

Investigating Semantic Relationship between the Scene and the Sound within the Scene by Dynamic Causal Modeling

Xiaojing Wang and Junhai Xu*

School of Computer Science and Technology, Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin 300072, P.R. China

*Corresponding author

Abstract—The purpose of the paper was to explore the dynamic information flow among brain regions, which took part in the perception of the semantic relationship between the scene and the sound of the object within the scene. 15 healthy volunteers were recruited to observe the 4 categories (32 subclasses) of the scene pictures and listened to the 8 categories (64 kinds) of sound corresponding to each scene, and the data were collected and analyzed by functional magnetic resonance imaging. Dynamic causal modeling method embedded in spm8 to analyze the connection among brain regions involved in different experimental tasks. Six candidate models were constructed, and an optimal model was determined by Bayesian model selection. In the optimal model, the flow of information between the parahippocampal place area (PPA) and the lateral occipital complex (LO) under the adjustment of the scene condition was interactive, and the sound condition influenced the activity of the STS. The results showed that the existence of the information flow among the brain regions were selective to the object sound or the scene, which was essential for the processing of the semantic relationship between the human brain's perceived sound and the scene.

Keywords—functional magnetic resonance imaging; dynamic causal modeling; semantic relationship; scene recognition

I. INTRODUCTION

Within the development of functional magnetic resonance imaging (fMRI) technology, a large number of studies have explored the brain specificity and investigated the response of different brain regions corresponding to different experimental tasks [1, 2]. However, this isolated brain-specific response (brain specificity) approach seemed to be inconsistent with the real brain response, so more and more studies have focused on exploring connections between brain regions [3]. Research on the connection of different brain areas under different tasks is more meaningful for human to understand the function of human brain. There are two main methods of brain connection: functional connection and effective connection. Functional connection is undirected, and it is concerned about whether there is the connection among the regions and how much the connection strength is, but do not care about the direction of connection. However, effective connection is directional, it can not only characterize the connection of the regions under the experiment context, but also describe the flow of information among brain regions. Previous studies have revealed two pathways for scene recognition [4-6]: the global scene and the

object in the scene [7]. The reason for the existence of the latter channel was due to the semantic correlation between the scene and the object within the scene, for example, you can quickly and accurately determine that the scene is the bathroom when you see only the bathtub. Furthermore, studies have shown that there is the relationship between the objects and its sound so that the brain region could process information about the same object in two different modes. The article mainly described the dynamic causal model among the human brain during perception the semantic relation between the scene and the sound.

II. EXPERIMENTAL PROCEDURE AND METHODS

A. Experimental Procedure

Block design was used throughout the experiment, no the localizer. The experiment had a total of four runs, each run had 12 blocks (random presentation). Each block had 8 trails, a trail lasted 2.5s. Subjects were asked to name each item silently during the presentation of scene pictures and sound.

The selected regions of interest in this study were obtained by taking the AAL template from the wfupickatlas software together with the actual activation results after correction by FDR ($p < 0.05$, corrected). lateral occipital complex (LOC) is localized in the lateral-ventral occipitotemporal region^[7], and then divided into anterior and posterior segments, namely the posterior fusiform (pF) and the lateral occipital (LO). (parahippocampal place area) PPA was localized in the posterior parahippocampal-collateral sulcus region. (superior temporal sulcus) STS was localized in the posterior part of the superior temporal sulcus.

21 subjects (right-hand, 11 females, 21 ± 2.21 years old) were enrolled in the study. They were with normal hearing and vision and did not have a history of neurological or psychiatric illness. All subjects signed an agreement before the MR scan, which followed the local ethics committee's requirements. After the experiment, each participant got a certain reward.

The scanning device is Siemens Skyra 3.0T scanner using the 8-channel head coil in the Binzhou Medical College, Yantai Affiliated Hospital. Structural anatomical T1-weighted images were obtained using 3D MPRAGE sequences (TR = 1900 ms, TE = 2.52 ms, TI = 1100 ms, voxel size = $0.9375 * 0.9375 * 1 \text{ mm}^3$, matrix size = $256 * 256 * 160$). The T2 image was

scanned with an EPI sequence (TR = 2000 ms, TE = 30 ms, voxel size = 2.5 * 2.5 * 4.6 mm³, matrix size = 64 * 64 * 33).

B. Dynamic Causal Modeling Analysis

DCM is a generalized Bayesian framework for inferring the connections among pre-defined cortical regions based on the time courses of these regions. Specifically, DCM can estimate and infer how a nervous system affects another nervous system, and how it is affected by experimental conditions. This connection analysis method is a verification tool that has been used extensively over the last decade for the fMRI studies of block design. In the previous step, we defined the regions of interest, and extracted the time series from these regions of interest. Then we defined the candidate models representing the possible connections between these ROIs. To compare predictive responses (i.e., predictions with defined models) to measurement responses (time series of extractions), different parameters of these models at neuronal levels were estimated using the hemodynamic forward model. For each particular model, three different sets of parameters need to be estimated: (1) Driving input parameters. Reflecting how the cortical region responds to external stimuli; (2) endogenous connection parameters. Reflecting the basic effective connection between regions (i.e., how a region influences another); (3) modulatory parameters. Quantify how experimental conditions affect effective connectivity. These parameters are represented by Hertz (HZ). An endogenous connection reflects how the increase in activity in a source region affects the changing rate of activity in a target region. Positive (negative) value indicates that the activity of the target area would increase (decrease) as the changing rate of a region increases. For the modulatory connection, the estimated parameter values are added directly to the endogenous parameters, reflecting the context influences the connection.

Model evidence is the probability of observing the measured data given a model. The Bayesian model selection (BMS) is used to identify the model that best predicts the likelihood of the observed data based on comparing their model evidences. The superiority of one model over other models is described by the exceedance probability, which denotes how much the model better fits the given data than others. Bayesian model averaging (BMA) is used to obtain a single model with new connection parameters, which correspond to the average connections for all models.

C. Data Processing

The timing difference between the functional images was corrected by resampling the time slices to the first slice of each volume at the slice timing, then volumes was normalized to the Montreal Neurological Institute (MNI) template. It was vital to note that we recombined the 12 original conditions to 2 conditions (i.e. sound condition and scene condition) in the first-level analysis.

III. EXPERIMENTAL RESULTS

A. Time-series Extraction

DCM analysis focused on connectivity among four pre-defined ROIs: PPA, LO, pF, STS. Time series were

concatenated across the four runs. For each ROI of each subject, we set a 10-mm radius sphere in the center of the peak of activation at the group level, and then extract the principle eigenvariate at uncorrected threshold of $p = 0.05$. Although both sides of each ROI were activated, but for the sake of maintaining our models as simple as possible, we selected only the left hemisphere for DCM analysis due to that DCM is gradually ineffective as the number in the models increases [8, 9].

B. Construction of Model space

Construction of the basic models based on four ROIs. Sound and scene conditions are input to the system as external stimuli. The internal connections between all the regions of the model were bidirectional, and the next step was to determine the modulatory connection. Importantly, effectivity connectivity between regions do not necessarily imply anatomical connections in DCM analysis. To test our specific hypotheses, we constrain the parameters of the model as much as possible based on some strong priori. Finally, six candidate models were built as shown in Figure I. Then these six models are specified, estimated, compared.

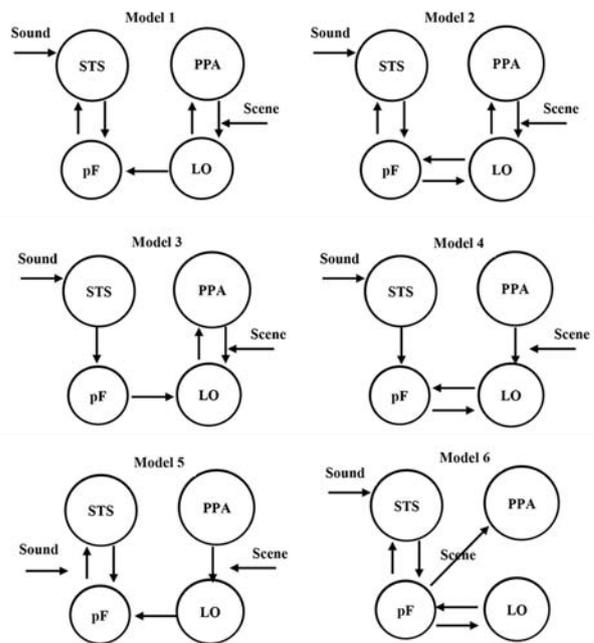


FIGURE I. THE ILLUSTRATION OF THE MODEL SPACE.

C. Model Comparison

The models were compared with the random effects BMS in spm8, and the exceedance probabilities and posterior probabilities (the likelihood of a model for given data) were calculated using the DCM tool (DCM10). It should be noted that the exceedance probability would change as the increasing size of the model space, and all exceedance probabilities sum to 1 over all models tested. For instance, an exceedance probability of 0.9 denotes that this model performs 90% better than other given models for testing given data. The BMS result showed that model 1 exceeds 46% over other models. The

suboptimal model was model 6 with the exceedance probability was 12.6%. The exceedance probability of the remaining models was all less than 12% (the details were specified in Figure II).

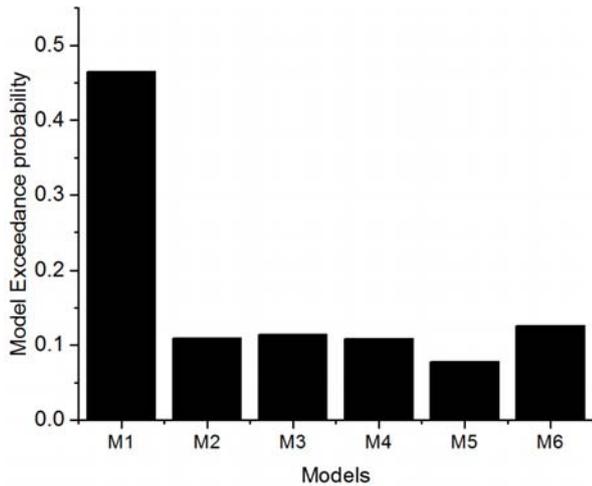


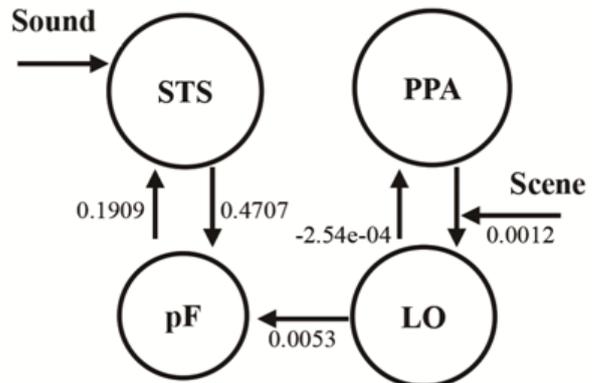
FIGURE II. EXCEEDANCE PROBABILITIES OF THE FIVE MODELS TESTED

The BMA used six models to obtain a single model with average connection parameters (endogenous and tuning parameters). The results were shown in Table I, Figure IIIa and 3b. The endogenous parameters (baseline) from PPA to LO (0.0012HZ), from pF to LO (0.0053HZ) and from pF to STS (0.1909HZ) were positive, suggesting that the enhancement of Ppa activity enhanced the activity of LO, and the enhancement of pF activity enhanced the activity of LO as well as the activity of STs. On the contrary, the connection parameters from LO to PPA (-2.54e-04HZ) and from STS to pF (-0.4707HZ) were negative, indicating that an increase in the activity of all source regions resulted in a decrease in the activity of the target region.

TABLE I. MEAN AND STANDARD DEVIATION (IN ITALICS) OF ENDOGENOUS AND MODULATORY PARAMETERS FOR CONNECTIONS ACROSS ALL SUBJECTS AND ALL MODELS

Connection		Mean	Standard deviation
Endogenous parameters			
From	To		
PPA	LO	0.0012	<i>0.5385</i>
LO	PPA	-2.5438e-04	<i>0.5380</i>
pF	LO	0.0053	<i>0.5379</i>
pF	STS	0.1909	<i>0.4457</i>
STS	pF	-0.4707	<i>0.5024</i>
Modulatory parameters			
From	To		
PPA	LO	0.0006	<i>0.3799</i>
LO	PPA	0.0026	<i>0.3783</i>

(a) Endogenous connections



(b) Modulatory connections

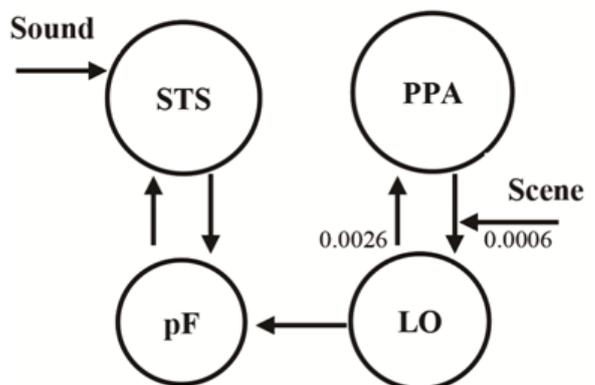


FIGURE III. (A) BMA PARAMETER ANALYSIS RESULTS FOR ENDOGENOUS CONNECTIONS, EXPRESSED IN HZ. (B) BMA ANALYSIS RESULTS FROM MODULATORY CONNECTIONS (EXPRESSED IN HZ)

As for the modulatory parameters under the scene conditions, the significant positive adjustment can be seen: from PPA to LO (0.0006 Hz) and from LO to PPA (0.0026 Hz). These results showed that the scene context affected the effectivity connection between Ppa and LO. Once the modulatory parameters were added to the endogenous parameters, the result indicated that the scene perception resulted in a decrease in the positive connection from Ppa to LO, however, the negative connection from LO to PPA became positive.

IV. DISCUSSIONS

In this paper, we used the dynamic causal modeling analysis to explore the connection of human brain under the condition of sound and scene. We found the existence of the intrinsic connection from PPA to LO, from LO to PPA, from pF to LO, from pF to STS, from STS to pF, and the connection from PPA to LO was influenced by the scene conditions, in particular, positive connections became weaker. The connection from LO to PPA was affected by the condition of the scene, which changed from a negative connection to a positive connection. DCM was performed in the left hemisphere of the regions of interest. There are the evidences that the existence of an anatomical connection between the

occipital region and the high-order region (e.g., fusiform, parahippocampal gyri) [10-12], which adjust the recognition of object, face, and scene [13, 14]. Thus, the result provided an empirical support to the anatomical connection facilitated interaction between connected regions to some extent.

In conclusion, we could show that there are connection interactions between PPA, LO, pF and STS in brain regions involving in perceiving sounds and scenes. This interaction is under the influence of experimental conditions, and the result implies that anatomically interconnected brain regions may contribute to functional interaction.

ACKNOWLEDGMENT

We are grateful to professor MacEvoy for offering us the experimental stimuli which contain the scene of traffic[7].

REFERENCES

- [1] AM K, A DM, LQ U, Z S, DG G, PT R, DS M, FX C, MP M. Development of Anterior Cingulate Functional Connectivity from Late Childhood to Early Adulthood. *Cerebral Cortex*, 2009,19(3): 640-657.
- [2] Schwarz C, Bolz J. Functional specificity of a long-range horizontal connection in cat visual cortex: a cross-correlation study. *The Journal of Neuroscience*, 1991,11(10): 2995-3007.
- [3] Friston KJ. Functional and effective connectivity in neuroimaging: A synthesis. *Human brain mapping*, 1994,2(1-2): 56-78.
- [4] Schyns PG, Oliva A. From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, 1994,5(4): 195-200.
- [5] Peyrin C, Baciú M, Segebarth C, Marendaz C. Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage*, 2004,23(2): 698-707.
- [6] Epstein RA, Higgins JS. Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex*, 2007,17(7): 1680-1693.
- [7] MacEvoy SP, Epstein RA. Constructing scenes from objects in human occipitotemporal cortex. *Nature neuroscience*, 2011,14(10): 1323-1329.
- [8] Peyrin C, Michel CM, Schwartz S, Thut G, Seghier M, Landis T, Marendaz C, Vuilleumier P. The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Cognitive Neuroscience Journal of*, 2010,22(12): 2768-2780.
- [9] Stephan KE, Penny WD, Moran RJ, Ouden HEMD, Daunizeau J, Friston KJ. Ten simple rules for dynamic causal modeling. *Neuroimage*, 2010,49(4): 3099-3109.
- [10] Catani M, Jones DK, Donato R, Ffytche DH. Occipito-temporal connections in the human brain. *Brain A Journal of Neurology*, 2003,126(Pt 9): 2093-2107.
- [11] Catani M, Schotten MTD. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 2008,44(8): 1105-1132.
- [12] Powell HWR, Guye M, Parker GJM, Symms MR, Boulby P, Koepp MJ, Barker GJ, Duncan JS. Noninvasive in vivo demonstration of the connections of the human parahippocampal gyrus. *Neuroimage*, 2004,22(2): 740-747.
- [13] Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A*, 1999,96(16): 9379-9384.
- [14] Kravitz DJ, Saleem KS. The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, 2013,17(1): 26-49.