Decoding Music Genres Based on High Resolution Brain Activity Information

Qinhan Hou, Gaoyan Zhang*
*College of Intelligence and Computing, Tianjin key lab of cognitive computing and application, Tianjin University, China
E-mail: zhanggaoyan@tju.edu.cn

Abstract— **Decoding stimuli from the brain is important for understanding the brain's processing mechanism of external information, which can also promote the development of brain-computer interface. Most of the existing researches focused on the decoding of audiovisual information. Few studies investigated the decoding of music stimuli, and the decoding accuracy is also not satisfactory. This paper uses a public 7-Tesla fMRI image dataset, which collects the high resolution blood oxygen dependent level (BOLD) signals when 20 subjects listen to 5 music genres. After fMRI data preprocessing, two feature selection methods were used. One is based on a prior template (called MASK) including the Heschl's gyrus (HG), the anterior superior temporal gyrus (aSTG), and the posterior superior temporal gyrus (pSTG). The other one is whole brain analysis of variance (ANOVA). Then, the gradient boosting decision tree (GBDT) algorithm is used to train the decoding models to discriminate different music genres. Results showed that among the five genres, ambient music is easier than the other four categories to be decoded. Compared with the previous study that used the same dataset and the same prior template but combined with the classifier of support vector machine[1], the GBDT algorithm improved the accuracies in most genres from around 45% to around 65%. Compared with the MASK method, the ANOVA method improved the decoding accuracy to larger than 74% for all genres. Analysis of contributive regions in the ANOVA method shows that insular and parietal regions are additionally recruited in decoding of music genres, which may be related to music understanding and emotion expression. In summary, the findings in this paper help to understand the brain mechanism of music processing in depth. Meanwhile, the decoding model proposed in this study can also be used to classify other fMRI stimuli.**

I. Introduction

The human brain can be seemed like a system to consistently process external stimulation and this process is a kind of information encoding and decoding. Using the encoding model, the computer can detect the brain function activity patterns caused by specific types of stimuli; while using the decoding model, feature extraction and classification recognition can be performed based on this activity pattern to find the correct type of stimulation. Exploring brain decoding helps us understand the information processing mechanism in the brain and it can also promote the development of brain computer interface.

As Frank Tong said in his review paper [2], the previous research of the brain decoding process was more based on the visual aspect. Although some recent studies have paid attention to auditory signal decoding, the music signal decoding is much less[3][4][5]. One most different thing between a normal auditory signal and a part of the music signal is that music often includes the emotional element. Zulkurnaini once studied the neural potential response of 28 subjects when they listened to the Quran and classical music. By using EEG data for comparison, he found that the human brain's alpha-band brain waves are stronger when listening to the Quran, so listening to the Quran is more likely to produce a sense of relaxation and alertness than listening to classical music[6]. Through this research, we know that in addition to simple physiological reactions, music can also cause emotional reactions. Besides, past studies have shown that music exposure also enhances emotional and cognitive functioning in healthy subjects and various clinical patient groups, and has a significant helping effect in the recovery of certain diseases, such as stroke [7] These neuroscience studies on music are mainly based on the encoding process. In terms of music decoding, Toiviainen et al. used Lasso regression analysis on 3T fMRI brain data to explore the response of different areas of the brain to different characteristics of music, such as music brightness, fullness, and complexity[8]. Casey had researched the music decoding mode with 7 T fMRI data to discriminate five different music genres, while the classification accuracies of his model were not satisfactory (most are lower than 50% except for ambient genre) [1]. Therefore, further studies are needed to explore the effective decoding model and thus understand the music processing mechanism in depth.

In decoding studies, high resolution brain imaging data is useful to improve the decoding performance. In general, fMRI data, compared with Electroencephalography (EEG) and magnetoencephalography (MEG), has a better spatial resolution. Besides, in the past low magnetic field (such as 3- Tesla) studies, the activity state of microvessels was often covered by the normal blood flow of surrounding large blood vessels, resulting in inaccurate positioning. In contrast, the high-field (7-Tesla) fMRI developed in recent years can greatly improve the accuracy of positioning so that it has a great advantage in spatial positioning. Considering this advance of high-field fMRI, this paper selects 7T open fMRI images as experimental data[9], which has also been used in Casey's study[1].

Casey mainly used support vector machine(SVM) with linear kernel for classification, but we believe that the kernel function is not the proper one. In fact, it is hard to find the most suitable kernel function for unknown data. So we proposes to use gradient boosting decision tree (GBDT) algorithm to replace linear kernel SVM, which has high-level anti-noise characteristics to adapt brain activation images. So we will first compare the results given by us using GBDT to Casey's, show its improvement in classification performance. The main goal of this paper is to construct an effective discriminative model to decode different music genres with better accuracy and further investigate the contributive regions in this process to understand the music processing mechanism in the brain.

II. Materials and Methods

A. Dataset

We used the public OpenFMRI dataset published by Hanke in 2015[9]. To get this dataset, Hanke chose 20 subjects, all of whom have normal hearing and no mental problems. These subjects were asked to listen to 25 pieces of music stimulation (each piece of music lasted 6 seconds, the sampling rate was 44.1 kHz). The 25 pieces of music include five types of music,

A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	1	5	0	2	6	0	6	1	0	5	7	5	0	1	6	3	3	4	6
2	5	1	4	6	3	6	1	3	7	0	3	2	2	5	1	6	7	7	7
0	0	4	6	4	2	5	7	2	2	4	1	6	7	2	7	0	4	6	5
3	7	3	2	6	0	4	0	4	4	1	2	0	1	4	0	7	1	2	2
7	3	6	5	1	7	2	3	0	5	6	6	7	5	6	2	2	0	3	
5	2	2	7	0	1	1	4	6	1	3	5	3	6	3	3	1	5	0	/
4	4	0	1	2	4	7	5	7	3	7	0	4	4	0	5	4	2	1	
1	6	7	3	5	5	3	2	5	6	2	4	1	3	7	4	5	6	5	

Runs

B

SEQ 0	A C M C R S S M M S C C S R M A S A R C A M R R A
SEQ 1	S A S R S M C R C S C A A C C M A M R A R R M M S
SEQ 2	C A R A A C S A S R C R S M S S C M R R M M A M C
SEQ 3	R S S R M S C R C A R A S M M A C M C C S A A M R
SEQ 4	C S M C A M S A S S C M A A C R M M R R A R S R C
SEQ 5	C M A S C R S S R M R R A A M C S M M S A R C A C
SEQ 6	R S R M S C C A A C S S A M M A R C R A S M C M R
SEQ 7	R M S C C A R C S M M R S R A C M A A S S A M C R
C - country / M - metal / R - Rock'n'roll / A - ambient / S - symphonic	

Fig. 1 (A) The playing sequence of music. (B) The sequence order for each time of scan.

namely ambient music, country music, metal music, rock and roll, and symphony. Each piece of music has undergone rhythm alignment, and a 50ms sine wave is added at the beginning and end to mark the beginning and end.

The MRI scanning parameters are as follows: a whole-body 7-Tesla Siemens MAGNETOM magnetic resonance scanner equipped with a local circularly polarized head transmit and a 32 channel brain receive coil (Nova Medical, Inc., Wilmington, MA, USA). 36 axial slices (thickness 1.4 mm, 1.4×1.4 mm in-plane resolution, 224 mm field-of-view (FoV), anterior-to-posterior phase encoding direction) with a 10% inter-slice gap were recorded in ascending order. This configuration represents a good compromise between spatial resolution, volume coverage, and volume acquisition time [10]. The parameters of these functional MRI data are T2*-weighted echo-planar images (gradient-echo, 2s repetition time (TR), 22ms echo time, 0.78 ms echo spacing, 1488 Hz/Px bandwidth, generalized autocalibrating partially parallel acquisition (GRAPPA) acceleration factor 3, 24 Hz/Px bandwidth in phase encoding direction).

In the process of listening to musical stimuli, Hanke defined a round of stimulus as each volunteer listening to each of the 25 pieces of music in a De Bruijn cycle (Figure 1). Every time after listening to the music clip, there will be a random delay of 4, 6, or 8 seconds. Each round of stimulation lasted for a total of 300 seconds, and considering the delayed performance of hemodynamic, a total of 306 seconds of data were recorded.

The scanning interval (TR) of the experiment is 2s, so each round of scan will get 153 fMRI images. Each subject will go through 8 rounds of such stimulation.

B. Pre-processing

The original format of fMRI cannot be directly used for machine learning training, so a series of pre-processing operations are required. The significance of the pre-processing operation is to eliminate possible errors and to perform a series of registration and formatting of the image data so that it can be used in a unified machine learning process. The pre-processing process mainly includes the correction of head movement and acquisition time, space registration, and normalization. All of these operations of this experiment is completed by the SPM toolkit (https://www.fil.ion.ucl.ac.uk/spm/), which is a neuroimaging data processing tool open-sourced by UCL. The version used in the paper is SPM12.

C. Feature Selection Methods

As the parameters listed above, the whole-brain image has numerous dimensions: 853776 dimensions. Thus it needs to choose proper feature selection methods to reduce the dimensions for a subsequent experiment. We choose two methods to reduce the dimensions, one is the prior MASK, and the other one is Analysis of variance (ANOVA).

The MASK method is a hypothesis-driven method to extract features. The MASK method is to obtain a 0-1 matrix based on the specific brain region template which is selected based on prior knowledge. Through the operation of this MASK matrix and the image matrix, the target image containing only specific voxels can be obtained.

ANOVA is a data-driven method, which calculates the inter-category differences based on F statistical test. During this experiment, we will use the ANOVA method to calculate the inter-category differences in fMRI activities among the five music genres and then extract the top 5% of differential voxels from the whole brain. The higher the evaluation value calculated according to the ANOVA method, means that these voxels have different activation states due to different music. The bigger the changes are, the more their changes lead to changes in the model.

ANOVA and MASK both reduce the dimension of our data, from 800,000 to about four to five thousand. We will choose these two methods according to different experimental requirements.

D. Construction of Decoding Models

In our machine learning stage, we chose the Gradient Boosting Decision Tree (GBDT) algorithm, which is an ensemble algorithm. Its basic feature is to train many weak classifiers to combine to achieve the effect of a strong classifier. Considering that the dimension of data in this study is large and the features are not obvious, it can be regarded as weakly separable data. Therefore, training multiple weak classifiers with an ensemble algorithm can save computing resources and play an anti-overfitting effect. The weak classifier of GBDT usually adopts the Classification and Regression Tree (CART), which was proposed by Breiman et al. in 1984[11]. CART tree is a kind of binary tree. Its generation depends on each feature to split. This classification divides the input feature space into finite units and determines the predicted probability distribution on these units. The split standard adopted when this tree is generated is the Gini Index.

$$Gini(p) = 1 - \sum_{\mathrm{k}=1} \mathrm{p}_{\mathrm{k}}^{2} \tag{2}$$

The meaning of this formula is that in a classification task with a total category of K, the probability that the sample belongs to the kth category is p_k. Then according to the given sample set D, its Gini coefficient is:

$$Gini(D) = 1 - \sum_{k=1}^{K} \left(\frac{|C_k|}{|D|}\right)^2 \tag{3}$$

Here, C_k is the subset of samples belonging to the kth class in D, and K is the number of classes. If the sample set D is divided into two subsets D_1 and D_2 according to whether feature A takes a possible value a, then the formula is as follows:

$$D_1 = (\mathrm{x}, \mathrm{y}) \in \mathrm{D}|\mathrm{A}(\mathrm{x}) = \mathrm{a}, D_2 = D - D_1$$

$$Gini(\mathrm{D}, \mathrm{A}) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{4}$$

The Gini coefficient D obtained in this way represents the uncertainty of the set D after dividing the subtree by A = a.

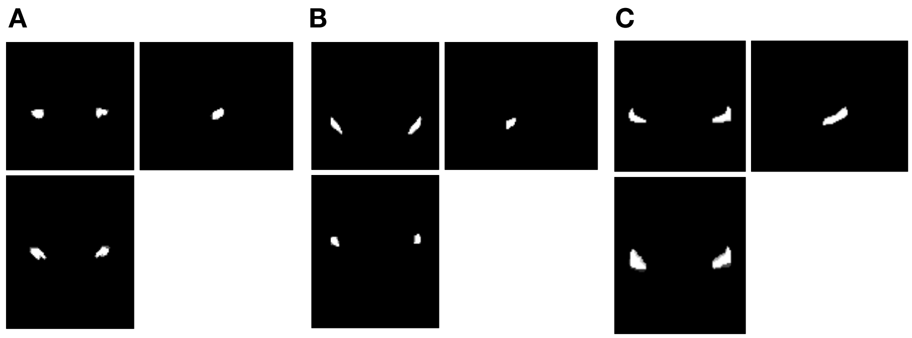

Fig. 2 Three masks for the interest regions. (A) HG (B) aSTG (C) pSTG.

The larger the Gini index, the greater the uncertainty of the sample set. Therefore, the division standard of the CART tree is to make the Gini coefficient of features after a certain numerical division as small as possible. After generating the CART tree according to the principle of optimal segmentation, a pruning operation is needed to make the decision tree as simple as possible and adapt to other data. The following loss function is used in the pruning process to calculate the pruning loss:

$$C_\alpha(T) = C(T) + \alpha|T| \tag{5}$$

When T is any subtree, $C(T)$ is the prediction error for the training data (such as the Gini index mentioned above), $|T|$ is the number of leaf nodes of the subtree, and α is a parameter to weigh training fitness of the data and the complexity of the model. Breiman had proved that the tree can be pruned by a recursive method[11]. This recursion increases α from small and obtains the optimal parameters by obtaining the optimal loss function to obtain an optimal subtree sequence.

In the case of training multiple CART trees, the model will use the boosting method to iterate. This method is essentially an additive model, that is, a linear combination of each base classifier; at the same time, it also belongs to the forward distribution algorithm, through each round of training model to further optimize the next round of models, through continuous iteration to reduce the deviation of the entire model. The initialization boosting method will assign the same weight of 1/N to each basic model, and then assign larger weights to instances that fail to train during each training process, that is, focus on strengthening instances that are more difficult to predict. In the end, the prediction function with a better prediction effect is more weighted, and vice versa, the weighted voting method is finally used to classify the test data. Besides, GBDT adopts the loss function gradient descent strategy, each time a new model is established based on the gradient descent direction of the loss function of the previous model.

III. Experiments and Results

A. Analyze 1: Decoding based on MASK of specific regions

First, we experimented with extracting features by the MASK method. The prerequisite of the MASK method is to determine the selected brain area. Casey used the same data set as this study for research in 2017[1]. He chose the posterior superior temporal gyrus (pSTG), the anterior superior temporal gyrus (aSTG), and the Heschl's gyrus (HG, the transverse temporal gyrus) area. Therefore, in this study, we also chose these three brain regions as feature regions based on his prior experience. After that, we compared the results to Casey's results to verify his conclusions and improved the classification accuracy based on his experiments.

The next step will be to make MASK maps of these three regions to extract voxels for research. The operation of extracting the MASK is completed by the SPM12 toolkit. We first got the Havard-Oxford brain template (https://identifiers.org/neurovault.collection:262), which contains the detailed partition information of the brain, and we extracted the three regions

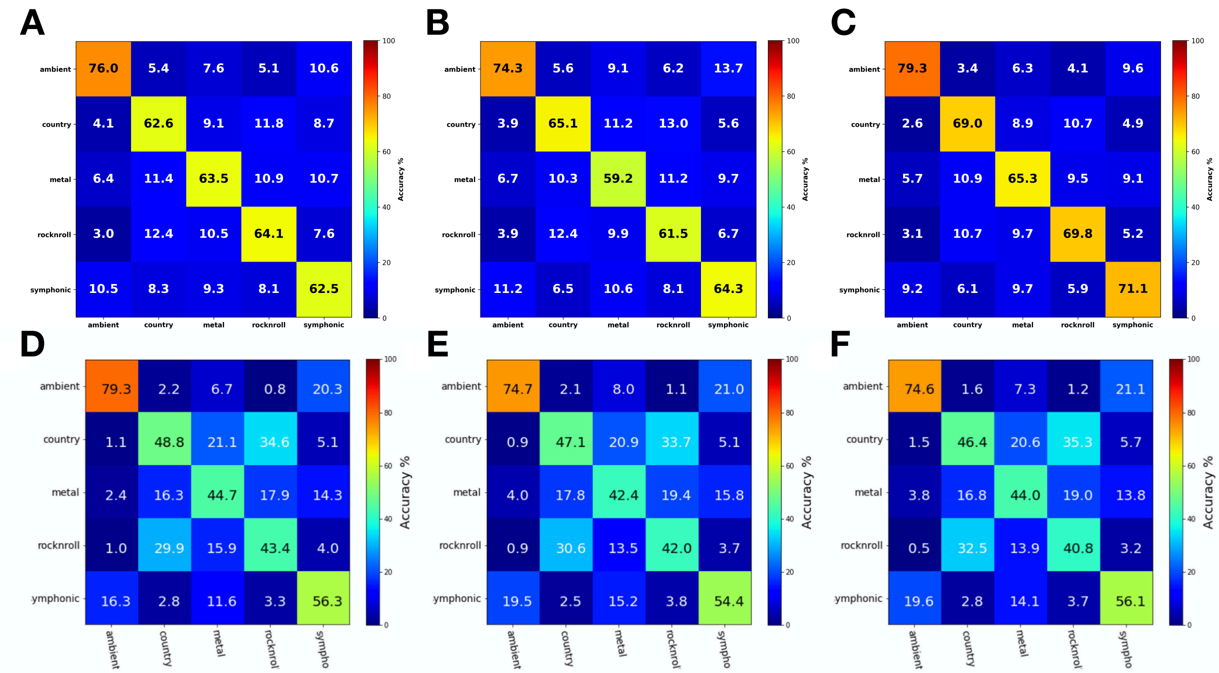

Fig. 3 The comparison between results of this study and Casey's. The result is displayed in the form of a confusion matrix obtained by testing the full dataset of the trained classifier. (A) HG, this study. (B) aSTG, this study. (C)pSTG, this study. (D) HG, Casey's. (E) aSTG, Casey's. (F) pSTG, Casey's.

mentioned above according to the voxel intensity to get the mask of the three brain regions. After extracting the masks of the three regions, we imported the brain functional images of 20 subjects into Python and masked them according to the regions respectively. We obtained 11700 brain maps with respective dimensions. They are 1830, 521 and 604 dimensions for pSTG, aSTG and HG.

After the mask operation, the data of the three regions were respectively divided into 10 groups for training using the GBDT algorithm (parameters are as above) and the 10-fold cross-validation was performed. The confusion matrix was obtained as a result, and we compared it with Casey's result, see Fig. 3. It can be seen that the classification effect of ambient sound performs best in the three regions, reaching 76.0%, 74.3% and 79.3% respectively in HG, aSTG and pSTG. The classification effect of the rest of the music genres is generally higher than 65%. Among the misclassifications, Rocknroll, Metal, and Country music have a greater probability of misclassification, which is between 10% and 15%. Besides, the highest misclassification rate occurs in the aSTG area, between symphony and environmental sounds, raise to 13.7%. In Casey's 2017 study[1], a linear kernel SVM was used to train the types of songs for three identical regions. In terms of accuracy, the three regions have high-resolution accuracy for environmental sounds, reaching 79.3%, 74.7% and 74.6% in the HG, aSTG and pSTG regions respectively, which are close to the results of this paper. As for other music, Casey's classification accuracy rate is low, especially for human voice music (such as rock, metal, and country music) only reached 40% ~ 48% accuracy. In terms of misclassification, Casey also found that vocal music is prone to confusion. For example, the misclassification probability between country music and rock music has reached more than 30%, and the misclassification probability of symphony and environmental sounds has reached about 20%.

B. Analyze 2: Decoding based on Whole Brain Feature Extraction using ANOVA

The second experiment done in this paper is the whole-brain feature searching process using the ANOVA method. This experiment used images masked by a whole-brain average mask as the original data, with a dimension of about 120,000. Using the ANOVA method to take the first 5% of the voxels in the

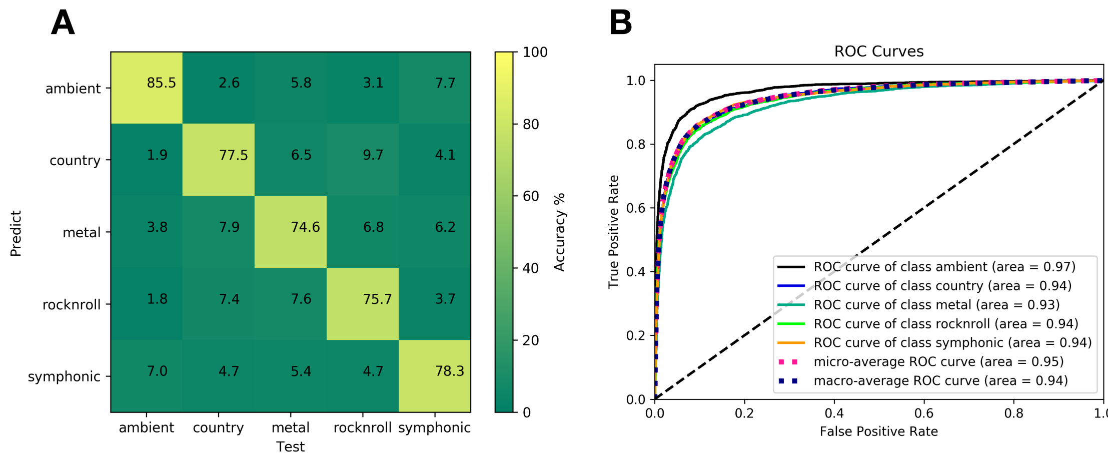

Fig. 4 The confusion matrix and ROC curves for classifier trained during Analyze 2. (A) Confusion matrix. (B) ROC curves figure.

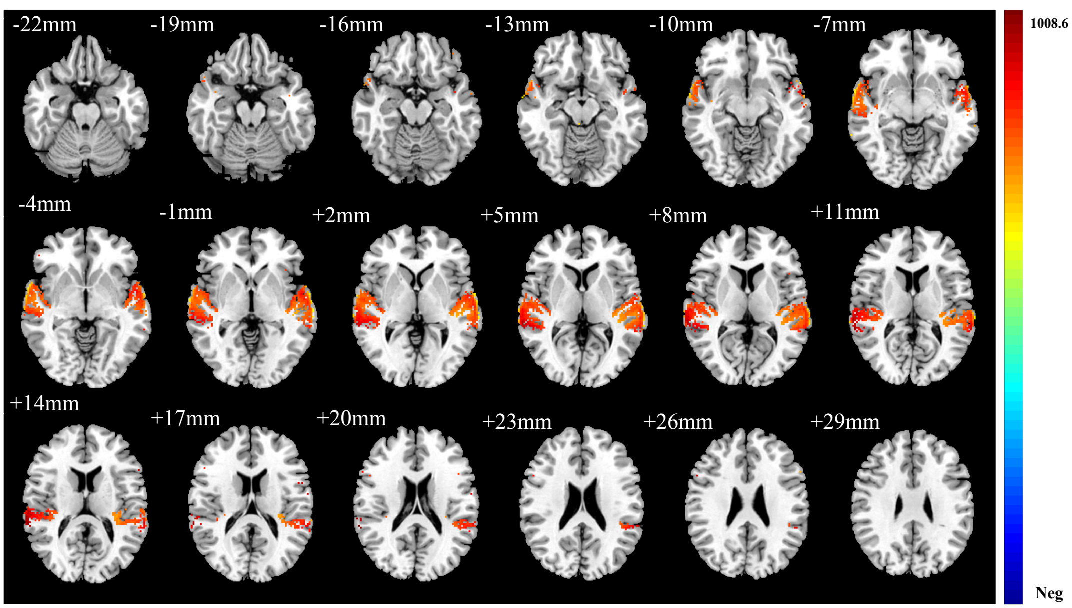

Fig. 5 The contributive regions in decoding from whole brain.

image, that is, around 6000 voxels as features to train the classifier with the GBDT algorithm.

The result of this experiment is better than the accuracy rate of the previous experiment for the classification of specific areas. It can be seen from the confusion matrix (See Fig. 4) that the classification accuracy rate of all genres has reached more than 75%, while the recognition accuracy rate of ambient sounds has exceeded 85%. It can also be seen in the ROC graph that the AUC has reached more than 0.90 for all genres. It could be easily found that the feature voxels extracted from the whole brain through the ANOVA method perform better than voxels in specific regions, which also proves that voxels other than the above three regions are involved in the process of recognizing music. Next, we will explore the specific location of these voxels.

Similarly, the voxel points corresponding to the obtained

Table 1 The number of active voxels in the insula and parietal lobe.

	Total Number of Active Voxels	Insula	Parietal Lobe
Left Cerebrum	1722	127	64
Right Cerebrum	1719	186	19

ANOVA feature vector are projected onto the average image, and the feature voxels obtained by the ANOVA method can be obtained. We used the REST toolbox (https://www.nitrc.org/projects/rest/) to open the projected image to get the rendered image and XJView (https://www.alivelearn.net/xjview/) for the detailed report. From the image rendered by REST (See Fig. 5).

It could be seen that almost voxels activated are in the temporal gyrus, but there are also some in other areas. The detailed report of XJView shows that in the ANOVA feature voxels, in addition to the superior temporal gyrus and transverse temporal gyrus regions extracted in the previous experiment, a large number of voxels located in the parietal and insular lobes are selected as feature voxels (see Table 1). Therefore, we can conclude that in the process of brain activity involved in music distinction, in addition to the traditional auditory cortex, the insula and parietal lobe are also involved in this process.

IV. Discussion

A. *Analyze 1: Specific Area Decoding Accuracy Experiment*

In the first experiment of this paper, through previous research published by Casey in 2017[1], we verified that the superior temporal gyrus and the transverse temporal gyrus in the temporal lobe area played a great role in the process of distinguishing music genres. Compared with Casey, this paper’s experiment got a high classification rate. It can be guessed that the weak performance features reduce the classification ability of linear SVM, and the ensemble algorithm used in this paper has higher adaptability to this kind of data with weakly separable and large amounts of dimensions. As for the misclassifications still existing, it could be guessed that they are due to the similarity of certain physical characteristics, which leads to similar music listening. For example, the symphony music and the ambient music are both without human voice, and they have similar pitch, so that the decoding pattern of them is similar in human brain.

This result validates the results of many previous studies, such as Casey's 2012 study on the sound quality conditions of brain-encoded music [12] and Guntupali's 2013 Ph.D. thesis on whole-brain spatial representation [13]. Besides, Argye et al. also mentioned in 2017 that damage to the pSTG area will seriously affect the ability to recognize music [14]. According to the results, it could be curtained that the temporal lobe, especially the superior temporal gyrus do attend the process of distinct music genres. The result may help those who could not distinct music to find the causes and possible treatments.

B. *Analyze 2: Decoding based on Whole Brain Feature Extraction*

In the second experiment of this paper, we carried out the feature extraction of the whole brain space and found that in addition to the traditional auditory cortex involved in the music recognition process, there are also a large number of voxels in the parietal and insular lobes. Regarding the physiological function of the parietal lobe, Brownsett found through EEG data research that the parietal lobe is involved in the control process of human language understanding and reading and writing [15], while Sarkheil used fMRI data to find that the parietal lobe decodes dynamic facial expressions in the human brain [16]. Insular lobe was once considered to be the brain area that controls the addictive mechanism, but Li research found that music and dance can promote the development of insular lobe and enhance human empathy [17], and He also found that music can promote and enhanced the connectivity of insular lobe which may improve the clinical manifestations of schizophrenia[18]. Therefore, this paper uses fMRI data to prove that the temporal lobe, insular lobe, and parietal lobe are all involved in the music type recognition process. The joint participation of multiple regions proves that the process of listening to music not only involves the identification of physical information in the auditory cortex but also involves human advanced cognitive processes.

C. *Which Methods to Choose: MASK or ANOVA ?*

As we used two methods to do the feature extraction, the MASK and ANOVA, even with different intentions, we could do a little comparison between them.

Obviously, MASK is used to extract specific brain regions, and it is suitable for research on a specific region. However, selecting a suitable brain template is the key to this method. Different brain templates may cause different brain regions to be selected, that is, different feature points, which may lead to differences in results. In contrast, the ANOVA is more like an adaptive method to automatically calculate feature points.

In the experiment, we found that the accuracy of the feature points calculated by the ANOVA method is better than which given by MASK. It could be explained that ANOVA uses voxels activated in the whole brain, while the MASK only concentrate on one specific region. As we found in our conclusion, the activation area when listening to a certain kind of music is not so concentrated, so finding feature points in the whole brain is more effective for classification, which is actually a process of removing noise.

Therefore, when selecting methods for feature extraction, using ANOVA to search the whole brain could be effective, but it could be more complex while doing data pre-processing and time-consuming. When it comes to concentrating on one specific area, find a suitable brain template and do the MASK.

V. Conclusion

This paper proposes a qualitative and quantitative decoding study on the process of human brain music discrimination in the context of brain decoding. The experimental process is based on machine learning to perform pattern recognition on fMRI data to obtain the distinguishing ability of specific voxels for music genres. We first verified if three special areas (anterior superior temporal gyrus, posterior superior temporal gyrus and transverse temporal gyrus) selected by prior research participate in the music genres classification process, and we finally used GBDT algorithm to get a better model with higher accuracy rate. It proves that these three areas had attended to the process when human beings classifying music genres. And after that, we also used ANOVA method to search feature areas in whole-brain, the result shows that apart from those three areas mentioned in the first experiment, the insular lobe and parietal lobe will also activate when listening to specific music genre. At the same time, some studies have shown that the left pSTG area has a certain contribution to participating in language comprehension function [19], and another study through MEG research has shown that the HG area of musicians has a larger gray matter content when receiving stimulation than ordinary people's HG. (Expanded by 130%) [20]. Combined with these research results, we can continue to explore whether music appreciation can promote brain development in specific areas, and the contribution of these promotion effects to children's understanding of language and perception of emotions.

In conclusion, this paper found three brain lobes (superior temporal gyrus and transverse temporal gyrus in the temporal lobe, insular lobe and parietal lobe) are activated when listening and distinguishing different music genres. The result may help to better understanding of the human brain decoding process while listening to music and promote the auditory aspect of brain-computer interface development. The following research may focus on decoding physical features of music, and how to generalize the specific decoding model.

Acknowledgment

This study is supported by National Natural Science Foundation of China (No.61503278 and No.61876126). The authors thank Hanke for opening the fMRI dataset.

References

[1] Casey, Michael A. "Music of the 7Ts: Predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical Music Features." *Frontiers in psychology* 8 (2017): 1179.

[2] Tong, Frank, and Michael S. Pratte. "Decoding patterns of human brain activity." *Annual review of psychology* 63 (2012): 483-509.

[3] Hong KS, Santosa H. Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hear Res*. 2016;333:157-166.

[4] L. Wang et al., "Decoding dynamic auditory attention during naturalistic experience," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 974-977.

[5] Zhang F, Wang JP, Kim J, Parrish T, Wong PC. Decoding multiple sound categories in the human temporal cortex using high resolution fMRI. *PLoS One*. 2015;10(2):e0117303.

[6] Zulkurnaini, Noor Ashikin, et al. "The comparison between listening to Al-Quran and listening to classical music on the brainwave signal for the alpha band." *2012 Third International Conference on Intelligent Systems Modelling and Simulation*. IEEE, 2012.

[7] Teppo Särkämö, Mari Tervaniemi, Sari Laitinen, Anita Forsblom, Seppo Soinila, Mikko Mikkonen, Taina Autti, Heli M. Silvennoinen, Jaakko Erkkilä, Matti Laine, Isabelle Peretz, Marja Hietanen, Music listening enhances cognitive recovery and mood after middle cerebral artery stroke, *Brain*, Volume 131, Issue 3, March 2008, Pages 866–876

[8] Toiviainen, Petri, et al. "Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data." *Neuroimage* 88 (2014): 170-180.

[9] Hanke, Michael, et al. "High-resolution 7-Tesla fMRI data on the perception of musical genres–an extension to the studyforrest dataset." *F1000Research* 4.174 (2015): 174.

[10] Hanke, Michael, et al. "A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie." *Scientific data* 1.1 (2014): 1-18.

[11] Breiman, Leo, et al. *Classification and regression trees*. CRC press, 1984.

[12] Casey, Michael, et al. "Population codes representing musical timbre for high-level fMRI categorization of music genres." *Machine Learning and Interpretation in Neuroimaging*. Springer, Berlin, Heidelberg, 2012. 34-41.

[13] Guntupalli, Jyothi Swaroop. Whole brain hyperalignment: intersubject hyperalignment of local representational spaces. Dartmouth College, 2013.

[14] Hillis, Argye E., Christopher Rorden, and Julius Fridriksson. "Brain regions essential for word comprehension: Drawing inferences from patients." *Annals of neurology* 81.6 (2017): 759.

[15] Sonia L. E. Brownsett, Richard J. S. Wise, The Contribution of the Parietal Lobes to Speaking and Writing, *Cerebral Cortex*, Volume 20, Issue 3, March 2010, Pages 517–523

[16] Sarkheil, Pegah, et al. "Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions." *Social cognitive and affective neuroscience* 8.8 (2013): 950-957.

[17] Gujing, Li, et al. "Increased insular connectivity and enhanced empathic ability associated with dance/music training." *Neural plasticity* 2019 (2019).

[18] He, Hui, et al. "Music intervention leads to increased insular connectivity and improved clinical symptoms in schizophrenia." *Frontiers in neuroscience* 11 (2018): 744.

[19] Ge, Jianqiao, et al. "Cross-language differences in the brain network subserving intelligible speech." *Proceedings of the National Academy of Sciences* 112.10 (2015): 2972-2977.

[20] Schneider, Peter, et al. "Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians." *Nature neuroscience* 5.7 (2002): 688-694.