World Scientific
www.worldscientific.com

# Towards Tunable Consensus Clustering for Studying Functional Brain Connectivity During Affective Processing

Chao Liu*,¶, Basel Abu-Jamous*,‖, Elvira Brattico†,‡,** and Asoke K. Nandi*,§,††

*Department of Electronic and Computer Engineering
Brunel University London, London, UK

†Center for Music in the Brain (MIB)
Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

‡The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

§Distinguished Visiting Professor
The Key Laboratory of Embedded Systems and Service Computing
College of Electronic and Information Engineering
Tongji University, Shanghai, P. R. China
¶Chao.Liu@brunel.ac.uk
‖Baselwaelaref.Abujamous@brunel.ac.uk
**Elvira.Brattico@clin.au.dk
††Asoke.Nandi@brunel.ac.uk

In the past decades, neuroimaging of humans has gained a position of status within neuroscience, and data-driven approaches and functional connectivity analyses of functional magnetic resonance imaging (fMRI) data are increasingly favored 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. We propose a tunable consensus clustering paradigm that aims at overcoming the clustering methods selection problem as well as reliability issues in neuroimaging by means of first applying several analysis methods (three in this study) on multiple datasets and then integrating the clustering results. To validate the method, we applied it to a complex fMRI experiment involving affective processing of hundreds of music clips. We found that brain structures related to visual, reward, and auditory processing have intrinsic spatial patterns of coherent neuroactivity during affective processing. The comparisons between the results obtained from our method and those from each individual clustering algorithm demonstrate that our 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. The consensus clustering method is implemented in the R package "UNCLES" available on http://cran.r-project.org/web/packages/UNCLES/index.html.

Keywords: Consensus clustering; Bi-CoPam; model-free analysis; fMRI; affective processing; functional connectivity.

---

## 1. Introduction

After the rapid rise of human neuroimaging in the 90s ("the decade of the brain")[1] with funding initiatives, papers in such prominent journals as Nature, Science, and lavish attention from the press, there is currently a vein of skepticism towards neuroimaging findings. Papers making the headlines announced significant brain activity in dead salmon or evidenced magical high correlations between behavioral and brain data.[2] Although the fascination of brain data can still blur critical thinking in front of a crudely-built mock brain scanner,[3] 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.[4] 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 functional magnetic resonance imaging (fMRI) and leading to inconsistent findings.[5] 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 (for reviews, cf. Ref. 6), and inter-subject correlation analysis[7] 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.[8]

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,[9] the anterior cingulated cortex, the insula,[10] the orbitofrontal cortex[11,12] and the reward circuit[13] 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,[14] eigenvector centrality mapping,[15] network science[16–18] and ICA[19] have been only marginally adopted to investigate functional connectivity during listening to music. Clustering analysis[20,21] 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,[22] hierarchical clustering,[23] artificial neural network[24]-based self-organizing maps (SOMs),[25] graph clustering,[22,26] and fuzzy clustering.[27] When using clustering analysis, 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 propose a new consensus clustering[28] strategy, called the binarization of consensus partition matrices (Bi-CoPaM).[29] Rather than ranking the different clustering algorithms, Bi-CoPaM 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, we selected three widely used clustering algorithms in neuroimaging, namely the $K$-means, hierarchical clustering and SOM,[23,30] to be fed into the Bi-CoPaM pipeline and produced consensus results. The Bi-CoPaM pipeline extracts clusters 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 such a pipeline, 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.

## 2. Materials and Methods

### 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 were musicians who possessed formal musical training, on average, for $16.1 \pm 6$ (SD) years. The others were nonmusicians 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 $2.5 \pm 3.4$ (SD) hours per day and nonmusicians declared that they listened to music several hours per day. The study were 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 Refs. 31–33.

### 2.2. *Music*

Four stimulus categories were used in the fMRI experiment. These categories consisted 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). Participants were asked to bring four music pieces for each category, giving 16 pieces in total. (For more details on selecting the appropriate music stimuli, cf. Ref. 31). Two 18-second long music excerpts with 500 ms fade-in and fade-out were selected from each music piece using Adobe Audition and on the basis of a listening test. This yielded 32 excerpts with eight in each stimulus category.

### 2.3. *fMRI experiment*

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 30 dB of gradient noise attenuation. Thirty-three oblique slices covering the whole brain (field of view 20 mm; $64 \times 64$ matrix; thickness 4 mm; spacing 0 mm) were acquired using an interleaved gradient echo-planar imaging (EPI) sequence with TR equal to 3 s, echo time 32 ms and flip angle 90°, sensitive to blood-oxygen-level dependent (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 emotion aspects of the stimuli. Following the end of the stimuli was a 3 s 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 hand. 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; $256 \times 256$ matrix; thickness 1 mm; spacing 0 mm) were acquired for about 9 min. (For more details about this experimental procedure and other scanning technical specifications, cf. Ref. 31)

### 2.4. *Data preprocessing and preparation*

The whole-brain images were preprocessed by statistic parametric mapping 8 (SPM8, http://www.fil.ion.ucl.ac.uk/spm) and voxel-based morphometry (VBM, http://dbm.neuro.uni-jena.de/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 more details, cf. Ref. 31). For preparing the data for the consensus analysis, we followed two steps 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 ($228453 \times 1$) by using a standard brain mask. The above step was 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 s at a sampling rate
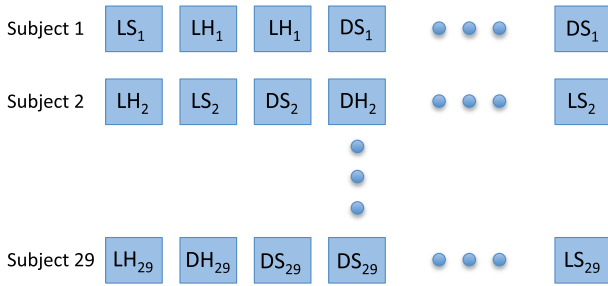
**Fig. 1.** Structure of music excerpts where DH stands for disliked happy music, DS for disliked sad music, LH for liked happy music, and LS for liked sad music. Note that each category has different music excerpts for different subjects. For example, $LS_1$ listened by subject 1 is different from $LS_2$ 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 $DS_1$.

of 3 s), and each corresponding to instances when the participants were listening to music clips. Figure 1 illustrates the order each music excerpt was presented (partial).

For one excerpt, we used a $228,453 \times 6$ (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 $1856 \times 228,453 \times 6$ (or 7) data points that were used in the following clustering analysis.

### 2.5. *Consensus clustering (Bi-CoPaM)*

The Bi-CoPaM method is a tunable consensus clustering method which recruits various single clustering algorithms while considering various datasets to identify the subset of objects (e.g. voxels) that are consistently correlated.[29] Given $L$ datasets and $C$ clustering methods, the Bi-CoPaM can be applied by the following four main steps:

(i) Partition generation: each of the $C$ clustering methods is applied to each of the $L$ datasets yielding $R = C \times L$ partitions.

(ii) Relabeling: because clustering is unsupervised, there are no labels for the clusters in the different partitions, i.e. the $i$th cluster in one partition is not guaranteed to match the $i$th cluster in another partition. Relabeling reorders, the clusters in the partitions so that they become aligned. Min–min approach was used to perform relabeling.

(iii) Fuzzy consensus partition matrix (CoPaM) generation: the relabeled partitions are averaged to produce a fuzzy CoPaM in which each voxel has a fuzzy membership value in each of the clusters based on the number of individual partitions that assigned it to it.

(iv) Quenching/Binarization: the fuzzy CoPaM is binarized to produce the final binary partition. Different threshold binarization (DTB) technique assigns a voxel to a cluster if and only if its fuzzy membership value in that cluster is higher than its closest competing cluster fuzzy membership value by the value of the parameter $\delta$. The parameter $\delta \in [0,1]$ controls the tightness of the cluster where $\delta = 0.0$ is the least tight (most sparse) and $\delta = 1.0$ is the tightest.

One important feature of the Bi-CoPaM is that the results are 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 $\delta$ values while others would resist higher levels of tightening. In general, the role of many of the clusters that become empty at low $\delta$ 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 $\delta$ 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.

### 2.6. *Degree of freedom in Bi-CoPaM*

The Bi-CoPaM method requires presetting the number of clusters ($K$). Moreover, the tightening parameter $\delta$ needs to be optimized. $M-N$ scatter plots technique tackles those issues. Each time that Bi-CoPaM is applied with different $K$ values, DTB binarization is performed with a range of $\delta$ values (e.g. from 0.0 to 1.0 with 0.1 steps). All of the individual clusters that appear in the results are scattered on a 2D plot where the horizontal axis ($M$) represents the average mean square error (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. The average MSE per voxel for the $k$th cluster can be calculated by

$$\text{MSE}_{\text{cluster}(k)} = \frac{1}{LN_k} \sum_{l=1}^{L} \sum_{n \in C_k} \|\mathbf{x}_n^l - \mathbf{z}_k^l\|^2, \quad (1)$$

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, $x_n^l$ is the normalized BOLD signal vector of the $n$th voxel in the cluster from the $l$th dataset, $z_k^l$ is the average normalized BOLD signal vector of the voxels in the $k$th cluster from the $l$th dataset, and $\|\cdot\|$ is the norm of a vector. The objective is to maximize the number of voxels included in the clusters while minimizing the dispersion within the clusters measured by the MSE metric.

As all of the individual clusters generated under all of the different combinations of parameters are scattered on this $M$–$N$ plot as shown in Fig. 2.

The cluster closest to the top left corner 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 of selecting clusters and removing those with overlaps with the selected ones are repeated iteratively up to a preset maximum number of clusters or earlier when the scatter plots are empty. The final number of clusters



Fig. 2.   (Color online) Example of $M$–$N$ scatter plots technique.

is not predetermined 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.

## 2.7.   *Test the over representation*

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. The formula for calculating the probability is

$$P(X = k \mid N, K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad (2)$$

where the bracket is the binomial coefficient.[34] Hypergeometric test is often used to identify which sub-population is over or under represented in a sample. In this experiment, we use the hypergeometric cumulative distribution function (CDF) to compute the p value. The hypergeometric CDF is

$$P(X \geq k \mid N, K) = \sum_{i=k}^{n} P(X = i). \quad (3)$$

The $P$ in formula (3) 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. We utilized this test 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.
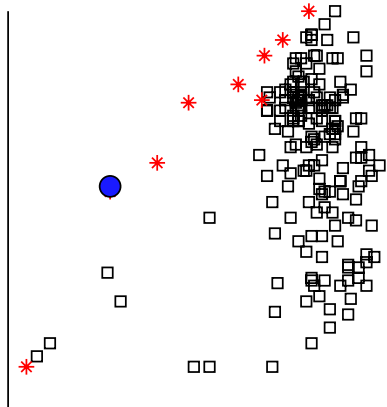
## 2.8. Clustering experiment

### 2.8.1. Clusters generation by Bi-CoPaM

Each of 1856 excerpts data (normalized to 0 mean and unit variance) was clustered by $K$-means, hierarchical and SOM with $k$ equals to 10, 25, 50, and 100. These clustering results generated by different algorithms with four different cluster numbers were combined using the Bi-CoPaM paradigm and selected by $M$–$N$ scatter plot, yielding the consensus clustering results. The brain regions within clusters were extracted out by using automated anatomical labeling (AAL) atlas.[35]

### 2.8.2. Filtering

Two types of filtering were used in order to make the results more reliable. Firstly, the original clusters were filtered by discarding those voxels with weak responses (voxels whose time series have a small variance), since the data used to be clustered were normalized 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 were discarded. After repeating this process for all the subjects, we obtained 29 thresholded partitions. Then if more than 70% percent of the subjects showed a strong response at a certain voxel, this voxel was retained for the following analysis. After this step, the clusters only contained voxels having strong BOLD responses. A sensitivity test was also carried out with respect to the filtering parameters described above. Secondly, the resulting clusters from the previous step were filtered by using the hypergeometric (HyperGeo) distribution test that discarded 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 chose $p$ equals to 0.001 to distinguish the large clusters from the isolated voxels.

### 2.8.3. Excerpts pattern analysis

By utilizing the fact that the clustering experiment in this study is based on the BOLD response shapes corresponding to stimuli, we inspected the differences in response shapes to test the hypothesis that different music categories (DH, DS, LH, LS) or different group of participants (musician versus nonmusician) would elicit distinct shapes. Once the final clusters were 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 statistical test on the response shapes. Figure 3 illustrates the excerpt data averaging process.

The averaged excerpt data were further clustered into groups with each one having distinct response shapes. HyperGeo tests were carried out between groups of musicians and nonmusicians, liked versus disliked stimuli, as well as sad versus happy stimuli. If in a particular cluster one stimulus category or participants' group was significantly represented, then this category or group would be declared to tend to have the corresponding response shape.

### 2.8.4. Analysis on music excerpts causing "strong" response

We designed another analysis pipeline based on the time series of the clustering results, which searches the musical categories that tend to elicit strong BOLD signal response. In the case of a particular subject, for a cluster (e.g. Visual or Reward or Auditory) that contains $V$ voxels and for each voxel having a time series of 450 time points (scans), it
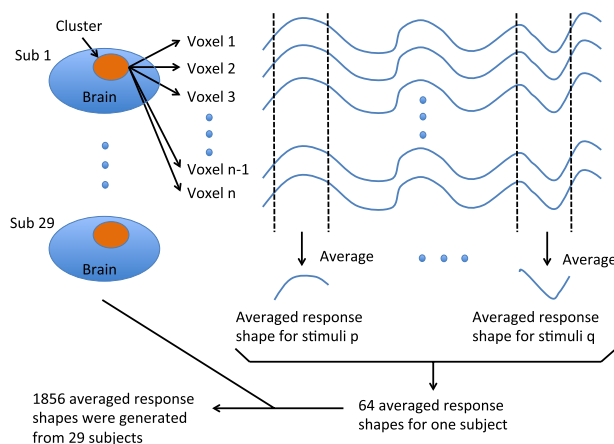


Fig. 3. Illustration of excerpt data averaging. For each subject and each excerpt, the time series of the voxels within each final cluster were averaged to obtain the mean time profile for this stimulus within this cluster. Repeating the averaging process on all the data for 29 subjects gives 1856 (= 29 subjects $\times$ 64 excerpts) averaged time profiles for each cluster.
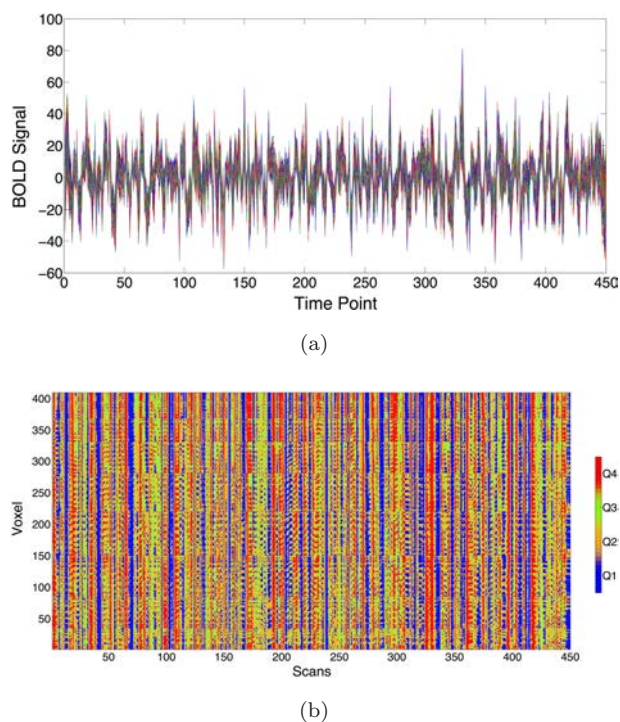
(a)



(b)

Fig. 4. (Color online) Visualization of cluster data from
one random subject. (a) is the raw time series of the vox-
els 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 visi-
ble vertical patterns through the whole scan session indi-
cate the synchronized responses among the voxels within
this cluster.

gave $V \times 450$ time points in total. Each of these data
values (amplitudes), corresponding to $V \times 450$ time
points, is assigned a number from 1 to 4 correspond-
ing to its position in quantile 1 to quantile 4. Figure 4
is the example of visualizing the quantile data of a
random cluster.

After obtaining the quantile data matrix ($V \times$
450) for a particular subject, we examined the modes
of the quantile values within the time windows
covering the duration of each stimulus (music
excerpt). If in a certain time window, the mode would
be 4, then this would be taken to mean that the
"strong" response (red colored in Fig. 3) dominates
during this stimulus. Hence, we extracted out the
excerpt categories eliciting the "strong" responses
for that particular subject, and we repeated this 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 as
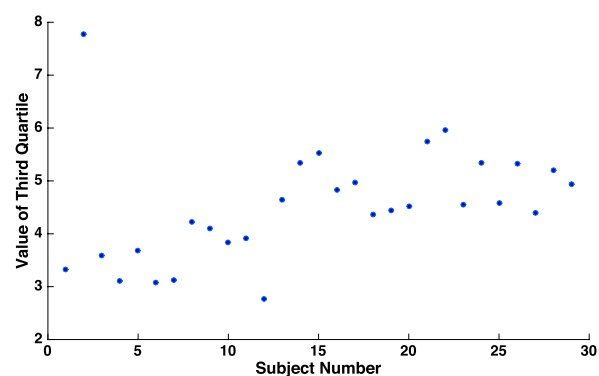


Fig. 5. The threshold values/third quartiles for strong
response in Reward cluster of all the 29 subjects.

shown in Fig. 5, we coded subjects' response val-
ues in the range from 1 to 4, irrespective of the
range of the original responses, and without ref-
erence 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 indepen-
dent subject-wise, category-wise, and excerpt-wise.
Afterwards, we tested the distribution differences
among the four stimulus categories and the two par-
ticipant groups to inspect which type of stimuli or
subject group elicited stronger responses than others
within the same cluster type (e.g. Visual or Reward
or Auditory).

### 2.9. Pipelines

Figure 6 demonstrates the whole experiment pipeline
in this paper.



Fig. 6. Experiment pipeline.

## 2.10. *Availability*

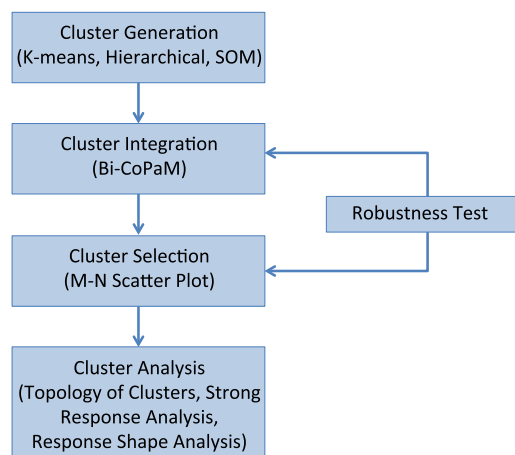The Bi-CoPAM method and the M-N plots technique are implemented in the *R* package "UNCLES" available on: http://cran.r-project.org/web/packages/UNCLES/index.html.

## 3. Results

### 3.1. *Topology of clusters*

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, thus complying with expectations based on knowledge of brain physiology.[36] Among these clusters, we found that emotion and sensory-related brain areas were grouped into three clusters separately (see Fig. 7) which correspond well with the literature studying music emotions with the model-based approach.[30,31] 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.

### 3.2. *Comparison among multiple clustering algorithms combinations and single algorithm*

In order to evaluate the advantage of using the Bi-CoPaM with multiple methods over the Bi-CoPaM with a single method, we compared the clustering results obtained by the following four experiment scenarios: (1) Bi-CoPaM with $k$-means, (2) hierarchical clustering, (3) SOM, and (4) the combination of all the aforementioned methods. The Venn diagrams (Fig. 8) show the relationships among the results obtained by different clustering algorithm settings for cluster A Visual, cluster B Reward, and cluster C Auditory.

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 failed to



| 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) |

(a) Cluster A Visual



| 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) |

(b) Cluster B Reward



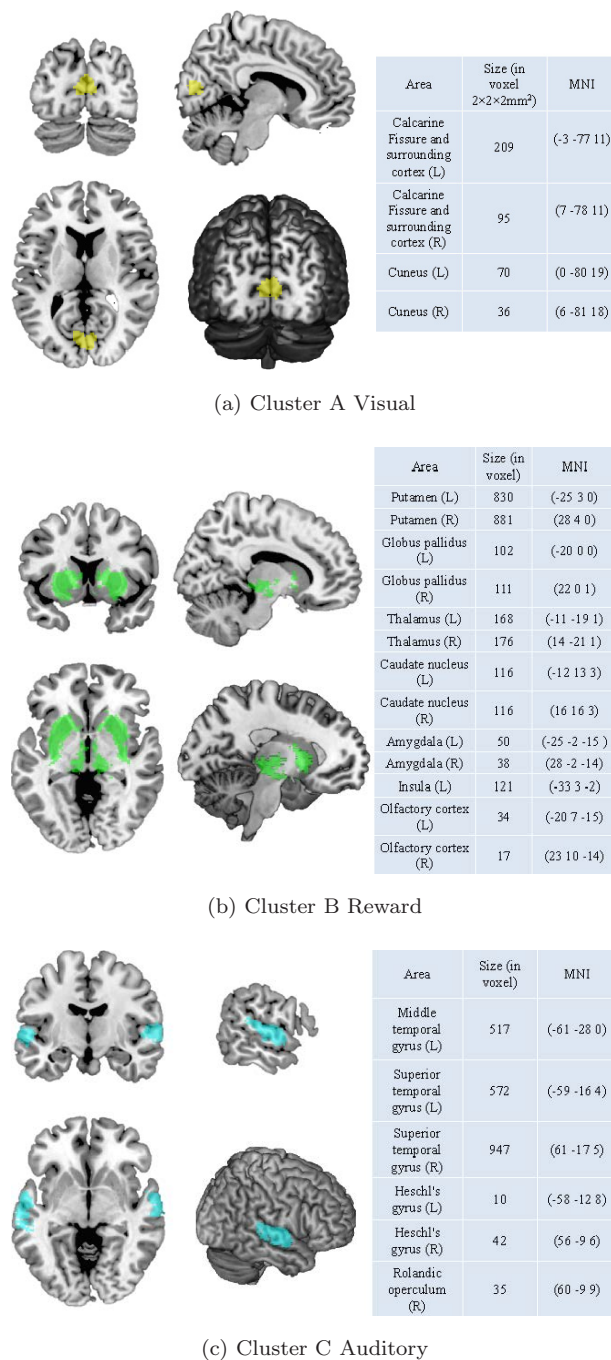| Area | Size (in voxel) | MNI |
|---|---|---|
| Middle temporal gyrus (L) | 517 | (-61 -28 0) |
| Superior temporal gyrus (L) | 572 | (-59 -16 4) |
| Superior temporal gyrus (R) | 947 | (61 -17 5) |
| Heschl's gyrus (L) | 10 | (-58 -12 8) |
| Heschl's gyrus (R) | 42 | (56 -9 6) |
| Rolandic operculum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

(c) Cluster C Auditory

Fig. 7. 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.
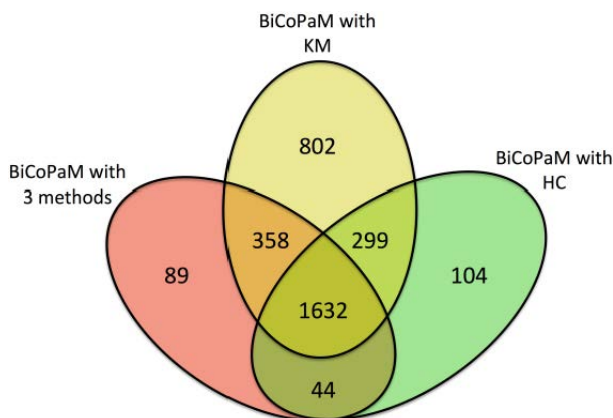
form a cluster within subcortical reward regions. Bi-CoPaM with all the three methods not only contained the common parts of the results from the other two experiment paradigms, but also included 988 voxels that are not included in the results obtained
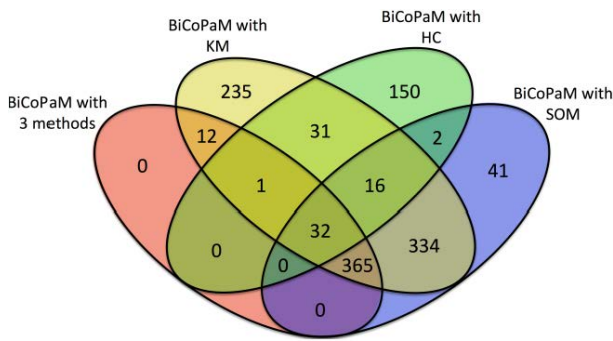
(a) Cluster A Visual
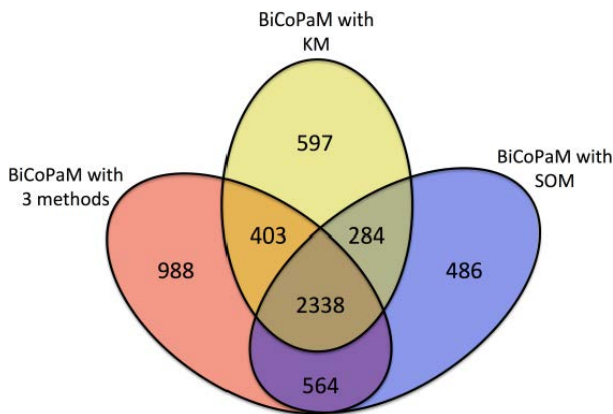


(b) Cluster B Reward



(c) Cluster C Auditory

Fig. 8. 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.

with either KM or SOM method. Cluster C Auditory did not appear in the results by the single SOM algorithm and similarly, the Bi-CoPaM with all the three methods successfully found 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, of the four experimental scenarios, Bi-CoPaM with all the three methods provides the most complete set of clusters.

### 3.3. *Filtering sensitivity test*

We ran the filtering (Material and Methods, Sec. 2.8, Filtering) with different parameter combinations (percentage of total data and percentage of mean variance) and inspected the filtered cluster size to see the filtering impact on the results. From Tables 1 to 3, 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.

### 3.4. *Robustness test against subject functional data variability*

Considering that the brain activity to a certain stimulus varied among different individuals, it raises the problem whether the algorithm could perform well regarding the functional data variation among participants. Since it has been demonstrated that Bi-CoPaM with three methods provides the most complete set of clusters, we carried out the test to investigate the robustness of Bi-CoPaM with three clustering methods against data variability on aforementioned three clusters. We generated two groups of subsets for test. One was created by randomly selecting 75% of the musicians (10 out of 13) and nonmusicians (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 was repeated 10 times and these 10 subsets formed group A. Similarly, we chose a different ratio of 90% of the musicians (12 out of 13) and nonmusicians (14 out of 16) as well as 90% of the excerpts for each participant, when the data were randomly selected. Repeating the selection with new ratio for 10 times formed group B consisting approximately 80% of all the data. Then we applied Bi-CoPaM with three clustering algorithms on these subsets and recorded

Table 1.   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 2.   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 3.   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 |

the clustering results. Table 4 lists the results from group A and Table 5 lists the results from group B. Note that for the final size of cluster Reward and Auditory, there were unspecified voxels included within AAL atlas which are not reported in Tables 4 and 5.

Tables 4 and 5 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 but was never missed in the second test reported in Table 5. On one hand, this proved our Bi-CoPaM method 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.

### 3.5.  *Statistical test of temporal features of clusters*

We carried out the response shape analysis in the three clusters Visual, Reward, and Auditory with

Table 4. 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 | | | | | | | | | | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 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 | 0.60 | 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 | 0.60 | 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 5. Results of test group B (80% of the full data).

| Clusters | | Trial | | | | | | | | | | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 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. | 0.80 | 0.59 | 0.82 | 0.66 | 0.67 | 0.54 | 0.68 | 0.50 | 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 | 0.50 | 0.73 | 0.79 | 0.57 | 0.76 | 0.50 | 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 | 0.40 | 0.19 | 0.8 | 0.572 |

respect to the contrasts between experimental conditions and groups: liked versus disliked, happy versus sad, musician versus nonmusician. In cluster A Visual, we found the response difference in musician versus nonmusician group as shown in Fig. 9. For the response shape (initially reduces and then steadily rises till the end of the stimuli) shown in the figure, nonmusician is significantly over represented with $p$ value equal to 0.00053. Additionally, no significant difference in the shape of the BOLD responses was found on the contrasts liked versus disliked and happy versus 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.

### 3.6. Music categories and participant groups that cause "strong" BOLD response level

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 6) with respect to contrasts liked versus
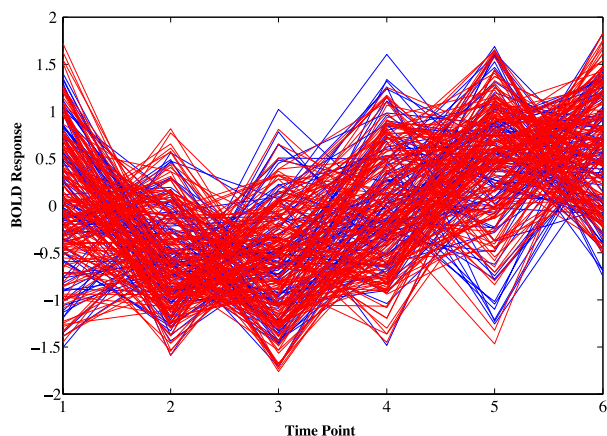
Fig. 9. (Color online) Response pattern in cluster A Visual. There are 248 excerpt responses (248 six-points temporal profiles) that were grouped together. Of these similar BOLD response shapes, 87 come from musicians (blue lines) and 161 come from nonmusicians (red lines). The $p$ value for nonmusician in this distribution is 0.00053, indicating nonmusicians are more likely to elicit the above response shape than musicians.

Table 6. $M$ stands for group musician and $N$ stands for group nonmusician.

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

disliked, happy versus sad and musician versus nonmusician.

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. Discussion

In the current study, we tested a novel data-driven approach aimed at integrating 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 by Bi-CoPaM, 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 nonmusicians.

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.[19,24] 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 the their time series 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 generated and selected the nontrivial 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,[37,38] 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 Bi-CoPaM utilizing all three clustering methods. To this end, 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 using Bi-CoPaM with multiple methods over using a single specific method.

Our innovative use of Bi-CoPaM methodology allows us to find clusters including functionally and anatomically related neural networks responding to emotional music (for a broad discussion of the model-based findings obtained with the same dataset and of their relation with the acoustic features present in the music stimuli, cf. Ref. 33).

After the cluster generation and selection, emotion-related brain structures responsible for rewarding and pleasurable sensations such as the basal ganglia, thalamus, insula,[39] and other areas involved with processing of auditory features such as the Heschl's gyrus, the Rolandic operculum and the superior temporal gyrus[40] 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 Bi-CoPaM methodology 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,[16] no reward circuit activity was found. Our study confirms findings on the neural structures related to musical emotions obtained with model-based approaches.[31,39,41–45] 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).[46]

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 nonmusicians). Unlike the traditional statistical tests for fMRI data that compare the response strength differences using a general linear model,[47] which often have been questioned,[48] 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.[49,50] With the current results, we replicated those findings without falling into the risks of circular analysis.[8] Moreover, we evidenced a difference between musicians and nonmusicians 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 nonmusicians accumulated and achieved as a consequence of listening to 18 s of emotionally-loaded music. A similar result of tightened connectivity in visual clusters was found by Luo *et al.*[51] 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 nonmusicians on other regions, a tendency for a larger recruitment of visual clusters seems to be present in nonmusicians, similarly to the current study.

On the other hand, when the same dataset was analyzed 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 decline, in musicians over nonmusicians, 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 nonmusicians for all the stimulus conditions, whereas in the study by Brattico *et al.*[33] 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 nonmusicians 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 nonmusicians, in the line of previous findings by Chapin *et al.*[52] or James *et al.*[53] As discussed already in Ref. 54 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.

Here, we first clustered the data not with one clustering algorithm but with three clustering algorithms independently. Then Bi-CoPaM generates consensus among the three sets of clusters. This takes out the risks of capturing artefacts of an individual clustering algorithm. Furthermore, we analyzed the data not with one set of parameters but with many sets of parameters. For example, Bi-CoPaM explores 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, our study provided a robust solution for obtaining the consistently strong activation patterns in neuroimaging studies of affect.

## Acknowledgment

## References

1. E. G. Jones and L. M. Mendell, Assessing the decade of the brain, *Science* **284**(5415) (1999) 739.
2. E. Vul, C. Harris, P. Winkielman and H. Pashler, Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition, *Psychol. Sci.* **4**(3) (2009) 274–290.
3. S. S. Ali, M. Lifshitz and A. Raz, Empirical neuroenchantment: From reading minds to thinking critically, *Front. Hum. Neurosci.* **8** (2014) 357.
4. A. Abbott, Brain imaging studies under fire, *Nature* **457** (2009) 245.
5. K. S. Button, J. P. A. Ioannidis, C. Mokrysz *et al.*, Power failure: Why small sample size undermines the reliability of neuroscience, *Nat. Rev. Neurosci.* **14**(5) (2013) 365–376.
6. M. P. van den Heuvel and H. E. H. Pol, Exploring the brain network: A review on resting-state fMRI functional connectivity, *Eur. Neuropsychopharmacol.* **20**(8) (2010) 519–534.
7. U. Hasson, Y. Nir, I. Levy, G. Fuhrmann and R. Malach, Intersubject synchronization of cortical

activity during natural vision, *Science* **303**(5664) (2004) 1634–1640.

8. N. Kriegeskorte, M. A. Lindquist, T. E. Nichols, R. A. Poldrack and E. Vul, Everything you never wanted to know about circular analysis, but were afraid to ask, *J. Cereb. Blood Flow Metab.* **30**(9) (2010) 1551–1557.

9. N. Gosselin, I. Peretz, E. Johnsen and R. Adolphs, Amygdala damage impairs emotion recognition from music, *Neuropsychologia* **45**(2) (2007) 236–244.

10. A. R. Damasio, T. J. Grabowski, A. Bechara *et al.*, Subcortical and cortical brain activity during the feeling of self-generated emotions, *Nat. Neurosci.* **3**(10) (2000) 1049–1056.

11. S. Khalfa, D. Schon, J.-L.Anton and C. Liégeois-Chauvel, Brain regions involved in the recognition of happiness and sadness in music, *Neuroreport.* **16**(18) (2005) 1981–1984.

12. A. J. Blood and R. J. Zatorre, Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion, *Proc. Natl. Acad. Sci. USA* **98**(20) (2001) 11818–11823.

13. M. T. Mitterschiffthaler, C. H. Y. Fu, J. A. Dalton, C. M. Andrew and S. C. R. Williams, A functional MRI study of happy and sad affective states induced by classical music, *Hum. Brain Mapp.* **28**(11) (2007) 1150–1162.

14. C. Karmonik, A. K. Brandt, S. H. Fung, R. G. Grossman and J. T. Frazier, Graph theoretical connectivity analysis of the human brain while listening to music with emotional attachment: feasibility study, in *Conf. Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conf.* (2013) Vol. 2013, pp. 6526–6529.

15. S. Koelsch and S. Skouras, Functional centrality of amygdala, striatum and hypothalamus in a "small-world" network underlying joy: An fMRI study with music, *Hum. Brain Mapp.* **35**(7) (2014) 3485–3498.

16. R. W. Wilkins, D. A. Hodges, P. J. Laurienti, M. Steen and J. H. Burdette, Network science and the effects of music preference on functional brain connectivity: From Beethoven to Eminem, *Sci. Rep.* **4** (2014) 6130.

17. M. Ahmadlou, H. Adeli and A. Adeli, Graph theoretical analysis of organization of functional brain networks in ADHD, *Clin. EEG Neurosci.* **43**(1) (2012) 5–13.

18. M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task, *Clin. Neurophysiol.* **125**(4) (2014) 694–702.

19. F. Cong, V. Alluri, Q. Lin *et al.*, Semi-blind independent component analysis of functional MRI elicited by continuous listening to music, in *ICASSP 2013* Vol. 2013, pp. 1310–1314.

20. T. W. Liao, Clustering of time series data — a survey, *Pattern Recognit.* **38** (2005) 1857–1874.

21. L.-D. Shi, Y.-H. Shi, Y. Gao, L. Shang and Y.-B. Yang, XCSc: A novel approach to clustering with extended classifier system, *Int. J. Neural Syst.* **21**(1) (2011) 79–93.

22. G. Bello-Orgaz, H. D. Menéndez and D. Camacho, Adaptive $K$-means algorithm for overlapped graph clustering, *Int. J. Neural Syst.* **22**(05) (2012) 1250018.

23. C. Goutte, P. Toft, E. Rostrup, F. Nielsen and L. K. Hansen, On clustering fMRI time series, *Neuroimage* **9**(3) (1999) 298–310.

24. M. Ahmadlou and A. Hojjat, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integr Comput Aided Eng.* **17**(3) (2010) 197–210.

25. S. Haykin, Neural Networks — A comprehensive Foundation, 3rd edn. (Pearson, Prentice Hall, Singapore, 1999).

26. M. Ahmadlou and H. Adeli, Functional community analysis of brain: A new approach for EEG-based investigation of the brain pathology, *Neuroimage* **58**(2) (2011) 401–408.

27. R. Baumgartner, C. Windischberger and E. Moser, Quantification in functional magnetic resonance imaging: Fuzzy clustering versus correlation analysis, *Magn. Reson. Imaging* **16**(2) (1998) 115–125.

28. R. Ghaemi, N. Sulaiman, H. Ibrahim and N. Mustapha, A survey: Clustering ensembles techniques, *Eng. Technol.* **38** (2009) 636–645.

29. B. Abu-Jamous, R. Fa, D. J. Roberts and A. K. Nandi, Paradigm of tunable clustering using binarization of consensus partition matrices (Bi-CoPaM) for gene discovery, *PLoS One* **8**(2) (2013) e56432, doi: 10.1371/journal.pone.0056432.

30. S. J. Peltier, T. A. Polk and D. C. Noll, Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm, *Hum. Brain Mapp.* **20**(4) (2003) 220–226.

31. E. Brattico, V. Alluri, B. Bogert *et al.*, A functional MRI study of happy and sad emotions in music with and without lyrics, *Front. Psychol.* **2** (2011) 308.

32. S. Saarikallio, S. Nieminen and E. Brattico, Affective reactions to musical stimuli reflect emotional use of music in everyday life, *Music Sci.* **17**(1) (2012) 27–39.

33. E. Brattico, B. Bogert, V. Alluri, M. Tervaniemi and T. Eerola, It's sad but I like it: The neural dissociation between musical emotions and liking in experts and laypersons, *Front. Hum. Neurosci.* **9**(676) (2015) 1–21, doi: 10.3389/fnhum.2015.00676.

34. J. A. Rice, *Mathematical Statistics and Data Analysis*, Thomson/Brooks/Cole (CA, USA, 2007).

35. N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou *et al.*, Automated anatomical labeling of activations in SPM using a macroscopic anatomical

parcellation of the MNI MRI single-subject brain, *Neuroimage* **15**(1) (2002) 273–289.

36. S. A. Huettel, A. W. Song and G. McCarthy, *Functional Magnetic Resonance Imaging*, 2nd edn. (Sinauer Associates, 2009).

37. M. Welvaert and Y. Rosseel, On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data, *PLoS One* **8**(11) (2013) e77089.

38. C. M. Bennett and M. B. Miller, How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci.* **1191** (2010) 133–155.

39. V. N. Salimpoor and R. J. Zatorre, Neural interactions that give rise to musical pleasure, *Psychol. Aesthet. Creat. Arts* **7**(1) (2013) 62–75.

40. V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams and E. Brattico, Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm, *Neuroimage* **59**(4) (2012) 3677–3689.

41. L. F. Barrett and T. D. Wager, The structure of emotion evidence from neuroimaging studies, *Curr. Dir. Psychol. Sci.* **15**(2) (2006) 79–83.

42. S. Koelsch, Towards a neural basis of music-evoked emotions, *Trends Cogn. Sci.* **14**(3) (2010) 131–137.

43. S. Koelsch, T. V. Fritz, D. Y. Cramon, K. Müller and A. D. Friederici, Investigating emotion with music: An fMRI study, *Hum. Brain Mapp.* **27**(3) (2006) 239–250.

44. S. Koelsch, Brain correlates of music-evoked emotions, *Nat. Rev. Neurosci.* **15**(3) (2014) 170–180.

45. E. Brattico, B. Bogert and T. Jacobsen, Toward a neural chronometry for the aesthetic experience of music, *Front. Psychol.* **4** (2013) 206.

46. D. Tomasi and N. D. Volkow, Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder, *Biol. Psychiatry* **71**(5) (2012) 443–450.

47. K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith and R. S. J. Frackowiak, Statistical parametric maps in functional imaging: A general linear approach, *Hum. Brain Mapp.* **2**(4) (1995) 189–210.

48. J.-B. Poline, M. Brett, The general linear model and fMRI: Does love last forever? *Neuroimage* **62**(2) (2012) 871–880.

49. V. N. Salimpoor, I. van den Bosch, N. Kovacevic, A. R. McIntosh, A. Dagher and R. J. Zatorre, Interactions between the nucleus accumbens and auditory cortices predict music reward value, *Science* **340**(6129) (2013) 216–219.

50. V. Alluri, E. Brattico, P. Toiviainen *et al.*, Musical expertise modulates functional connectivity of limbic regions during continuous music listening, *Psychomusicology Music Mind Brain* **25**(4) (2015) 443–454.

51. C. Luo, S. Tu, Y. Peng *et al.*, Long-term effects of musical training and functional plasticity in salience system, *Neural Plasticity*, 2014 (180138), 13 pp., doi: 10.1155/2014/180138.

52. H. Chapin, K. Jantzen, J. A. S. Kelso, F. Steinberg and E. Large, Dynamic emotional and neural responses to music depend on performance expression and listener experience, *PLoS One* **5**(12) (2010) e13812, doi: 10/1371/journal.pone.0013812.

53. C. E. James, J. Britz, P. Vuilleumier, C. A. Hauert and C. M. Michel, Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts, *Neuroimage* **42**(4) (2008) 1597–1608.

54. M. Reybrouck and E. Brattico, Neuroplasticity beyond sounds: Neural adaptations following long-term musical aesthetic experiences, *Brain Sciences* **5**(1) (2015) 69–91, doi: 10.3390/brainsci5010069.