in Table 5-4. Figure 5-4 illustrates the topology of clusters from naturalistic listening. Cluster C4 includes various parts of the orbital frontal cortex (inferior, middle, and superior) extending to a small part of the left middle and inferior temporal gyrus. Cluster C5 mainly contains structures related to visual processing, e.g., occipital gyrus, cuneus, and fusiform gyrus. Cluster 6 comprises structures that have a similar topology to the cluster C3 in gender judgment condition, namely the auditory and limbic system plus a small part of cerebellum. Cluster C8 includes bilateral anterior cingulate and paracingulate gyrus as well as bilateral insula, where the left insula is very small compared to the right part, plus various positions of inferior frontal cortex (triangular, orbital, and opercular). Cluster C9 is a combination of the auditory processing related structures (middle and superior gyrus, Heschl's gyrus) and right insula. Cluster C11 is the smallest one including three structures within the right hemisphere.

5.3.2 Cluster topology interaction

In order to compare the differences of clusters between two experiment sessions, Venn diagram is used to show the overlaps and exclusive brain structures between any two sets of clusters from corresponding experiment sessions, giving a more straightforward view than comparing the two sets of clusters individually. Nevertheless, the area information (anatomical labels, size, and MNI coordinates) is also extracted out to provide finer information when biological insights are discussed.

5.3.2.1 Liking judgment and naturalistic listening

The comparison between liking judgment and naturalistic listening (Figure 5-5, Table 5-5) shows only one overlapped area in cluster C5 in both conditions (C5_N and C5_L) is found. The structures included are scattered across various brain regions such as postcentral gyrus, inferior frontal cortex, supramarginal gyrus, precentral gyrus, and other two structures having only one voxel each. The majority of the brain structures included in each experiment session are not overlapped.

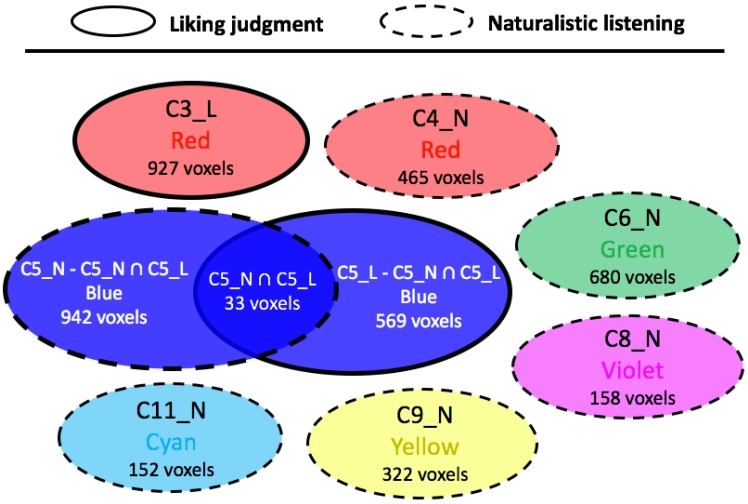


Figure 5-5. Cluster topology interaction between liking judgment and gender judgment. Solid circles represent clusters from liking judgment condition and dashed circles represent clusters from gender judgment condition. The colour label within each circle indicates the colour of the corresponding cluster and the number of voxels indicates the size of different clusters/brain regions.

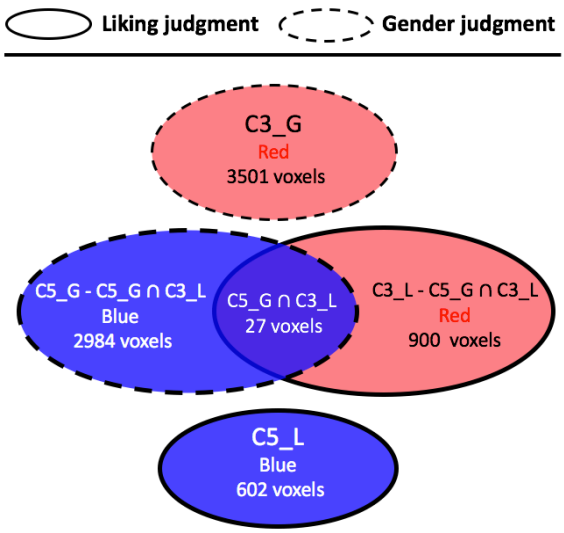


Figure 5-6. Cluster topology interaction between liking judgment and gender judgment. Solid circles represent clusters from liking judgment condition and dashed circles represent clusters from gender judgment condition.

5.3.2.2 Liking judgment and gender judgment

For the comparison between liking judgment and gender judgment (Figure 5-6, Table 5-6), it shows almost no overlapping between the two sets of clusters from liking judgment condition and gender judgment condition. Only 27 voxels representing part of the visual processing system are shared by cluster C3 in liking judgment (C3_L) and C5 in gender judgment (C5_G).

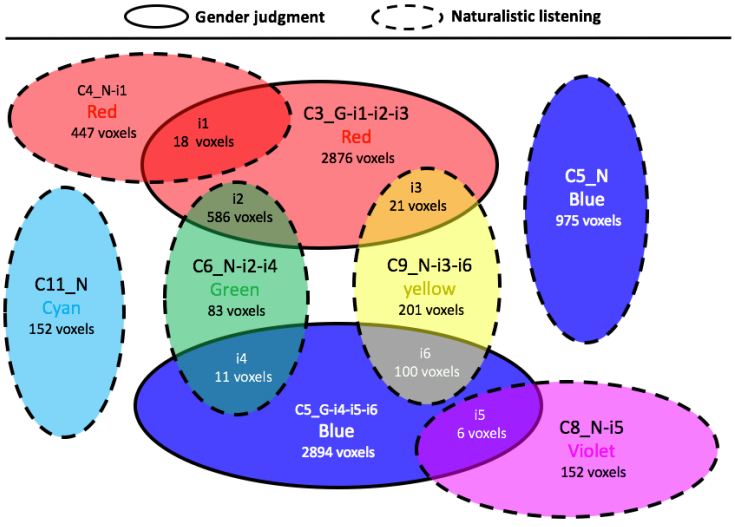


Figure 5-7. Cluster topology interaction between gender judgment and naturalistic listening. Solid circles represent clusters from gender judgment condition and dashed circles represent clusters from naturalistic listening condition.

5.3.2.3 Gender judgment and naturalistic listening

In contrast to the previous two comparisons (liking judgment and naturalistic listening, liking judgment and gender judgment), the cluster topology interaction between gender judgment and naturalistic listening shows several overlapping areas (Figure 5-7, Table 5-7). All of the two clusters in gender judgment (C3_G and C5_G) have overlaps with four out of six clusters in naturalistic listening (C4_N, C6_N, C8_N, and C9_N). Among all the overlapped areas (i1 to i6), i2 and i6 contain large amount of voxels. Area i2 is the intersection between the auditory-limbic systems in gender judgment and naturalistic listening respectively, plus the difference in cerebellum. Area i6 contains the overlapped auditory cortex between the two experiment conditions compared. Other overlapped areas (i1, i3, i4, and i5) are smaller and contain no more than two structures as shown in Table 5-7.

Table 5-2. Anatomical information for clusters in liking judgment session.

C3(red)		
Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Supramarginal gyrus (R)'	[59,-28,32]	195
'Postcentral gyrus (R)'	[46,-26,44]	162
'Postcentral gyrus (L)'	[-53,-21,37]	133
'Supramarginal gyrus (L)'	[-58,-28,33]	64
'Supplementary motor area (R)'	[6,5,52]	75
'Inferior parietal, but supramarginal and angular gyri (L)'	[-49,-27,42]	53
'Inferior frontal gyrus, opercular part (L) '	[-55,9,16]	35
'Middle temporal gyrus (L)'	[-53,-55,6]	67
'Precentral gyrus (R)'	[48,-3,40]	45
'Rolandic operculum (R)'	[51,2,13]	28
'Inferior parietal, but supramarginal and angular gyri (R)'	[45,-38,49]	25
'Supplementary motor area (L)'	[-1,1,52]	24
'Median cingulate and paracingulate gyrus (R)'	[4,8,42]	10
'Precentral gyrus (L)'	[-55,6,26]	8
'Rolandic operculum (L)'	[-54,4,12]	1
'Median cingulate and paracingulate gyrus (L)'	[-1,0,47]	2

C5 (blue)		
Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Middle occipital gyrus (R)'	[33,-86,9]	143
'Middle occipital gyrus (L)'	[-38,-82,12]	162
'Superior occipital gyrus (R)'	[23,-89,20]	68
'Fusiform gyrus (L)'	[-29,-68,-12]	68
'Inferior occipital gyrus (R)'	[41,-81,-6]	44
'Superior occipital gyrus (L)'	[-18,-91,32]	25
'Cuneus (R)'	[18,-96,11]	15
'Calcarine fissure and surrounding cortex (R)'	[7,-88,4]	17
'Fusiform gyrus (R)'	[30,-72,-11]	20
'Inferior occipital gyrus (L)'	[-33,-76,-9]	10
'Lingual gyrus (L)'	[-22,-72,-12]	13
'Lingual gyrus (R)'	[25,-66,-1]	10
'Calcarine fissure and surrounding cortex (L)'	[4,-90,5]	5
'Cuneus (L)'	[-10,-89,37]	2

Table 5-3. Anatomical information for clusters in gender judgment session.

C3(red)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (L)'	[-40,12,-22]	596
'Temporal pole, superior temporal gyrus (R)'	[41,12,-23]	409
'Crus I of cerebellum (L)'	[-30,-80,-25]	374
'Parahippocampal gyrus (L)'	[-17,-10,-22]	234
'Superior temporal gyrus (L)'	[-48,-5,-8]	256
'Lobule III of cerebellum (R)'	[12,-34,-19]	88
'Vermis III of cerebellum'	[2,-40,-12]	77
'Crus I of cerebellum (R)'	[40,-76,-28]	218
'Temporal pole, middle temporal gyrus (L)'	[-42,14,-31]	126
'Lobule III of cerebellum (L)'	[-9,-37,-18]	67
'Lobules IV-V of cerebellum (L)'	[-12,-39,-14]	149
'Parahippocampal gyrus (R)'	[21,-3,-24]	134
'Amygdala (L)'	[-22,0,-20]	61
'Insula (R)'	[40,3,-8]	139
'Temporal pole, middle temporal gyrus (R)'	[44,14,-29]	83
'Vermis IV-V of cerebellum'	[1,-55,-1]	64
'Amygdala (R)'	[29,1,-24]	34
'Hippocampus (L)'	[-19,-17,-16]	62
'Thalamus (L)'	[-3,-21,8]	66
'Vermis I-II of cerebellum'	[1,-38,-20]	8
'Inferior frontal gyrus, orbital part (R)'	[36,23,-18]	55
'Crus II of cerebellum (L)'	[-8,-87,-26]	59
'Olfactory cortex (L)'	[0,12,-9]	15
'Vermis VI of cerebellum'	[-1,-73,-10]	17
'Lobules IV-V of cerebellum (R)'	[10,-44,-9]	27
'Gyrus rectus (L)'	[0,28,-18]	21
'Olfactory cortex (R)'	[30,10,-20]	1
'Gyrus rectus (R)'	[1,23,-17]	4
'Inferior temporal gyrus (L)'	[-56,-55,-24]	31
'Hippocampus (R)'	[21,-3,-21]	3
'Caudate nucleus (L)'	[-2,15,0]	2
'Fusiform gyrus (L)'	[-18,-35,-19]	8
'Inferior temporal gyrus (R)'	[54,-55,-26]	13

C5(blue)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[41,-2,3]	415
'Superior temporal gyrus (R)'	[55,-24,6]	662
'Superior temporal gyrus (L)'	[-53,-18,4]	405
'Rolandic operculum (R)'	[49,-10,13]	231
'Rolandic operculum (L)'	[-47,-11,10]	182
'Heschl"s gyrus (R)'	[45,-18,9]	108
'Heschl"s gyrus (L)'	[-43,-18,9]	72
'Insula (L)'	[-42,0,0]	158
'Middle temporal gyrus (L)'	[-56,-25,-1]	233
'Postcentral gyrus (R)'	[57,-15,38]	181
'Inferior frontal gyrus, opercular part (R)'	[49,14,11]	88
'Inferior frontal gyrus, opercular part (L) '	[-55,10,11]	72
'Precentral gyrus (R)'	[58,2,33]	115
'Supramarginal gyrus (R)'	[53,-31,25]	56
'Supplementary motor area (R)'	[9,-3,63]	30
'Putamen (R)'	[33,7,7]	3

Table 5-4. Anatomical information for clusters in naturalistic listening.

C4(red)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Middle frontal gyrus, orbital part (R)'	[33,52,-14]	88
'Inferior frontal gyrus, orbital part (R)'	[39,31,-18]	77
'Inferior frontal gyrus, orbital part (L)'	[-40,37,-16]	76
'Middle frontal gyrus, orbital part (L)'	[-27,56,-13]	49
'Gyrus rectus (R)'	[5,56,-20]	31
'Temporal pole, middle temporal gyrus (L)'	[-48,15,-30]	27
'Superior frontal gyrus, orbital part (R)'	[13,57,-19]	23

'Superior frontal gyrus, orbital part (L)'	[-15,60,-15]	23
'Gyrus rectus (L)'	[-3,59,-21]	19
'Inferior frontal gyrus, medial orbital (R)'	[7,66,-12]	17
'Inferior frontal gyrus, medial orbital (L)'	[-3,65,-10]	8
'Middle temporal gyrus (L)'	[-60,-1,-27]	20
'Inferior temporal gyrus (L)'	[-57,-4,-29]	7

C5(blue)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Calcarine fissure and surrounding cortex (L)'	[-7,-74,10]	227
'Calcarine fissure and surrounding cortex (R)'	[11,-76,8]	166
'Cuneus (R)'	[11,-83,23]	135
'Cuneus (L)'	[-3,-82,21]	101
'Lingual gyrus (R)'	[13,-65,1]	72
'Superior occipital gyrus (R)'	[20,-89,22]	54
'Middle occipital gyrus (L)'	[-34,-83,18]	78
'Lingual gyrus (L)'	[-10,-67,0]	63
'Superior occipital gyrus (L)'	[-19,-89,26]	49
'Fusiform gyrus (L)'	[-29,-70,-11]	25
'Middle occipital gyrus (R)'	[29,-81,16]	4
'Inferior occipital gyrus (L)'	[-32,-78,-12]	1

C9(yellow)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Superior temporal gyrus (R)'	[57,-17,3]	177
'Superior temporal gyrus (L)'	[-55,-11,2]	69
'Middle temporal gyrus (L)'	[-59,-27,-2]	39
'Insula (R)'	[43,-9,-1]	24
'Heschl"s gyrus (R)'	[46,-19,7]	11
'Heschl"s gyrus (L)'	[-50,-16,7]	2

C6(green)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (L)'	[-37,10,-23]	153
'Parahippocampal gyrus (L)'	[-16,-10,-23]	107
'Lobule III of cerebellum (R)'	[12,-33,-19]	53
'Temporal pole, superior temporal gyrus (R)'	[39,9,-22]	83
'Lobule III of cerebellum (L)'	[-11,-35,-20]	31
'Superior temporal gyrus (L)'	[-46,-6,-9]	63
'Lobules IV-V of cerebellum (L)'	[-12,-36,-17]	44
'Vermis III of cerebellum'	[1,-39,-12]	24
'Parahippocampal gyrus (R)'	[19,-7,-21]	41
'Amygdala (L)'	[-21,1,-21]	15
'Insula (R)'	[42,6,-10]	33
'Vermis I-II of cerebellum'	[0,-37,-19]	5
'Amygdala (R)'	[29,3,-27]	6
'Hippocampus (L)'	[-14,-6,-21]	9
'Temporal pole, middle temporal gyrus (L)'	[-40,9,-28]	5
'Lingual gyrus (L)'	[-11,-36,-7]	7

C11(violet)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[41,20,-3]	29
'Anterior cingulate and paracingulate gyrus (L)'	[-1,37,8]	27
'Inferior frontal gyrus, orbital part (R)'	[41,29,-7]	23
'Inferior frontal gyrus, triangular part (R)'	[47,24,4]	24
'Inferior frontal gyrus, triangular part (L)'	[-46,22,2]	24
'Inferior frontal gyrus, opercular part (R)'	[51,15,5]	19
'Anterior cingulate and paracingulate gyrus (R)'	[3,41,7]	7
'Inferior frontal gyrus, orbital part (L)'	[-45,22,-4]	3
'Insula (L)'	[-39,22,1]	2

C11(cyan)

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Angular gyrus (R)'	121	[50,-59,32]
'Middle frontal gyrus (R)'	20	[31,16,46]
'Inferior parietal, but supramarginal and angular gyri (R)'	11	[52,-58,44]

Table 5-5. Cluster topology interaction between liking judgment and naturalistic listening.

C5_N ∩ C5_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Superior occipital gyrus (R)'	[21,-91,21]	13
'Middle occipital gyrus (L)'	[-38,-80,16]	9
'Fusiform gyrus (L)'	[-29,-74,-11]	4
'Calcarine fissure and surrounding cortex (R)'	[7,-90,4]	3
'Middle occipital gyrus (R)'	[29,-80,16]	2
'Superior occipital gyrus (L)'	[-22,-90,30]	1
'Lingual gyrus (R)'	[22,-70,2]	1

C5_N − C5_N ∩ C5_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Calcarine fissure and surrounding cortex (L)'	[-7,-74,10]	227
'Calcarine fissure and surrounding cortex (R)'	[11,-76,8]	163
'Cuneus (R)'	[11,-83,23]	135
'Cuneus (L)'	[-3,-82,21]	101
'Lingual gyrus (R)'	[13,-65,1]	71
'Lingual gyrus (L)'	[-10,-67,0]	63
'Superior occipital gyrus (L)'	[-19,-89,26]	48
'Middle occipital gyrus (L)'	[-34,-83,18]	69
'Superior occipital gyrus (R)'	[20,-88,23]	41
'Fusiform gyrus (L)'	[-29,-70,-11]	21
'Inferior occipital gyrus (L)'	[-32,-78,-12]	1
'Middle occipital gyrus (R)'	[29,-81,15]	2

C5_L − C5_N ∩ C5_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Middle occipital gyrus (R)'	[33,-86,9]	141
'Middle occipital gyrus (L)'	[-38,-82,12]	153
'Superior occipital gyrus (R)'	[24,-89,20]	55
'Fusiform gyrus (L)'	[-30,-68,-12]	64
'Inferior occipital gyrus (R)'	[41,-81,-6]	44
'Superior occipital gyrus (L)'	[-18,-91,32]	24
'Cuneus (R)'	[18,-96,11]	15
'Fusiform gyrus (R)'	[30,-72,-11]	20
'Inferior occipital gyrus (L)'	[-33,-76,-9]	10
'Calcarine fissure and surrounding cortex (R)'	[7,-88,3]	14
'Lingual gyrus (L)'	[-22,-72,-12]	13
'Lingual gyrus (R)'	[26,-66,-1]	9
'Calcarine fissure and surrounding cortex (L)'	[4,-90,5]	5
'Cuneus (L)'	[-10,-89,37]	2

Table 5-6. Cluster topology interaction between liking judgment and gender judgment.

C5_G ∩ C3_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Postcentral gyrus (R)'	[54,-20,35]	13
'Inferior frontal gyrus, opercular part (L) '	[-57,8,12]	7
'Supramarginal gyrus (R)'	[55,-26,29]	3
'Precentral gyrus (R)'	[58,4,26]	2
'Rolandic operculum (R)'	[54,4,16]	1
'Supplementary motor area (R)'	[10,-2,58]	1

C5_G − C5_G ∩ C3_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[41,-2,3]	415
'Superior temporal gyrus (R)'	[55,-24,6]	662
'Superior temporal gyrus (L)'	[-53,-18,4]	405
'Rolandic operculum (R)'	[49,-10,13]	230
'Rolandic operculum (L)'	[-47,-11,10]	182
'Heschl"s gyrus (R)'	[45,-18,9]	108
'Heschl"s gyrus (L)'	[-43,-18,9]	72

'Insula (L)'	[-42,0,0]	158
'Middle temporal gyrus (L)'	[-56,-25,-1]	233
'Postcentral gyrus (R)'	[57,-14,39]	168
'Inferior frontal gyrus, opercular part (R)'	[49,14,11]	88
'Inferior frontal gyrus, opercular part (L) '	[-54,10,11]	65
'Precentral gyrus (R)'	[58,2,33]	113
'Supramarginal gyrus (R)'	[53,-31,24]	53
'Supplementary motor area (R)'	[9,-4,63]	29
'Putamen (R)'	[33,7,7]	3

C5_G − C5_G ∩ C3_L

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Supramarginal gyrus (R)'	[59,-28,32]	192
'Postcentral gyrus (R)'	[45,-26,44]	149
'Postcentral gyrus (L)'	[-53,-21,37]	133
'Supramarginal gyrus (L)'	[-58,-28,33]	64
'Supplementary motor area (R)'	[6,5,52]	74
'Inferior parietal, but supramarginal and angular gyri (L)'	[-49,-27,42]	53
'Middle temporal gyrus (L)'	[-53,-55,6]	67
'Inferior frontal gyrus, opercular part (L) '	[-54,9,17]	28
'Precentral gyrus (R)'	[48,-4,41]	43
'Rolandic operculum (R)'	[50,2,13]	27
'Inferior parietal, but supramarginal and angular gyri (R)'	[45,-38,49]	25
'Supplementary motor area (L)'	[-1,1,52]	24
'Median cingulate and paracingulate gyrus (R)'	[4,8,42]	10
'Precentral gyrus (L)'	[-55,6,26]	8
'Rolandic operculum (L)'	[-54,4,12]	1
'Median cingulate and paracingulate gyrus (L)'	[-1,0,47]	2

Table 5-7. Cluster topology interaction between gender judgment and naturalistic listening.

i1

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, middle temporal gyrus (L)'	[-49,14,-28]	9
'Inferior frontal gyrus, orbital part (R)'	[39,26,-20]	9

i2

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (L)'	[-37,10,-23]	143
'Parahippocampal gyrus (L)'	[-16,-9,-23]	106
'Lobule III of cerebellum (R)'	[12,-33,-19]	46
'Temporal pole, superior temporal gyrus (R)'	[40,9,-22]	68
'Lobule III of cerebellum (L)'	[-11,-35,-20]	30
'Parahippocampal gyrus (R)'	[19,-7,-22]	39
'Superior temporal gyrus (L)'	[-46,-6,-9]	49
'Lobules IV-V of cerebellum (L)'	[-13,-35,-18]	36
'Vermis III of cerebellum'	[2,-39,-12]	16
'Amygdala (L)'	[-21,1,-21]	15
'Amygdala (R)'	[29,3,-27]	6
'Insula (R)'	[43,3,-9]	14
'Vermis I-II of cerebellum'	[0,-37,-19]	3
'Hippocampus (L)'	[-14,-6,-21]	9
'Temporal pole, middle temporal gyrus (L)'	[-40,9,-28]	5
'Olfactory cortex (L)'	[-2,10,-14]	1

i3

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Superior temporal gyrus (L)'	[-52,-9,-1]	11
'Insula (R)'	[43,-11,-2]	10

i4

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[43,3,-8]	9

'Superior temporal gyrus (L)'	[-48,-2,-6]	2

i5

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[39,19,-2]	4
'Inferior frontal gyrus, opercular part (R)'	[52,17,4]	2

i6

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Superior temporal gyrus (R)'	[55,-18,4]	50
'Superior temporal gyrus (L)'	[-54,-13,2]	24
'Heschl"s gyrus (R)'	[45,-18,7]	8
'Insula (R)'	[45,-7,-1]	8
'Middle temporal gyrus (L)'	[-57,-27,0]	8
'Heschl"s gyrus (L)'	[-50,-16,7]	2

C6_N − i2 − i4

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (R)'	[35,8,-21]	15
'Vermis III of cerebellum'	[-1,-38,-12]	8
'Lobule III of cerebellum (R)'	[12,-35,-18]	7
'Temporal pole, superior temporal gyrus (L)'	[-40,11,-20]	10
'Superior temporal gyrus (L)'	[-45,-6,-9]	12
'Insula (R)'	[41,14,-11]	10
'Lobules IV-V of cerebellum (L)'	[-11,-41,-10]	8
'Lingual gyrus (L)'	[-11,-36,-7]	7
'Vermis I-II of cerebellum'	[1,-37,-19]	2
'Lobule III of cerebellum (L)'	[-4,-38,-16]	1
'Parahippocampal gyrus (R)'	[16,-3,-18]	2
'Parahippocampal gyrus (L)'	[-20,-28,-24]	1

C3_G − i1 − i2 − i3

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (L)'	[-41,13,-22]	453
'Temporal pole, superior temporal gyrus (R)'	[41,13,-23]	341
'Crus I of cerebellum (L)'	[-30,-80,-25]	374
'Crus I of cerebellum (R)'	[40,-76,-28]	218
'Superior temporal gyrus (L)'	[-48,-5,-8]	196
'Parahippocampal gyrus (L)'	[-18,-11,-22]	128
'Temporal pole, middle temporal gyrus (L)'	[-41,14,-31]	112
'Vermis III of cerebellum'	[2,-40,-12]	61
'Lobules IV-V of cerebellum (L)'	[-11,-40,-13]	113
'Parahippocampal gyrus (R)'	[22,-2,-25]	95
'Insula (R)'	[40,4,-9]	115
'Temporal pole, middle temporal gyrus (R)'	[44,14,-29]	83
'Amygdala (L)'	[-22,0,-20]	46
'Vermis IV-V of cerebellum'	[1,-55,-1]	64
'Lobule III of cerebellum (L)'	[-7,-39,-16]	37
'Lobule III of cerebellum (R)'	[12,-35,-19]	42
'Thalamus (L)'	[-3,-21,8]	66
'Hippocampus (L)'	[-19,-19,-15]	53
'Amygdala (R)'	[29,1,-24]	28
'Crus II of cerebellum (L)'	[-8,-87,-26]	59
'Inferior frontal gyrus, orbital part (R)'	[35,22,-17]	46
'Vermis VI of cerebellum'	[-1,-73,-10]	17
'Lobules IV-V of cerebellum (R)'	[10,-44,-9]	27
'Olfactory cortex (L)'	[0,12,-9]	14
'Vermis I-II of cerebellum'	[1,-38,-20]	5
'Gyrus rectus (L)'	[0,28,-18]	21
'Inferior temporal gyrus (L)'	[-56,-55,-24]	31
'Olfactory cortex (R)'	[30,10,-20]	1
'Gyrus rectus (R)'	[1,23,-17]	4
'Hippocampus (R)'	[21,-3,-21]	3
'Caudate nucleus (L)'	[-2,15,0]	2
'Fusiform gyrus (L)'	[-18,-35,-19]	8
'Inferior temporal gyrus (R)'	[54,-55,-26]	13

C5_G − i4 − i5 − i6

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[40,-2,3]	394
'Superior temporal gyrus (R)'	[55,-24,6]	612
'Superior temporal gyrus (L)'	[-53,-18,4]	379
'Rolandic operculum (R)'	[49,-10,13]	231
'Rolandic operculum (L)'	[-47,-11,10]	182
'Heschl''s gyrus (R)'	[45,-18,9]	100
'Insula (L)'	[-42,0,0]	158
'Heschl''s gyrus (L)'	[-43,-18,9]	70
'Middle temporal gyrus (L)'	[-56,-25,-1]	225
'Postcentral gyrus (R)'	[57,-15,38]	181
'Inferior frontal gyrus, opercular part (R)'	[49,14,11]	86
'Inferior frontal gyrus, opercular part (L) '	[-55,10,11]	72
'Precentral gyrus (R)'	[58,2,33]	115
'Supramarginal gyrus (R)'	[53,-31,25]	56
'Supplementary motor area (R)'	[9,-3,63]	30
'Putamen (R)'	[33,7,7]	3

C9_N − i3 − i6

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Superior temporal gyrus (R)'	[58,-16,2]	127
'Superior temporal gyrus (L)'	[-58,-11,2]	34
'Middle temporal gyrus (L)'	[-60,-27,-2]	31
'Heschl''s gyrus (R)'	[48,-20,7]	3
'Insula (R)'	[43,-7,0]	6

C4_N − i1

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Temporal pole, superior temporal gyrus (L)'	[-40,12,-22]	596
'Temporal pole, superior temporal gyrus (R)'	[41,12,-23]	409
'Crus I of cerebellum (L)'	[-30,-80,-25]	374
'Parahippocampal gyrus (L)'	[-17,-10,-22]	234
'Superior temporal gyrus (L)'	[-48,-5,-8]	256
'Crus I of cerebellum (R)'	[40,-76,-28]	218
'Lobule III of cerebellum (R)'	[12,-34,-19]	88
'Lobules IV-V of cerebellum (L)'	[-12,-39,-14]	149
'Vermis III of cerebellum'	[2,-40,-12]	77
'Lobule III of cerebellum (L)'	[-9,-37,-18]	67
'Temporal pole, middle temporal gyrus (L)'	[-41,14,-31]	117
'Parahippocampal gyrus (R)'	[21,-3,-24]	134
'Amygdala (L)'	[-22,0,-20]	61
'Insula (R)'	[40,3,-8]	139
'Temporal pole, middle temporal gyrus (R)'	[44,14,-29]	83
'Vermis IV-V of cerebellum'	[1,-55,-1]	64
'Amygdala (R)'	[29,1,-24]	34
'Hippocampus (L)'	[-19,-17,-16]	62
'Thalamus (L)'	[-3,-21,8]	66
'Crus II of cerebellum (L)'	[-8,-87,-26]	59
'Vermis I-II of cerebellum'	[1,-38,-20]	8
'Inferior frontal gyrus, orbital part (R)'	[35,22,-17]	46
'Olfactory cortex (L)'	[0,12,-9]	15
'Vermis VI of cerebellum'	[-1,-73,-10]	17
'Lobules IV-V of cerebellum (R)'	[10,-44,-9]	27
'Gyrus rectus (L)'	[0,28,-18]	21
'Olfactory cortex (R)'	[30,10,-20]	1
'Inferior temporal gyrus (L)'	[-56,-55,-24]	31
'Gyrus rectus (R)'	[1,23,-17]	4
'Hippocampus (R)'	[21,-3,-21]	3
'Caudate nucleus (L)'	[-2,15,0]	2
'Fusiform gyrus (L)'	[-18,-35,-19]	8
'Inferior temporal gyrus (R)'	[54,-55,-26]	13

C8_N − i5

Anatomical labels	MNI coordinates [x, y, z]	Size [voxels]
'Insula (R)'	[41,-2,3]	411
'Superior temporal gyrus (R)'	[55,-24,6]	662
'Superior temporal gyrus (L)'	[-53,-18,4]	405
'Rolandic operculum (R)'	[49,-10,13]	231
'Rolandic operculum (L)'	[-47,-11,10]	182

'Heschl"s gyrus (R)'	[45,-18,9]	108
'Heschl"s gyrus (L)'	[-43,-18,9]	72
'Insula (L)'	[-42,0,0]	158
'Middle temporal gyrus (L)'	[-56,-25,-1]	233
'Postcentral gyrus (R)'	[57,-15,38]	181
'Inferior frontal gyrus, opercular part (R)'	[49,14,11]	86
'Inferior frontal gyrus, opercular part (L) '	[-55,10,11]	72
'Precentral gyrus (R)'	[58,2,33]	115
'Supramarginal gyrus (R)'	[53,-31,25]	56
'Supplementary motor area (R)'	[9,-3,63]	30
'Putamen (R)'	[33,7,7]	3

5.4 Neuroscientific insights and discussion

In this real fMRI study, the consensus clustering framework is employed on fMRI data from three different sessions to study whether and how the conscious evaluation of the music heard in terms of aesthetic properties would modulate the emotion- and reward-related brain connectivity. We obtained distinct neural networks subserving music enjoyment during three levels of attentional and intentional engagement with the music, from naturalistic listening, to descriptive, non-evaluative listening up to intentional evaluative listening. Results support our hypothesis on the role of intentionality in auditory-limbic connectivity during music enjoyment. The obtained clusters clearly point at auditory-limbic connectivity between areas such as thalamus, superior temporal gyrus, amygdala, and parahippocampal gyrus, or between orbitofrontal regions or between supratemporal regions, insula and putamen, only during unfocussed, unintentional listening, namely when participants were asked to either classify the gender of the voice in the music excerpts or to simply passively listen to them. When participants were asked to decide whether they liked or not the music excerpt, only two clusters of intercommunicating brain regions were found: one including regions related to cognitive processing of sounds (middle temporal gyrus, rolandic operculum, inferior frontal gyrus), and regions related to action observation and motor preparation (supplementary motor areas, precentral gyrus); the other cluster comprises higher-order structures involved with visual processing (cuneus, lingual gyrus, middle, inferior and superior occipital gyri, fusiform gyrus).

The choice of consensus clustering strategy is an answer to the recent criticisms on the most common analysis methods in fMRI (Eklund et al., 2016). Typically, an fMRI study on a certain stimulation paradigm would adopt a single method of analysis and statistical thresholding and if

a second study on the same stimulation paradigm would utilize another method of analysis divergent results would occur. Considering that methods of analysis and statistics have proliferated in the field, it is paramount to avoid a scattered picture of the results gained by fMRI (Kriegeskorte et al., 2009). The consensus clustering framework allows us to merging many analysis methods and to obtaining robust and reproducible clusters from various datasets. In Chapter 4, we point out that different clustering algorithms (K-mean, SOM and hierarchical) produce partly divergent sets of clustering results, whereas Bi-CoPaM with all the three algorithms generates consensus among them, thus reducing the risks of capturing artefacts from single clustering algorithm. Furthermore, current results show nontrivial clusters covering large continuous brain regions, confirming the robustness of the method. Remarkably, unlike some algorithms that artificially introduce the spatial constraints to the clustering generation process (Blumensath et al., 2013; Craddock et al., 2012), our spatial information free strategy guaranteed that the voxels in fMRI data were clustered purely based on the similarities of their BOLD time series rather than on their topologies in the brain. It should be mentioned though that, similarly to other methods, the proposed consensus clustering framework for studying functional connectivity does not provide information on the temporal succession of increased connectivity in each obtained clusters.

The first finding of this study is the separation of clusters of correlated neural activity between the three experimental conditions. The conditions not requiring a conscious evaluation of liking of the music excerpts were most similar to each other in terms of shared voxels in the resulting clusters as evidenced by the topology interaction analyses. While the naturalistic listening condition showed similar functional connectivity only between visual areas than the liking judgment condition (particularly parts of the bilateral middle and superior occipital gyri), and it showed similarly correlated neural activity than that to the gender judgment condition only in fronto-parietal areas (particularly the right postcentral gyrus and the pars opercularis of the left inferior frontal gyrus), it did share four clusters with the gender judgment condition, meaning that it had similar increased connectivity between auditory (temporal pole, bilateral superior and middle temporal gyri, Heschl's gyrus), frontal (orbital part of inferior frontal gyrus), and limbic areas (parahippocampal gyrus, amygdala, insula, hippocampus).

When comparing the clusters obtained for each of the three experimental conditions, the connectivity of motor-related with the Rolandic operculum was much more evident for the Gender judgment condition as opposed to the Liking judgment condition, with only one shared voxel between the two. While the Rolandic operculum has been related to musical pleasure in previous studies (Green et al., 2012; Koelsch et al., 2006), it is also implicated in both overt and covert singing and speaking (Jeffries et al., 2003; Riecker et al., 2000; Wildgruber et al., 1996). One can thus speculate that the focus on the vocal properties of the stimulus would have prompted participants to recruit sound production planning areas of the brain.

Chatterjee and Vartanian (2016) recently proposed that all art phenomena emerge from the interaction between three main mental/neural systems: a sensory-motor one (sensation, perception, motor system), a knowledge-meaning one (expertise, context, culture) and an emotion-evaluation one (reward, emotion, wanting/liking). Also Juslin (2013) viewed aesthetic judgment as the final outcome of a summation of different emotion-inductive mechanisms. In our previous work (Brattico et al., 2013; Brattico and Pearce, 2013; Nieminen et al., 2011; Reybrouck and Brattico, 2015), we proposed a detailed spatiotemporal road map of music aesthetic processes in the brain, suggesting a distinction between unconscious, low-level perceptual-emotional stages and reflective processes involving cognitive control and leading to the three main outcomes of an aesthetic experience, namely emotion, preference and judgment. The early and late emotional processes during a musical experience can be modulated by what Hodges (Hodges, 2016) has termed “focus”, namely the act of paying attention to the music. Here and in previous work (Brattico et al., 2013), we extended this concept to intentionality as an internal state predisposing to attentive watching/listening in the case of performance arts or contemplation in the case of static arts, also inspired by previous proposals (Bundgaard, 2015; Hargreaves et al., 2012). Based on these premises, we here hypothesize that the individual’s psychological state or internal context, particularly intentionality, is an important predictor of the emotion-related brain processes occurring during music listening.

Our findings of connected regions within the ventral and dorsal attention networks, including parietal regions (bilateral supramarginal and angular gyri), and frontal regions (bilateral precentral and postcentral gyri, supplementary motor area, cingulate cortex), and within visual structures (including bilateral middle, superior occipital gyri, cuneus, lingual gyrus) is in line with the notion

that focused attention on the music aimed at providing an evaluation, as in the liking judgment condition, recruited supramodal cognitive and attention areas of the brain. Interestingly, a previous study (Bogert et al., 2016) compared two conditions contrasting evaluative and descriptive judgments of the same musical material also showed regional activations in fronto-parietal and occipital brain structures (such as bilateral superior frontal gyrus, inferior occipital gyrus, lingual gyrus, fusiform gyrus) specifically for the condition requiring the explicit classification of the emotions perceived in the music (the choice was between the three categories of happy, sad and fearful). Crucially, while in music domain the intentionality for an emotional listening mode seems to down-regulate subcortical emotion-related neural activity, in the visual-art domain the effect seems divergent based on a neurophysiological study showing neural correlates of affective processes only when the participants were focused on giving an evaluative beauty judgment of abstract black-white patterns (Höfel and Jacobsen, 2007a).

In turn, the condition in which participants were asked to focus on descriptive aspects of the music excerpts, namely whether they contained one, few or many instruments, elicited more subcortical neural structures such as in the caudate, pallidum, and cortical areas previously linked to emotion processing, such as the inferior parietal lobule (see also Chapin et al., 2010; Flores-Gutiérrez et al., 2007; Satoh et al., 2011). Also in this study, the connected regions activated by the gender judgment task have been formerly clearly related to emotion processing: parahippocampal gyrus, amygdala, insula, hippocampus, thalamus, medial orbitofrontal cortex, caudate nucleus, and the vermis of the cerebellum, in a network closely communicating with ventral stream auditory regions such as the anterior superior temporal gyrus. A second network involved with the gender judgment task included sensorimotor regions coupled with the bilateral insula and the right putamen. Naturalistic listening produced coupled activity in several overlapping regions than those elicited by the gender judgment condition, although with a more scattered pattern showing six separate clusters over attention-, perception- and emotion-related areas. Overall, the degree of connectivity between striatal areas, ventrolateral prefrontal regions and auditory cortices have been repeatedly found in recent studies to be crucial for determining the subjective experiment of enjoyment in music (Blood and Zatorre, 2001; Martínez-Molina et al., 2016; Sachs et al., 2016; Salimpoor et al., 2013).

Notably, the activity and connectivity of sensorimotor areas (such as the precentral and postcentral gyri and the supplementary motor area) found for all the three experimental conditions and particularly for the gender judgement condition, have been consistently observed in response to music-induced emotions (Blood and Zatorre, 2001; Bogert et al., 2016; Mitterschiffthaler et al., 2007; (Brattico et al., 2015). The connected areas found here (particularly the opercular part of the inferior frontal gyrus and the inferior parietal lobule) also partially overlap with the action observation network, also termed “mirror neuron” system, that is activated both by motor production by an individual and by perception of motor acts by others (Morin and Grèzes, 2008; Rizzolatti et al., 1996). Some proposed theories in music psychology argue that motor mimicking of sounds resembling an emotional vocalization is a crucial mechanism for inducing emotions (Juslin, 2013; Juslin and Västfjäll, 2008).

In conclusion, the study prove that intentionality in judging the hedonic value of a musical piece is important in shaping neural connectivity to music, and specifically in connecting brain regions related to attention and cognition. In turn, when attention is focused on non-evaluative aspects of the music areas related to emotions and pleasures become more coupled. Here it is relevant to state that the findings are obtained with music unfamiliar to the participants. Based on previous findings obtained with another paradigm (Brattico et al., 2015), we might expect opposite effects with highly familiar music. The current results inform theories about aesthetic experience of music and call for further research in combined music and visual-art domains.

Chapter 6 Summary and Future Work

This thesis adopts the concept of consensus clustering to analyse brain functional connectivity, aiming to reduce the bias from the underlying hypothesis that single clustering algorithms have. A complete flexible framework of analysing brain functional connectivity using consensus clustering is designed and validated. We demonstrate that the proposed framework improves over traditional single clustering algorithm in terms of the completeness of clusters and features the capability of producing robust clustering results in terms of cross-participant variability, which are the essential requirements for every good analysis strategy. The framework is applied on two datasets from real fMRI experiments (Affect and Affect 2) in the thesis, which yield several novel and important neuroscientific findings. In this chapter, the whole thesis is briefly summarised first, followed by discussing the future directions of the consensus clustering analysis of brain functional connectivity.

6.1 Summary

6.1.1 Consensus clustering framework for analysis fMRI data

In this thesis, a novel consensus clustering framework for analysing fMRI data is designed to meet the call for consensus and consistent analysis of neuroimaging data. The framework starts from processing the raw BOLD time-series, then forms the appropriate subsets of data that correspond to the experimental paradigm, followed by clustering experiments and integration of multiple clustering experiments consisting of more than one clustering algorithms and many datasets. The evaluation procedure of the results is also embedded into the framework, including filtering parameters design, robustness test against individual variability, and the selection of non-trivial clusters. In the rest of this section, the important aspects of the framework are discussed.

The correct analysis starts with the data with good quality. In this thesis, the preprocessing of fMRI data is carried out following a well-established pipeline and have been used in several published works (Brattico et al., 2016, 2011). There is no perfect preprocessing pipeline for every fMRI study as the detailed procedure often depends on the type of the experiment (e.g. event-

related or block design) and the performance of the scanner. When extracting out the BOLD time-series corresponding to each stimulus, i.e. music excerpts in this thesis, the whole time-series is firstly detrended and filtered using a high-pass filter to exclude the low frequency components like scanner drift that might introduce artificial correlation or similarity between voxel time-series. This ensures the voxels are clustered purely based on the similarity of their BOLD activities, together with the strategy that no spatial constraints are put on either individual clustering experiments or consensus clustering results generation.

Regarding the individual clustering experiments, only the clustering algorithms that have been widely used in analysing neuroimaging studies are adopted, i.e., K-means, hierarchical clustering, and SOM in the thesis. These algorithms are relatively computationally efficient and can cope with the large-scale voxel-wise clustering strategy, compared with other methods such as spectral clustering that requires matrix decomposition. Grid computing is used to further reduce the time for obtaining a large amount of individual clustering experiments (at least thousands) by running multiple experiments at the same time, thanks to the feature that individual clustering results do not need to communicate with each other before consensus integration. And for each dataset, i.e., excerpt BOLD time-series, three methods with multiple number of clusters Ks are applied, providing a multi-perspective views of clustering solutions. These solutions form the basis for the consensus cluster generation in the later stage.

The clustering results integration starts from fuzzy consensus partition matrix formation. For a set of partition matrices with the same number of clusters K, a reference partition matrix is selected first and then the rest partition matrices are aligned and fused into the reference partition matrix. This step is often done by randomly permuting the order of the partition matrices and then adding them one by one sequentially (Abu-Jamous et al., 2015b, 2014). To minimise the random effect in this step, an intermediate quality ranking is done for all the partition matrices to be merged. The ranking is done by scattering all the clusters from a partition, denoted by a dot, in a two dimensional coordinate with horizontal axis representing the cluster-wise MSE value and vertical axis representing the size of clusters (normalised). Then a score equal to the mean distance between all the clusters and the top left corner (the criterion of best cluster) is calculated, indicating the general quality of this partition in terms of the balance between the size of clusters and their intra cluster variability. All the partition matrices are then ranked in a descend order according to the

score of each partition. Through this ranking, partitions with high scores are firstly merged to yield the fuzzy memberships for those high quality clusters. Then the partitions with low scores are merged later. This strategy avoids the scenario where a partition with poor clustering result is taken as the reference at the very first step in the random permutation of partition matrices, thus undermine the power of integrating multiple clustering solutions.

After obtaining the fuzzy consensus partition matrix, a binarisation step is adopted to assign the consensus membership for each voxel. The difference threshold binarisation approach is chosen as it considers the competitiveness of the clusters when assigning each data point to a certain cluster. Thus it generates tight clusters not only based on the absolute fuzzy membership value for each voxel but also considers the competitiveness of the clusters over the same voxel. The threshold for binarisation is a free parameter ranging from 0 to 1. To provide a complete set of binarisation results, here we do not arbitrarily set the threshold to a certain value but set multiple values between 0 and 1 at an equal spacing. Thus a collection of clusters with different tightness levels are generated. Because this step brings in a large numbers of clusters, the question of which ones are better arises. We address this issue by adopting a cluster selection procedure that follows the logic of ranking the partitions with respect to their qualities. To define the good clusters, we combine the methodological definition together with the neuroscientific interpretation, where a good cluster should have very small intra-cluster variability and cover a relatively large area of the brain. To satisfy this criterion, the good clusters should be as close to the top left corner as possible since the top left corner indicates a cluster with zero cluster-wise MSE and the largest size among all the scattered clusters. The selection starts with the cluster (dot) having the shortest distance between itself and the top left corner, followed by removing the clusters having conjunctions with it. Then for the rest of clusters, the one that is closest to the top left corner is selected and other clusters having overlaps with it are removed. Repeating these steps will yield unique clusters one by one that do not have any overlaps with each other. This procedure stops until no cluster is left or the number of clusters selected reaches a pre-defined value. As for how many clusters to be selected and analysed, researchers could inspect all the selected ones or choose the first C clusters, depending on the computational resources available.

The cluster filtering is a further refinement that makes clusters more focal by removing those voxels that are likely to be grouped together randomly or showing weak BOLD responses in most

of the participants. This is achieved by firstly applying the hypergeometric test on the area information of each cluster. If the ratio of a certain brain area (BA) within one particular cluster is similar to the percentage of the whole BA within the whole brain, then this can be achieved by randomly sampling a cluster having the same size of this cluster. In this case, this area will be removed as it poses a risk of being included randomly. Another filtering followed is to remove the voxels that having weak responses. This is due to the fact that during the clustering experiments, the BOLD time-series is normalised and thus lose part of the magnitude information. We want to keep the voxels showing not only highly correlated BOLD activities but also strong responses (magnitudes). Note that the systematic validation of the filtering parameters has been demonstrated in Chapter 4 and it provides a guide of avoiding inappropriate parameter that significantly change the topology of clusters. In summary, both of these filterings are not performed directly on the time-series, which might change the BOLD activity shapes. Rather they are mainly used for establishing the topology of clusters based on the neuroscientific facts.

For the whole consensus clustering analysis framework, we have done the following two experiments to demonstrate its validity and robustness respectively. We firstly compared the differences among various clustering algorithm combinations against single clustering algorithms. It shows that the combination of all the three clustering algorithms used in this thesis provides the most complete set of clusters that covers the crucial brain areas related to the fMRI experiment paradigm. Importantly, all the studies to date using consensus clustering analysis strategy only employ one algorithm. We believe our multiple algorithms integration greatly improve the accuracy, consistency, and completeness of clustering results. Secondly, the frameworks' robustness against individual BOLD response variability is tested. We find the designed framework can cope with the variability well and produce consistent and reasonable clustering solutions from different subsets of all the participants. This robustness is important as most fMRI experiments use limited number of participants due to the time and financial constraints. To draw sound conclusions from these participants, the most consistent parts of the clustering results have to be extracted out. In this test, we also find the benefits from recruiting more participants for the fMRI experiments. With more participants to counter the individual variability together with the framework itself, the clustering results become more solid and consistent. As a matter of fact, some undergoing brain

projects such as the WU-Minn Human Connectome Project has recruited 1200 healthy participants to carry out the neuroimaging study.

Besides the main framework, two statistical analyses are also designed to extract interesting information from the consensus clustering results, namely analysis of stimuli that tend to elicit stronger responses and response shape analysis, with the former related to the BOLD response magnitudes and the later related to the BOLD response shape. These two analysis are applied on Affect study (Chapter 4), yielding meaningful results that correspond well to the existing literature on study of affect processing. Note these two methods are not fixed in the main framework, one can decide to use them or not depending on what research questions are asked. For example, in Affect 2 study (Chapter 5), the main goal is to differentiate the brain functional connectivity among different levels of intentionality on listening to music, thus only the main framework is applied without carrying out the strong responses analysis and BOLD response shape analysis.

6.1.2 Neuroscientific insights

In addition to the proposed consensus clustering analysis framework, another important part of this research study presented in this thesis is the close collaborations with researchers from neuroscience field to investigate the neuroscientific meaning of various consensus clustering results. This starts from my academic visiting in Finland in my first year where I took part in neuroimaging data collection and processing and later through the whole period of time when the intensive data analysis and interpretation are done. We believe the close collaborations with the people from the neuroscience help us exploit the advantages of the framework by receiving valuable feedbacks at different stages during the development of the framework. The rest part of this section will summarise the novel neuroscientific findings we have obtained.

In Affect study, we find clusters including functionally and anatomically related neural networks responding to music with different emotions (happy or sad) and preferences (liked and disliked). Emotion-related brain structures responsible for rewarding and pleasurable sensations comprise of brain structures such as the basal ganglia, thalamus, insula. Other areas involved with processing of auditory features (Heschl's gyrus, the Rolandic operculum and the superior temporal gyrus) and visual information processing (Bilateral calcarine fissure and cuneus) are grouped into corresponding clusters separately. One of the most important findings of this study is that, without

any predetermined model that is compulsory in traditional GLM approach, the proposed consensus clustering analysis framework is able to obtain a single cluster including the anatomically connected subcortical and cortical structures of the reward circuit, responding to music with emotions and preferences. This is one of the few studies obtaining such finding with a data-driven method. Note that a recent study applied network theory, which is also a data-driven approach, to study affective music processing (Wilkins et al., 2014), but no reward circuit activity was found. Our study confirms findings on the neural structures related to musical emotions. The reward circuit is an important brain area that is found not only associated with affective processing but also with attention-deficit/hyperactivity disorder (ADHD). With the response shape analysis, we evidence a difference between musicians and non-musicians in the temporal course of the BOLD response for the interconnected cortical areas of visual cortex including the calcarine fissure and the cuneus. This finding suggests a larger involvement of visual processes that might be related to imagery or even to a relaxation state in non-musicians accumulated and achieved as a consequence of listening emotionally-loaded music. The music categories, i.e., the combination of emotion and preference, that tend to elicit strong BOLD responses are also evidenced within all the three functional networks related to affective processing as detailed in Chapter 4.

In Affect 2 study, we aim to depict whether the intentionality has effects on the neural functional connectivity during affective processing of musical emotions and preference, considering the fact that listening to music is not always intentional. Towards this aim, an fMRI experiment was designed to incorporate three levels of attentional and intentional engagements with the music (liking judgment, gender judgment, and naturalistic listening). The proposed framework is employed on each scanning session corresponding to one particular level. We find that different levels of attentional and intentional engagements with the music have different effects on auditory-limbic connectivity during affective processing of music. A very novel finding of this neuroimaging study is the separation of functionally connected neural networks between the three experimental conditions. Then by the designed cluster topology interaction illustration, the differences and concordance between any two experimental conditions can be inspected. Notably, the functional network encompassing sensorimotor areas (such as the precentral and postcentral gyri and the supplementary motor area) found for all the three experimental conditions and particularly for the gender judgement condition, have been consistently observed in response to music-induced

emotions. In summary, the study proves that intentionality in judging the hedonic value of a musical piece is important in shaping neural connectivity to music, and specifically in connecting brain regions related to attention and cognition. In turn, when attention is focused on non-evaluative aspects of the music areas related to emotions and pleasures become more coupled. Considering these findings are based on music that are unfamiliar to the participants, future fMRI experiment might incorporate the highly familiar music to participants to investigate the functional brain connectivity under the different levels of attentional and intentional engagements with the music.

6.2 Future work

6.2.1 Scalable clustering algorithm for large-scale dataset with number of clusters automatically detected.

The very first step of consensus clustering analysis is the partition generation from the single clustering algorithms. The problems of excessive computational load and limited clustering performance arise in the context of large-scale dataset such as fMRI data. Although some classical algorithms such as K-means are fast, the results of clustering largely depend on the user-defined number of clusters K. A scalable clustering paradigm to address the aforementioned problems with number of clusters K automatically detected has been developed and published. Since it has not been extensively integrated into the main consensus clustering analysis of fMRI data framework in the thesis, the method is described and discussed here for further enhancement of the framework.

The basic logic within the method is sampling and combination. After sampling, the data scale becomes smaller but the data distribution information is still kept. Then for each sample of the data, a clustering method called E-SMART (enhanced splitting merging awareness tactics) (Fa et al., 2014, 2013) is applied. E-SMART can partition the data into clusters with number of clusters K automatically detected, but it suffers from poor performance on large datasets. But on smaller sampled data the speed is much quicker. The clustering results of each sampled dataset are then combined to form the final partition for the whole dataset. The procedure is detailed in next section.

6.2.1.1 Algorithm

a. Sampling

Sampling is drawn randomly from the original dataset at the rate $1/s$. Here, we use sampling without replacement, so all s sampled subsets are generated with each one contains absolutely different data points from others and the union of them is the original dataset. Theoretically, the smaller s is, the more information of original dataset is kept. We recommend to set s as small as possible as long as the hardware can deal with the size of sampled data using E-SMART in a reasonable time.

b. Combination

In total, s partitions $\{\boldsymbol{P}_i | i = 1, \dots, s\}$ are generated by E-SMART. Each partition $P_i (i = 1 \dots s)$ has its number of clusters detected as K_i, then there will be K_i cluster centroids which are denoted as $\{\boldsymbol{C}_k | k = 1, \dots, s\}$. To combine these intermediate results into a final partition, we use the following steps:

1) Put all the cluster centroids $\{\boldsymbol{C}_k | k = 1, \dots, s\}$ into a new dataset $\boldsymbol{C}$, which is the assembly of all the cluster centroids detected.
2) Cluster centroid set $\boldsymbol{C}$ is further clustered by hierarchical clustering with Ward linkage.
3) Calculate the cluster number K for the whole dataset was calculated as the mode of $\{K_i\}$, denoted as K_m.
4) Retrieve the clustering results of step (2) by choosing cluster number K_m, yielding a new partition for the centre set C, denoted as $\boldsymbol{P}_c$.
5) Calculate the mean of each cluster in P_c, which yielding K_m centre points.
6) For each datum in original dataset, assign it to its nearest centre from K_m centre point obtained in step (5) using Euclidian distance.

6.2.1.2 Validation

Synthetic dataset

The method is validated using several synthetic datasets, whose dimension is 200000×7, that mimic the size of the real fMRI datasets (228453×6or7or8) used in the thesis. When the data is generated, white Gaussian noise $(0, \sigma^2)$ with σ equal to 0.01, 0.1, 0.2 and 0.3 are added. We firstly use these synthetic datasets for a quantitative evaluation of the method's capability of detecting the number of clusters and general accuracy. E-SMART is applied on samples and the original synthetic data and then the results of our method are compared with those obtained by K-means, which is relatively fast in the classical clustering family. For K-means, we arbitrarily set an interval for number of clusters ranging from 45 to 55 and run K-means on each subset with all the K value. In the data with number of clusters unknown, the K needs to be chosen from a wider

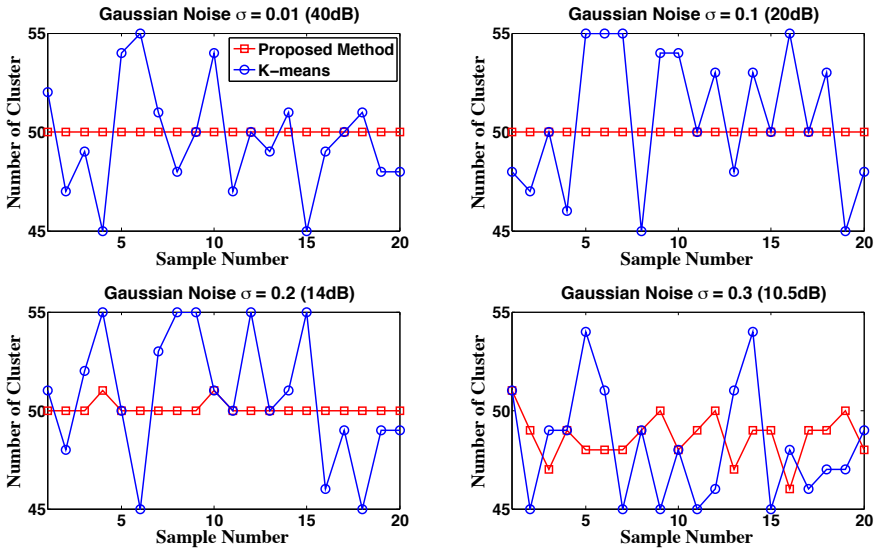


Figure 6-1. Number of clusters detected by E-SMART and k-means on each sample under different noise levels.

range. Silhouette index is used to evaluate the clustering results quality and determine the estimated K. The estimated K is chosen as the one that yields highest average Silhouette index. Then we compare the mode of these cluster numbers with the ground truth (K = 50) of the synthetic datasets. We use adjusted Rand index (ARI) (Rand, 1971) and normalized mutual information (NMI) (McDaid et al., 2011) as the metrics to evaluate the clustering membership accuracy on the synthetic dataset. The time we aim to compare includes two parts which are the time needed to specify appropriate cluster number and the actual execution time. The reason is that

here the detection of the number of clusters K is emphasised. So it is important to include the time needed to specifying the cluster number together with the execution time.

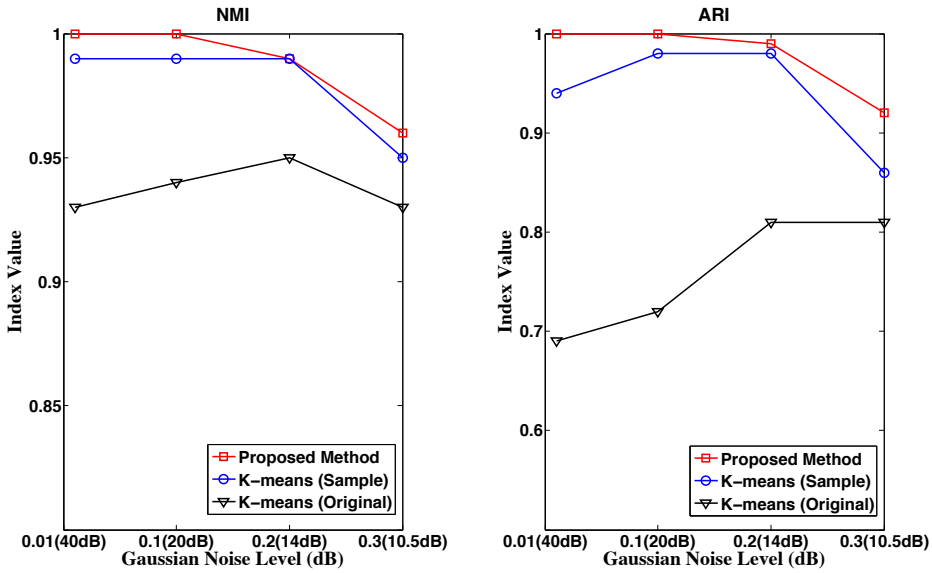


Figure 6-2. The normalized mutual information (NMI) and adjusted Rand index (ARI) comparison under different noise levels.

As shown in Figure 6-1, in the low noise level condition, the proposed method can detect the number of clusters correctly on all samples while the results from K-means fluctuate across the whole K value range we set. With the increased noise level, our method still generates more stable estimation of the number of clusters than that from K-means. The comparisons of the accuracies of assigning object membership are shown in Figure 6-2. We note that the proposed method has the perfect accuracy in the high SNR situation and very competent results compared to K-means on samples and K-means on original dataset under different noise level for the middle SNR situation, especially compared with the K-means on original data. Even when the SNR is low, the proposed method still achieves the highest accuracy both in NMI and ARI.

Table 6-1 compares the execution time of proposed method and K-means on the sampled subsets. Note the time for proposed method is the mean of the duration of the experiment on each of the 20 subsets. In a parallel manner, all these 20 subsets can be clustered simultaneously, so the total time is not calculated as the sum of the time on each subset. The K-means (Sample) is the estimated time of applying K-means with all the possible K values which should approximately range from 1 to $\sqrt{n/2}$ (~300 in this study) in this experiment. And the time values in K-means (Original) are the time for single run and evaluation on the whole synthetic dataset.

Table 6-1. Execution time on subsets under different noise level.

Noise Level	Method	Mean Time (sec)
$\sigma = 0.01$ $(40dB)$	Proposed method	277
	K-means (Sample)	5600
	K-means (Original)	9400
$\sigma = 0.1$ $(20dB)$	Proposed method	1700
	K-means (Sample)	6000
	K-means (Original)	9700
$\sigma = 0.2$ $(14dB)$	Proposed method	3050
	K-means (Sample)	6000
	K-means (Original)	9650
$\sigma = 0.3$ $(10.5dB)$	Proposed method	3400
	K-means (Sample)	6400
	K-means (Original)	10100

Real fMRI dataset

The real data come from an fMRI listening experiment related to the music emotions (Alluri et al., 2013; Brattico et al., 2011) carried out in the University of Helsinki. The whole fMRI experiment for one participant has 450 scans (TR=2s) including 32 music categories with each one repeated

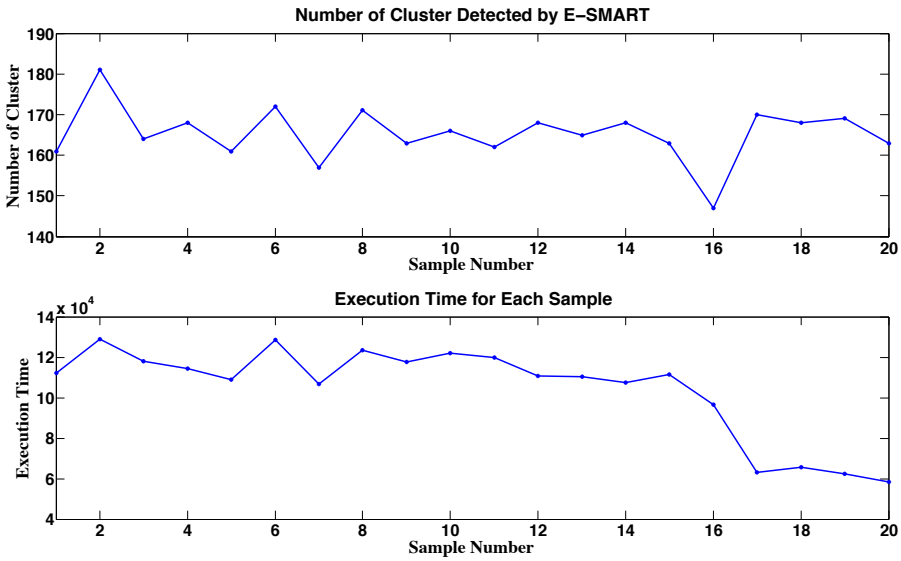


Figure 6-3. The number of clusters detected in real fMRI data and the execution time on each sample.

twice and each scan contains 228,453 voxels after preprocessing. In this paper, we use only one condition from one random subject in the experiment and apply our paradigm to it.

On the real fMRI data with no ground truth of the number of clusters, our method detects stable estimation of K which is around 170 (Figure 6-3). The execution time increases due to the large K in the data, compared with the time on samples with K equal to 50. But the speed is still competent compared with the case that use k-means to do the exhaustive search on real data.

6.2.1.3 Future application

One important feature of our proposed method is the experiments can be run simultaneously on multiple machines, as it does not need data communications between different sampled subsets before the combination. With the help of the power of distributed computation technique as used in this thesis, each worker node can handle more than one sample clustering tasks. So in ideal case, no matter how big the data is, the completion time for clustering the whole data is equal to the longest time needed by the algorithm for one sample. This algorithm is able to be extended in the future to do the random sampling repeatedly and combine all these clustering results, which would benefit from the diversity of the sampling, yielding more sound clustering results. Once this method is well established, the clustering results can be seen as the results from a single clustering algorithm and incorporated in the consensus clustering framework

6.2.2 From partition matrix to adjacency matrix

The consensus clustering framework in this thesis utilises the partition matrix as the medium for integrating clustering results from multiple algorithms on multiple datasets. Because the nature of partition matrix, partition matrices with different K have different number of rows (when row is used to represent the clusters). So they cannot be summed like how they are integrated with the same K. The adjacency matrix has been used in few fMRI data parcellation study using consensus clustering concept, although with only one clustering method and without tunable feature. By using adjacency matrix, clustering results from different K can be merged together as it does not include a matrix wise summation. Instead, the consensus is defined as how often two particular voxels belong to the same cluster in various clustering experiment. The following paragraph describes two main issues with adjacency matrix approach and their possible solution.

One is the high requirement for RAM as the size of adjacency matrix is $N \times N$ where N is the number of voxels. For the dataset in this thesis, it would be a 228453×228453 matrix, resulting in around 400GB for just storing the adjacency matrix in the RAM when the type of each element is stored as double type (8 Byte per value). The demand for memory can be reduced by using sparse matrix. However, the actual RAM needed might still be very large when a clustering with small K is used, i.e., each voxel and a large portion of the total voxels have been grouped in the same cluster. A simple and straightforward solution is using high performance computing facility with large amount of memory. Another solution is analysing the functional data parcellation-wise rather than voxel-wise, thus the value of N is reduced. Another issue is the consensus adjacency matrix needs to be clustered again to yield final clustering solutions. In the existing literature (Bellec et al., 2010; Kelly et al., 2012; Orban et al., 2015; Ryali et al., 2015), this is done by using K-means, hierarchical clustering, or spectral clustering on the consensus adjacency matrix to generate final clusters. Note this step introduces new free parameters such as number of clusters for the consensus adjacency matrix. Following the binarisation logic within the consensus clustering framework in this thesis, the final consensus adjacency matrix can be binarised to yield the binarised matrix, where the elements/membership are either 0 or 1, with respect to different binarisation threshold (e.g., from 0.1 to 1 at a step of 0.1). Then these binarised adjacency matrices can be treated as unweighted graphs and further clustered using community detection algorithm without specifying the number of clusters K.

In summary, by introducing the adjacency matrix as the medium of merging results from multiple clustering experiments, clustering solutions from different number of clusters K can be integrated. This can be further explored as an extension to the existing framework.

6.3 Closing comments

We believe that the developed consensus clustering framework for analysing brain functional connectivity is a progress of addressing the inconsistency and stability issues among the many clustering algorithms. Also, this framework is data-driven, providing the capability of dealing with functional MRI data from more and more complex experiment designs, where traditional model-based methods might fail or be very difficult to implement. In addition to the framework development, the exploration and investigation of two real fMRI data yield many novel findings in cognitive

neuroscience of music, more specifically, the affective processing of music. This work greatly enriches the fMRI data analysis strategies as well as the literature on understanding brain functions in the context of neuroscience.

Chapter 7 Reference

Abbott, A., 2009. Brain imaging studies under fire. Nature 457, 245. doi:Doi 10.1038/457245a

Abu-Jamous, B., Fa, R., Roberts, D.J., Nandi, A.K., 2015a. Application of the Bi-CoPaM method to five Escherichia Coli datasets generated under various biological conditions. J. Signal Process. Syst. 79, 159–166. doi:10.1007/s11265-014-0919-7

Abu-Jamous, B., Fa, R., Roberts, D.J., Nandi, A.K., 2015b. UNCLES: method for the identification of genes differentially consistently co-expressed in a specific subset of datasets. BMC Bioinformatics 16, 184. doi:10.1186/s12859-015-0614-0

Abu-Jamous, B., Fa, R., Roberts, D.J., Nandi, A.K., 2014. Comprehensive analysis of forty yeast microarray datasets reveals a novel subset of genes (APha-RiB) consistently negatively associated with ribosome biogenesis. BMC Bioinformatics 15, 322. doi:10.1186/1471-2105-15-322

Abu-Jamous, B., Fa, R., Roberts, D.J., Nandi, A.K., 2013. Paradigm of tunable clustering using Binarization of Consensus Partition Matrices (Bi-CoPaM) for gene discovery. PLoS One 8, e56432. doi:10.1371/journal.pone.0056432

Ahmadlou, M., Adeli, A., Bajo, R., Adeli, H., 2014. Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task. Clin. Neurophysiol. 125, 694–702. doi:10.1016/j.clinph.2013.08.033

Ahmadlou, M., Adeli, H., 2011. Functional community analysis of brain: a new approach for EEG-based investigation of the brain pathology. Neuroimage 58, 401–408. doi:10.1016/j.neuroimage.2011.04.070

Ahmadlou, M., Adeli, H., Adeli, a., 2012. Graph theoretical analysis of organization of functional brain networks in ADHD. Clin. EEG Neurosci. 43, 5–13. doi:10.1177/1550059411428555

Ahmadlou, M., Hojjat, A., 2010. Enhanced probabilistic neural network with local decision circles: a robust classifier. Integr. Comput. Aided. Eng. 17, 197–210.

Ali, S.S., Lifshitz, M., Raz, A., 2014. Empirical neuroenchantment: from reading minds to thinking critically. Front. Hum. Neurosci. 8, 357. doi:10.3389/fnhum.2014.00357

Alluri, V., Brattico, E., Toiviainen, P., Burunat, I., Bogert, B., Numminen, J., Kliuchko, M., 2015. Musical expertise modulates functional connectivity of limbic regions during continuous music listening. Psychomusicology Music. Mind, Brain 25, 443–454.

Alluri, V., Toiviainen, P., Jääskeläinen, 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, 3677–3689. doi:10.1016/j.neuroimage.2011.11.019

Alluri, V., Toiviainen, P., Lund, T.E., Wallentin, M., Vuust, P., Nandi, A.K., Ristaniemi, T., Brattico, E., 2013. From vivaldi to beatles and back: Predicting lateralized brain responses to music. Neuroimage 83, 627–636. doi:10.1016/j.neuroimage.2013.06.064

Ayad, H.G., Kamel, M.S., 2010. On voting-based consensus of cluster ensembles. Pattern Recognit. 43, 1943–1953. doi:10.1016/j.patcog.2009.11.012

Barrett, L.F., 2006. Are emotions natural kinds? Perspect. Psychol. Sci. 1, 28–58. doi:10.1111/j.1745-6916.2006.00003.x

Barrett, L.F., Wager, T.D., 2006. The structure of emotion evidence from neuroimaging studies. Curr. Dir. Psychol. Sci. 15, 79–83. doi:10.1111/j.0963-7214.2006.00411.x

Bassett, D.S., Bullmore, E., 2006. Small-world brain networks. Neuroscientist 12, 512–523. doi:10.1177/1073858406293182

Baumgartner, R., Windischberger, C., Moser, E., 1998. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. Magn Reson Imaging 16, 115–125.

Bellec, P., Perlbarg, V., Jbabdi, S., Pélégrini-Issac, M., Anton, J.-L., Doyon, J., Benali, H., 2006. Identification of large-scale networks in the brain using fMRI. Neuroimage 29, 1231–1243. doi:10.1016/j.neuroimage.2005.08.044

Bellec, P., Rosa-Neto, P., Lyttelton, O.C., Benali, H., Evans, A.C., 2010. Multi-level bootstrap

analysis of stable clusters in resting-state fMRI. Neuroimage 51, 1126–1139. doi:10.1016/j.neuroimage.2010.02.082

Bello-Orgaz, G., Menéndez, H.D., Camacho, D., 2012. Adaptive K-Means algorithm for overlapped graph clustering. Int. J. Neural Syst. 22, 1250018. doi:10.1142/S0129065712500189

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. 1191, 133–55. doi:10.1111/j.1749-6632.2010.05446.x

Bezdek, J.C., 1983. Pattern recognition with fuzzy objective function algorithms. SIAM Rev. 25, 442–442. doi:10.1137/1025116

Biswal, B., Van Kylen, J., Hyde, J.S., 1997. Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. NMR Biomed. 10, 165–170. doi:10.1002/(sici)1099-1492(199706/08)10:4/5%3C165::aid-nbm454%3E3.0.co;2-7

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. 34, 537–541. doi:10.1002/mrm.1910340409

Blood, a J., Zatorre, R.J., 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc. Natl. Acad. Sci. U. S. A. 98, 11818–23. doi:10.1073/pnas.191355898

Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J., Smith, S.M., 2013. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage 76, 313–324. doi:10.1016/j.neuroimage.2013.03.024

Bogert, B., Numminen-Kontti, T., Gold, B., Sams, M., Numminen, J., Burunat, I., Lampinen, J., Brattico, E., 2016. Hidden sources of joy, fear, and sadness: Explicit versus implicit neural processing of musical emotions. Neuropsychologia 89, 393–402. doi:10.1016/j.neuropsychologia.2016.07.005

Brattico, E., 2015. From pleasure to liking and back: Bottom-up and top-down neural routes to the aesthetic enjoyment of music., in: Art, Aesthetics, and the Brain. Oxford University Press,

New York, pp. 303–318.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., Tervaniemi, M., 2011. A functional MRI study of happy and sad emotions in music with and without lyrics. Front. Psychol. 2, 308. doi:10.3389/fpsyg.2011.00308

Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., 2016. It ' s sad but I like It : the neural dissociation between musical emotions and liking in experts and laypersons. Front. Hum. Neurosci. 9. doi:10.3389/fnhum.2015.00676

Brattico, E., Bogert, B., Jacobsen, T., 2013. Toward a neural chronometry for the aesthetic experience of music. Front. Psychol. 4, 206. doi:10.3389/fpsyg.2013.00206

Brattico, E., Pearce, M., 2013. The neuroaesthetics of music. Psychol. Aesthetics, Creat. Arts 7, 48–61. doi:10.1037/a0031624

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290. doi:10.1016/j.tics.2010.04.004

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S.B., Liotti, M., 2011. Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. Neuroimage 58, 250–258. doi:10.1016/j.neuroimage.2011.06.012

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–98. doi:10.1038/nrn2575

Bullmore, E.T., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198. doi:10.1038/nrn2575

Bundgaard, P.F., 2015. Feeling, meaning, and intentionality - a critique of the neuroaesthetics of beauty. Phenomenol. Cogn. Sci. 14, 781–801. doi:10.1007/s11097-014-9351-5

Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., Toiviainen, P., Snyder, J., 2015. Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. PLoS One 10, 1–18. doi:10.1371/journal.pone.0138238

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. doi:10.1016/j.neuroimage.2015.09.005

Button, K.S., Ioannidis, J.P. a, Mokrysz, C., Nosek, B. a, Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–76. doi:10.1038/nrn3475

Chapin, H., Jantzen, K., Kelso, J.A.S., Steinberg, F., Large, E., 2010. Dynamic emotional and neural responses to music depend on performance expression and listener experience. PLoS One 5. doi:10.1371/journal.pone.0013812

Chatterjee, A., Vartanian, O., 2016. Neuroscience of aesthetics. Ann. N. Y. Acad. Sci. 1369, 172–194. doi:10.1111/nyas.13035

Chen, H., Li, K., Zhu, D., Jiang, X., Yuan, Y., Lv, P., Zhang, T., Guo, L., Shen, D., Liu, T., 2013. Inferring group-wise consistent multimodal brain networks via multi-view spectral clustering. IEEE Trans. Med. Imaging 32, 1576–1586. doi:10.1109/TMI.2013.2259248

Chen, H., Yuan, H., Yao, D., Chen, L., Chen, W., 2006. An integrated neighborhood correlation and hierarchical clustering approach of functional MRI. IEEE Trans. Biomed. Eng. 53, 452–458. doi:10.1109/TBME.2005.869660

Chuang, K.H., Chiu, M.J., Lin, C.C., Chen, J.H., 1999. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means. IEEE Trans. Med. Imaging 18, 1117–1128. doi:10.1109/42.819322

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70, 66111. doi:10.1103/PhysRevE.70.066111

Cong, F., Alluri, V., Lin, Q., Toiviainen, P., Nandi, A.K., Brattico, E., Ristaniemi, T., 2013. Semi-blind independent component analysis of functional MRI elicited by continuous listening to music, in: ICASSP 2013. pp. 1310–1314.

Cordes, D., Haughton, V., Carew, J.D., Arfanakis, K., Maravilla, K., 2002. Hierarchical clustering to measure connectivity in fMRI resting-state data. Magn. Reson. Imaging 20, 305–317. doi:10.1016/S0730-725X(02)00503-9

Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., Meyerand, M.E., 2001. Frequencies contributing to functional connectivity in the cerebral cortex in 'resting-state' data. Am. J. Neuroradiol. 22, 1326–1333.

Cordes, D., Haughton, V.M., Arfanakis, K., Wendt, G.J., Turski, P.A., Moritz, C.H., Quigley, M.A., Meyerand, M.E., 2000. Mapping functionally related regions of brain with functional connectivity MR imaging. Am. J. Neuroradiol. 21, 1636–1644.

Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. 33, 1914–1928. doi:10.1002/hbm.21333

Crossley, N. A, Mechelli, A., Vértes, P.E., Winton-brown, T.T., Patel A. X., Ginestet C. E., McGuire, P. , Bullmore, E. T., 2013., Cognitive relevance of the community structure of the human brain functional coactivation network. Proc. Natl. Acad. Sci. 110, 15502–15502. doi:10.1073/pnas.1314559110

Damasio, a R., Grabowski, T.J., Bechara, a, Damasio, H., Ponto, L.L., Parvizi, J., Hichwa, R.D., 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat. Neurosci. 3, 1049–56. doi:10.1038/79871

DeGroot, M.H., 1987. A conversation with George Box. Stat. Sci. 2, 239–258. doi:10.1214/ss/1177013223

Dunn, J.C., 1974. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32–57. doi:10.1080/01969727308546046

Eickhoff, S.B., Bzdok, D., Laird, A.R., Roski, C., Caspers, S., Zilles, K., Fox, P.T., 2011. Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. Neuroimage 57, 938–949. doi:10.1016/j.neuroimage.2011.05.021

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. 201602413. doi:10.1073/pnas.1602413113

Ellison, D., Brattico, E., 2015. Affective versus cognitive responses to musical chords : An ERP

and behavioral study. Psychomusicology Music , Mind , Brain 25, 423–434.

Estivill-Castro, V., 2002. Why so many clustering algorithms. ACM SIGKDD Explor. Newsl. 4, 65–75. doi:10.1145/568574.568575

Evans, A., Kamber, M., Collins, D., MacDonald, D., 1994. Magnetic resonance scanning and epilepsy. Plenum Press, New York.

Evans, A.C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., Milot, S., Meyer, E., Bub, D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage 1, 43–53. doi:10.1016/1053-8119(92)90006-9

Fa, R., Abu-Jamous, B., Roberts, D.J., Nandi, A.K., 2013. Enhanced SMART framework for gene clustering using successive processing. 2013 IEEE Int. Work. Mach. Learn. Signal Process. 1–6. doi:10.1109/MLSP.2013.6661964

Fa, R., Roberts, D.J., Nandi, A.K., 2014. SMART: Unique Splitting-while-merging framework for gene clustering. PLOS One 9, e94141. doi:10.1371/journal.pone.0094141

Fadili, M.J., Ruan, S., Bloyet, D., Mazoyer, B., 2001. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. Med. Image Anal. 5, 55–67. doi:10.1016/S1361-8415(00)00035-9

Ferrarini, L., Veer, I.M., Baerends, E., Van Tol, M.J., Renken, R.J., Van Der Wee, N.J. a, Veltman, D.J., Aleman, A., Zitman, F.G., Penninx, B.W.J.H., Van Buchem, M. a., Reiber, J.H.C., Rombouts, S. a R.B., Milles, J., 2009. Hierarchical functional modularity in the resting-state human brain. Hum. Brain Mapp. 30, 2220–2231. doi:10.1002/hbm.20663

Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., Poline, J.-B., 2002. Improved detection sensitivity in functional MRI data using a brain parcelling technique, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI'02). pp. 467–474.

Flores-Gutiérrez, E.O., Díaz, J.L., Barrios, F.A., Favila-Humara, R., Guevara, M.Á., del Río-Portilla, Y., Corsi-Cabrera, M., 2007. Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. Int. J. Psychophysiol. 65, 69–84. doi:10.1016/j.ijpsycho.2007.03.004

Fortunato, S., 2009. Community detection in graphs. doi:10.1016/j.physrep.2009.11.002

Friston, K.J., Frith, C.D., Dolan, R.J., Price, C.J., Zeki, S., Ashburner, J.T., Penny, W.D., 2003. Human Brain Function. Academic Press.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? Neuroimage 10, 1–5. doi:10.1006/nimg.1999.0439

Friston, K.J., Holmes, a. P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210. doi:10.1002/hbm.460020402

Gao, J.H., Yee, S.H., 2003. Iterative temporal clustering analysis for the detection of multiple response peaks in fMRI. Magn. Reson. Imaging 21, 51–53. doi:10.1016/S0730-725X(02)00627-6

Ghaemi, R., Sulaiman, N., Ibrahim, H., Mustapha, N., 2009. A Survey : clustering ensembles techniques. Eng. Technol. 38, 636–645.

Ghosh, S.S., Keshavan, A., Langs, G., 2013. Predicting treatment response from resting state fMRI data: comparison of parcellation approaches. 2013 Int. Work. Pattern Recognit. Neuroimaging 225–228. doi:10.1109/PRNI.2013.64

Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., Boesiger, P., 1998. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn. Reson. Med. 40, 249–260. doi:10.1002/mrm.1910400211

Golland, P., Golland, Y., Malach, R., 2007. Detection of spatial activation patterns as unsupervised segmentation of fMRI data. Med. Image Comput. Comput. Assist. Interv. 10, 110–8. doi:10.1007/978-3-540-75757-3

Gosselin, N., Peretz, I., Johnsen, E., Adolphs, R., 2007. Amygdala damage impairs emotion recognition from music. Neuropsychologia 45, 236–244.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F., Hansen, L.K., 1999. On clustering fMRI time series. Neuroimage 9, 298–310. doi:10.1006/nimg.1998.0391

Green, A.C., Bærentsen, K.B., Stodkilde-Jorgensen, H., Roepstorff, A., Vuust, P., 2012. Listen, learn, like! Dorsolateral prefrontal cortex involved in the mere exposure effect in music. Neurol. Res. Int. 2012. doi:10.1155/2012/846270

Hargreaves, D.J., MacDonald, R., Miell, D., 2012. How do people communicate using music?, in: Musical Communication. Oxford University Press, Oxford-New York, pp. 1–24. doi:10.1093/acprof:oso/9780198529361.003.0001

Hasson, U., Furman, O., Clark, D., Dudai, Y., Davachi, L., 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57, 452–462. doi:10.1016/j.neuron.2007.12.009

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–40. doi:10.1126/science.1089506

Haykin, S., 1999. Neural Networks - A Comprehensive Foundation, Third Edit. ed. Singapore: Pearson, Prentice Hall.

Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Curr. Biol. 17, 323–328. doi:10.1016/j.cub.2006.11.072

Hodges, D.A., 2016. The neuroaesthetics of music, in: The Oxford Handbook of Music Psychology. Oxford University Press, Oxford, pp. 247–262.

Höfel, L., Jacobsen, T., 2007a. Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes? Int. J. Psychophysiol. 65, 20–31. doi:10.1016/j.ijpsycho.2007.02.007

Höfel, L., Jacobsen, T., 2007b. Electrophysiological indices of processing symmetry and aesthetics: a result of judgment categorization or judgment report? J. Psychophysiol. 21, 9–21. doi:10.1027/0269-8803.21.1.9

Huettel, S.A., Song, A.W., McCarthy, G., 2009. Functional Magnetic Resonance Imaging, Second Edi. ed, Magnetic Resonance Imaging. Sinauer Associates.

Jacobsen, T., 2014. Domain specificity and mental chronometry in empirical aesthetics. Br. J.

Psychol. 105, 471–473. doi:10.1111/bjop.12094

Jacobsen, T., Höfel, L., 2003. Descriptive and evaluative judgment processes: Behavioral and electrophysiological indices of processing symmetry and aesthetics. Cogn. Affect. Behav. Neurosci. 3, 289–299. doi:10.3758/CABN.3.4.289

James, C.E., Britz, J., Vuilleumier, P., Hauert, C.A., Michel, C.M., 2008. Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. Neuroimage 42, 1597–1608. doi:10.1016/j.neuroimage.2008.06.025

Jeffries, K.J., Fritz, J.B., Braun, A.R., 2003. Words in melody: an H-2 O-15 PET study of brain activation during singing and speaking. Neuroreport 14, 749–754. doi:10.1097/00001756-200304150-00018

Johnson, S.C., 1967. Hierarchical clustering schemes. Psychometrika 32, 241–254. doi:10.1007/BF02289588

Jones, E.G., Mendell, L.M., 1999. Assessing the Decade of the Brain. Science. 284, 739.

Juslin, P.N., 2013. From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. Phys. Life Rev. 10, 235–266. doi:10.1016/j.plrev.2013.05.008

Juslin, P.N., Västfjäll, D., 2008. Emotional responses to music: the need to consider underlying mechanisms. Behav. Brain Sci. 31, 559–621. doi:10.1017/S0140525X08005293

Kahnt, T., Chang, L.J., Park, S.Q., Heinzle, J., Haynes, J.-D., 2012. Connectivity-based parcellation of the human orbitofrontal cortex. J. Neurosci. 32, 6240–6250. doi:10.1523/JNEUROSCI.0257-12.2012

Karmonik, C., Brandt, A.K., Fung, S.H., Grossman, R.G., Frazier, J.T., 2013. Graph theoretical connectivity analysis of the human brain while listening to music with emotional attachment: feasibility study. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2013, 6526–9. doi:10.1109/EMBC.2013.6611050

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, New York.

Kaufman, L., Rousseeuw, P.J., 1987. Clustering by means of medoids. Stat. Data Anal. Based L 1-Norm Relat. Methods. First Int. Conf.

Kelly, C., Toro, R., Di Martino, A., Cox, C.L., Bellec, P., Castellanos, F.X., Milham, M.P., 2012. A convergent functional architecture of the insula emerges across imaging modalities. Neuroimage 61, 1129–1142. doi:10.1016/j.neuroimage.2012.03.021

Khalfa, S., Schon, D., Anton, J.-L., Liégeois-Chauvel, C., 2005. Brain regions involved in the recognition of happiness and sadness in music. Neuroreport 16, 1981–1984. doi:10.1097/00001756-200512190-00002

Koelsch, S., 2014. Brain correlates of music-evoked emotions. Nat. Rev. Neurosci. 15, 170–180. doi:10.1038/nrn3666

Koelsch, S., 2010. Towards a neural basis of music-evoked emotions. Trends Cogn. Sci. 14, 131–7. doi:10.1016/j.tics.2010.01.002

Koelsch, S., Fritz, T., Cramon, V., Yves, D., Müller, K., Friederici, A.D., 2006. Investigating emotion with music: an fMRI study. Hum. Brain Mapp. 27, 239–250. doi:10.1002/hbm.20180

Koelsch, S., Skouras, S., 2014. Functional centrality of amygdala, striatum and hypothalamus in a 'small-world' network underlying joy: an fMRI study with music. Hum. Brain Mapp. 35, 3485–3498. doi:10.1002/hbm.22416

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69. doi:10.1007/BF00337288

Kriegeskorte, N., Lindquist, M. a, Nichols, T.E., Poldrack, R. a, Vul, E., 2010. Everything you never wanted to know about circular analysis, but were afraid to ask. J. Cereb. Blood Flow Metab. 30, 1551–1557. doi:10.1038/jcbfm.2010.86

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540. doi:10.1167/8.6.88

Kühn, S., Gallinat, J., 2012. The neural correlates of subjective pleasantness. Neuroimage 61,

289–294. doi:10.1016/j.neuroimage.2012.02.065

Kumar, A., Daumé, H., 2011. A co-training approach for multi-view spectral clustering. Proc. 28th Int. Conf. Mach. Learn. 393–400.

Lange, N., Zeger, S.L., 1997. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. J. R. Stat. Soc. Ser. C (Applied Stat). 46, 1–29. doi:10.1111/1467-9876.00046

Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., Golland, P., 2012. Search for patterns of functional specificity in the brain: a nonparametric hierarchical Bayesian model for group fMRI data. Neuroimage 59, 1348–68. doi:10.1016/j.neuroimage.2011.08.031

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. Discovering structure in the space of fMRI selectivity profiles. Neuroimage 50, 1085–1098. doi:10.1016/j.neuroimage.2009.12.106

Liao, T.W., 2005. Clustering of time series data—a survey. Pattern Recognit. 38, 1857–1874. doi:10.1016/j.patcog.2005.01.025

Liao, W., Chen, H., Yang, Q., Lei, X., 2008. Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. IEEE Trans. Med. Imaging 27, 1472–1483. doi:10.1109/TMI.2008.923987

Lichtman, J.W., Pfister, H., Shavit, N., 2014. The big data challenges of connectomics. Nat Neurosci 17, 1448–1454. doi:10.1038/nn.3837

Lindquist, K. a, Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: a meta-analytic review. Behav. Brain Sci. 35, 121–143. doi:10.1017/S0140525X11000446

Liu, C., Abu-jamous, B., Brattico, E., Nandi, A., 2015a. Clustering consistency in neuroimaging data analysis, in: Fuzzy Systems and Knowledge Discovery (FSKD). pp. 1118–1122.

Liu, C., Abu-Jamous, B., Brattico, E., Nandi, A.K., 2017. Towards tunable consensus clustering for studying functional brain connectivity during affective processing. Int. J. Neural Syst. 27, 1650042. doi:10.1142/S0129065716500428

Liu, C., Fa, R., Abu-Jamous, B., Brattico, E., Nandi, A., 2015b. Scalable clustering based on enhanced-SMART for large-scale FMRI datasets. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2015–Augus, 962–966. doi:10.1109/ICASSP.2015.7178112

Liu, Y., Gao, J.H., Liu, H.L., Fox, P.T., 2000. The temporal response of the brain after eating revealed by functional MRI. Nature 405, 1058–1062. doi:10.1038/35016590

Luo, C., Tu, S., Peng, Y., Gao, S., Li, J., Dong, L., Li, G., Lai, Y., Li, H., Yao, D., 2014. Long-term effects of musical training and functional plasticity in salience system. Neural Plasticity, 2014, 180138.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of 5-Th Berkeley Symposium on Mathematical Statistics and Probability. pp. 281–297.

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R.J., Marco-Pallarés, J., 2016. Neural correlates of specific musical anhedonia. Proc Natl Acad Sci U S A. doi:doi:10.1073/pnas.1611211113

McDaid, A.F., Greene, D., Hurley, N., 2011. Normalized mutual information to evaluate overlapping community finding algorithms. Arxiv Prepr. arXiv1110.2515 1–3.

Meila, M., 2007. Comparing clusterings-an information based distance. J. Multivar. Anal. 98, 873–895. doi:10.1016/j.jmva.2006.11.013

Meyer, F.G., Chinrungrueng, J., 2005. Spatiotemporal clustering of fMRI time series in the spectral domain. Med. Image Anal. 9, 51–68. doi:10.1016/j.media.2004.07.002

Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., Thirion, B., 2012. A supervised clustering approach for fMRI-based inference of brain states. Pattern Recognit. 45, 2041–2049. doi:10.1016/j.patcog.2011.04.006

Mitterschiffthaler, M.T., Fu, C.H.Y., Dalton, J. a., Andrew, C.M., Williams, S.C.R., 2007. A functional MRI study of happy and sad affective states induced by classical music. Hum. Brain Mapp. 28, 1150–1162. doi:10.1002/hbm.20337

Morin, O., Grèzes, J., 2008. What is 'mirror' in the premotor cortex? A review. Neurophysiol. Clin. 38, 189–195. doi:10.1016/j.neucli.2008.02.005

Moser, E., Diemling, M., Baumgartner, R., 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. J. Magn. Reson. Imaging 7, 1102–1108.

Mueller, K., Fritz, T., Mildner, T., Richter, M., Schulze, K., Lepsien, J., Schroeter, M.L., Möller, H.E., 2015. Investigating the dynamics of the brain response to music: A central role of the ventral striatum/nucleus accumbens. Neuroimage 116, 68–79. doi:10.1016/j.neuroimage.2015.05.006

Newman, M., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. 103, 8577–82. doi:10.1073/pnas.0601602103

Nieminen, S., Istók, E., Brattico, E., Tervaniemi, M., Huotilainen, M., 2011. The development of aesthetic responses to music and their underlying neural and psychological mechanisms. Cortex 47, 1138–1146. doi:10.1016/j.cortex.2011.05.008

Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. 87, 9868–9872. doi:10.1073/pnas.87.24.9868

Orban, P., Doyon, J., Petrides, M., Mennes, M., Hoge, R., Bellec, P., 2015. The richness of task-evoked hemodynamic responses defines a pseudohierarchy of functionally meaningful brain networks. Cereb. Cortex 25, 2658–2669. doi:10.1093/cercor/bhu064

Pauling, L., Coryell, C.D., 1936. The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. Proc. Natl. Acad. Sci. U. S. A. 22, 210–216. doi:10.1073/pnas.22.4.210

Peciña, S., Smith, K.S., Berridge, K.C., 2006. Hedonic hot spots in the brain. Neuroscientist 12, 500–511. doi:10.1177/1073858406293154

Peltier, S.J., Polk, T. a, Noll, D.C., 2003. Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm. Hum Brain Mapp 20, 220–226. doi:10.1002/hbm.10144

Peña, D., 2001. George Box: An interview with the International Journal of Forecasting. Int. J. Forecast. 17, 1–9. doi:10.1016/S0169-2070(00)00061-3

Peña, J.., Lozano, J.., Larrañaga, P., 1999. An empirical comparison of four initialization methods for the K-Means algorithm. Pattern Recognit. Lett. 20, 1027–1040. doi:10.1016/S0167-8655(99)00069-0

Poline, J.-B., Brett, M., 2012. The general linear model and fMRI: does love last forever? Neuroimage 62, 871–80. doi:10.1016/j.neuroimage.2012.01.133

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J. a, Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. Neuron 72, 665–78. doi:10.1016/j.neuron.2011.09.006

Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. 66, 846–850. doi:10.2307/2284239

Reybrouck, M., Brattico, E., 2015. Neuroplasticity beyond Sounds: neural adaptations following long-term musical aesthetic experiences. Brain Sci. 5, 69–91. doi:10.3390/brainsci5010069

Rice, J.A., 2007. Mathematical Statistics and Data Analysis. Duxbury Press.

Riecker, a, Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport 11, 1997–2000. doi:10.1097/00001756-200006260-00038

Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Cogn. Brain Res. 3, 131–141. doi:10.1016/0926-6410(95)00038-0

Romesburg, C., 2004. Cluster Analysis for Researchers. LuLu Press, Morrisville, NC.

Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U. S. A. 105, 1118–1123. doi:10.1073/pnas.0706851105

Ryali, S., Chen, T., Padmanabhan, A., Cai, W., Menon, V., 2015. Development and validation of consensus clustering-based framework for brain segmentation using resting fMRI. J.

Neurosci. Methods 240, 128–140. doi:10.1016/j.jneumeth.2014.11.014

Saarikallio, S., Nieminen, S., Brattico, E., 2012. Affective reactions to musical stimuli reflect emotional use of music in everyday life. Music. Sci. 17, 27–39. doi:10.1177/1029864912462381

Sachs, M.E., Ellis, R.J., Schlaug, G., Loui, P., 2016. Brain connectivity reflects human aesthetic responses to music. Soc. Cogn. Affect. Neurosci. 11, 884–891. doi:10.1093/scan/nsw009

Salimpoor, V.N., van den Bosch, I., Kovacevic, N., McIntosh, A.R., Dagher, A., Zatorre, R.J., 2013. Interactions between the nucleus accumbens and auditory cortices predict music reward value. Science 340, 216–9. doi:10.1126/science.1231059

Salimpoor, V.N., Zatorre, R.J., 2013. Neural interactions that give rise to musical pleasure. Psychol. Aesthetics, Creat. Arts 7, 62–75. doi:10.1037/a0031819

Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E., 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. doi:10.1093/cercor/bhi016

Satoh, M., Nakase, T., Nagata, K., Tomimoto, H., 2011. Musical anhedonia: selective loss of emotional experience in listening to music. Neurocase 17, 410–417. doi:10.1080/13554794.2010.532139

Schwarz, A.J., Gozzi, A., Bifone, A., 2009. Community structure in networks of functional connectivity: Resolving functional organization in the rat brain with pharmacological MRI. Neuroimage 47, 302–311. doi:10.1016/j.neuroimage.2009.03.064

Shi, L.-D., Shi, Y.-H., Gao, Y., Shang, L., Yang, Y.-B., 2011. XCSc: a novel approach to clustering with extended classifier system. Int. J. Neural Syst. 21, 79–93. doi:10.1142/S0129065711002675

Sloboda, J.A., 2010. Music in everyday life: The role of emotions, in: Handbook of Music and Emotion: Theory, Research, Applications. Oxford University Press, New York, pp. 493–514.

Sloboda, J.A., O'Neill, S.A., 2001. Emotions in everyday listening to music., in: Music and

Emotion: Theory and Research. pp. 415–429. doi:10.1017/CBO9781107415324.004

Sporns, O., 2014. Contributions and challenges for network models in cognitive neuroscience. Nat. Neurosci. 17, 652–660. doi:10.1038/nn.3690

Stanberry, L., Nandy, R., Cordes, D., 2003. Cluster analysis of fMRI data using dendrogram sharpening. Hum. Brain Mapp. 20, 201–219. doi:10.1002/hbm.10143

Strehl, A., Ghosh, J., 2002. Cluster ensembles – a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3, 583–617. doi:10.1162/153244303321897735

Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain. Thieme Medical Publishers, New York.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., 2006. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27, 678–693. doi:10.1002/hbm.20210

Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.-B., 2014. Which fMRI clustering gives good brain parcellations? Front. Neurosci. 8, 1–13. doi:10.3389/fnins.2014.00167

Tomasi, D., Volkow, N.D., 2012. Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. Biol. Psychiatry 71, 443–50.

Tucholka, A., Thirion, B., Perrot, M., Pinel, P., Mangin, J.-F., Poline, J., 2008. Probabilistic anatomo-functional parcellation of the cortex: how many regions? Med. Image Comput. Comput. Interv. MICCAI 2008 11th Int. Conf.11, 399–406. doi:10.1007/978-3-540-85990-1

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289. doi:10.1006/nimg.2001.0978

Valente de Oliveira, J., Pedrycz, W., 2007. Advances in fuzzy clustering and its applications, advances in fuzzy clustering and its applications. John Wiley & Sons. doi:10.1002/9780470061190

Van den Heuvel, M.P., Mandl, R., Hulshoff Pol, H., 2008. Normalized cut group clustering of resting-state FMRI data. PLoS One 3, e2001. doi:10.1371/journal.pone.0002001

Van den Heuvel, M.P., Pol, H.E.H., 2010. Exploring the brain network: A review on resting-state fMRI functional connectivity. Eur. Neuropsychopharmacol. 20, 519–534. doi:10.1016/j.euroneuro.2010.03.008

Van den Heuvel, M.P., Stam, C.J., Boersma, M., Hulshoff Pol, H.E., 2008. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage 43, 528–39. doi:10.1016/j.neuroimage.2008.08.010

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, a., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, a. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, a. Z., Xu, J., Yacoub, E., 2012. The Human Connectome Project: A data acquisition perspective. Neuroimage 62, 2222–2231. doi:10.1016/j.neuroimage.2012.02.018

Venkataraman, A., Dijk, K.R. a Van, Buckner, R.L., Golland, P., 2013. Exploring functional connectivity in fMRI via clustering. Proc IEEE Int Conf Acoust Speech Signal Process. 2009, 441-444

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Psychol. Sci. 4, 274–290. doi:10.1111/j.1745-6924.2009.01132.x

Wang, D., Shi, L., Yeung, D.S., Heng, P.-A., Wong, T.-T., Tsang, E.C.C., 2005. Support vector clustering for brain activation detection. Med. Image Comput. Comput. Assist. Interv. 8, 572–579. doi:10.1162/15324430260185565

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. doi:10.1080/01621459.1963.10500845

Welvaert, M., Rosseel, Y., 2013. On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. PLoS One 8, e77089. doi:10.1371/journal.pone.0077089

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., Grodd, W., 1996. Functional lateralization of speech production at primary motor cortex: a fMRI study. Neuroreport 7, 2791–2795.

Wilkins, R.W., Hodges, D.A., Laurienti, P.J., Steen, M., Burdette, J.H., 2014. Network science and the effects of music preference on functional brain connectivity: from Beethoven to Eminem. Sci. Rep. 4, 6130. doi:10.1038/srep06130

Xu, R., Wunsch, D., 2005. Survey of clustering algorithms. IEEE Trans. Neural Networks 16, 645–678. doi:10.1109/TNN.2005.845141

Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkar, D., Hollinshead, M., Roffman, J., Smoller, J., Zollei, L., Polimeni, J., Fischl, B., Liu, H., Buckner, R., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165. doi:10.1152/jn.00338.2011.

Yeo, B.Y.T., Ou, W., 2004. Clustering fMRI time series. Mit 1–12.

Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C., Jin, C., Wee, C.Y., Yuan, Y., Lv, P., Yin, Y., Hu, X., Duan, L., Hu, X., Han, J., Wang, L., Shen, D., Miller, L.S., Li, L., Liu, T., 2013. DICCCOL: dense individualized and common connectivity-based cortical landmarks. Cereb Cortex 23, 786–800. doi:10.1093/cercor/bhs072