Coherence Research of Audio-Visual Cross-Modal Based on HHT

Xiaojun Zhu*, Jingxian Hu, and Xiao Ma
College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China
*Corresponding author, Email: zhuxiaojun@tyut.edu.cn,{hjxocean, tyut2010cstc}@163.com

***Abstract*—Visual and aural modes are two main manners that human utilize to senses the world. Their relationship is investigated in this work. EEG experiments involving mixed aural and visual modes are designed, utilizing Hilbert-Huang Transform (HHT) and electroencephalogram (EEG) signal processing techniques. During EEG data processing, I-EEMD method of similar weighted average waveform extension is proposed to decompose the EEG signals, specifically accounting for the problem of end effects and mode mixing existing in the traditional HHT. The main components of are obtained after decomposing the signals including mixed modes with I-EEMD respectively. The correlation coefficient of consistent and inconsistent mixed signal is calculated, and the comparison is made. Investigation on the comparison condition of the correlation coefficient indicates that there is coherence in both the visual and aural modes.**

***Index Terms*—EEG; Audio-visual; Coherence; HHT; EEMD**

I. INTRODUCTION

Human obtain information from the outside world with different sensory channels such as vision, hearing, touch, smell and taste. However, the role of different sensory modalities for human memory and learning are not independent of each other. The encouraging research results" Cross modal learning interaction of Drosophila" of Academician Aizeng Guo and Dr. Jianzeng Guo of Chinese Academy of Sciences proves that there are mutually reinforcing effects between modal of visual and olfactory in drosophila's learning and memory [1]. Then, whether the human' visual and auditory are able to produce a similar cross-modal collaborative learning effect? Can we take advantage of this learning effect to strengthen the information convey efforts, and then produce the effect of synergistic win - win and mutual transfer on memory? Human beings obtain and understand the information of the outside world by multiple sensory modals [2] [3]. However, the information from multiple modals sometimes may be consistent, and sometimes may be inconsistent. So that it requires the brain to treat and integrate the information and form a unified one. Since vision and hearing are the primary ways to percept the outside world for human [4], the coherence research for the information in the visual and audio channels is particularly important, and it also has the extraordinary significance for discovering the functional mechanism of the brain. Therefore, the coherence research of audio-visual information and its function in knowing the world and perceiving the environment will contribute to improving the lives of the handicapped whose visual or audio channel is defective, and make the reconstruction of some functions in their cognitive system come true [5].

Meanwhile, it will also give the active boost to improve the visual and audio effect of the machine and further develop the technology of human-computer interaction. Human brains' integration to the visual and auditory stimuli of the outside world is a very short but complicated non-linear process [6] [7]. In recent years, EEG is widely used in the visual and audio cognitive domain. EEG is the direct reflection of the brain electrical physiological activity, of which the transient cerebral physiological activities are included [8] [9] [10]. Accordingly, some researchers consider the transient process of the brain integrating the visual information and audio information mutually can cause the electric potential on the scalp surface to change [11]. Event-Related Potential (ERP) is the brain potential extracted from EEG and related to the stimulation activities. It can establish the relations between the brain responses and the events (visual or auditory stimulus), and capture the real-time brain information processing and treating process. Thus, in recent years, when the researchers in the brain science field and artificial intelligence field are studying the interaction in multiple sensory and crossing modals, the technology for analyzing ERP has be paid attention to unprecedentedly.

In this article, we will discuss the coherence between the visual EEG signal and the audio EEG signal from the perspective of signal processing based on Hilbert-Huang Transform (HHT) [12], and then investigate the mutual relations between the visual and audio modals. Firstly, this paper designs a visual and auditory correlation test experiment; evoked potential data under the single visual stimulus, the single audio stimulus, and the stimulus of audio-visual consistence and the stimulus of audiovisual inconsistence were collected respectively. Then, he main IMF components of single visual signal and single audio signal are decomposed by HHT, and analysis the coherence of visual and audio modals by calculating the correlation coefficient between these components and ERP signal under the stimulus of visual & audio modals. Our paper is organized as follows. Section II describes

doi:10.4304/jmm.9.2.230-237

Experiment and Records of Audio-visual Evoked. Section III describes Data Treatment Method in detail. We analyze the I-EEMD Processing and Analysis of Experimental Data, and provide the simulation results in Section IV. In Section V, We conclude this paper.

II. EXPERIMENT AND RECORDS OF AUDIO-VISUAL EVOKED EEG

A. Design of Experiment

The experiment contents include single visual experiment (experiment A), single audio experiment (experiment B) and stimulus experiment of audio-visual modals (experiment C). The stimulation experiment of audio-visual modals is also divided into experiment of consistent audio-visual modals (experiment C1) and experiment of inconsistent audio-visual modals (experiment C2). The materials for visual stimulus include seven elements in total, namely Chinese characters "ba", "ga", "a", the letters "ba", "ga", "a" and a red solid circle. The size and lightness of presented Chinese characters or the letters are consistent, and appeared by pseudorandom fashion; the materials for audio stimulus include four sound elements, namely the sound "ba", "ga", "a" and a short pure sound "dong". The sound file is edited by use of Adobe audition with the unified attribute of two-channel stereo, sampling rate of 44100 and resolution ratio of 16-bit. In the experiment of visual evoked potentials, the red solid cycle is the target stimulus, the pictures of other Chinese characters or letters are the non-target stimuli; In the experiment of audio evoked potentials, the short pure sound "tong" is the target stimulus, the other sounds are the non-target stimuli; In the audio-visual dual channels experiment, the visual pictures and the audio sounds combine randomly. When the picture is red solid cycle and the sound is short pure sound "dong", it is target stimulus, and other combinations are non-target stimuli. The experiment requires the subject to pushing a button to indicate their reactions for the tested target stimuli. This experiment model investigates the ERP data under the non-note model. Three groups of experiment stimulus are all OB (Oddball Paradigm) with the target stimulus rate of 20%. The lasting time for every stimulus is 350ms with the interval of 700ms. Three groups of experiments all include 250 single stimuli, of which 50 are target stimulus (trials). The software E-prime is used to implement this experiment.

B. Conditions to be Tested

20 healthy enrolled postgraduates with no history of mental illness (including 10 males and 10 females, right-handed, and their age from 22 to 27 years old) were selected as the subjects. All of them have normal binocular vision or corrected vision and normal hearing. Before the experiment, all of them have signed the informed consent form of the experiment to be tested voluntarily. Before the experiment, scalp of the subjects was kept clean. After the experiment, certain reward was given. Every subject participated in the experience for about one hour, including the experiment preparation and formal experiment process. In the process of the experiment, it was arranged that every subject had three minutes for resting, to prevent the data waveform of ERP being affected because of the subjects' overfatigue.

C. Requirements of Experiment and Electrode Selection

This EEG experiment was arranged to be completed in an independent sound insulation room. The subject faces toward the computer monitor and pronunciation speaker. The subject was 80cm away from the screen, the background color was black. In the process of experiment, the subjects were required to be relax, not nervous, keep the sitting posture well, concentrate, not to twist the head, stare at the computer screen with eyes, press the key "space" when the target stimulus appeared, and not to react to the non-target stimulus. EEG data was recorded by NEUROSCAN EEG system with 64-lead. The electrode caps with the suitable size were worn by the subjects. The electrode caps were equipped with Ag-AgCl electrodes. 10-20 system which is used internationally was employed for the electrode placement. Its schematic diagram is shown as figure 1. The conductive paste is placed between electrode and subject's scalp. It is required that the impedances of all leads should be lower than 5KΩ.

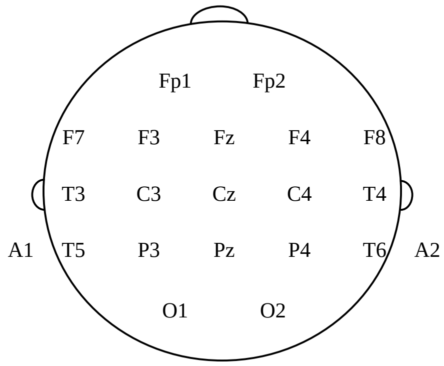


Figure 1. 10-20 Electrode Lead System

In this experiment, 64-lead EEG acquisition system of NEUROSCAN is adopted. However, according to the needs of the experiment, we only use 12 leads among them for analysis. According to the human brain scalp structure partitions and their functions, the visual activities mainly occur in occipital region, and the leads O1, O2 are chosen for analysis; The audio region is located at temporal lobe, and the leads T3, T4, F7, F8 related to the audio activities are chosen for analysis; In addition, the leads F3, F4, Fp1, Fp2 related to the stimulation classification at frontal lobe and the leads C3, C4 related to the whole brain information treatment process are chosen for analysis.

D. Records and Pre-treatment of EEG Signal

The process of EEG experiment is safe and harmless to human body, the time resolution ratio is also extremely high. Therefore, it plays a more and more important role in the field of cognitive science. The parts contained in EEG experiment designed by this paper include lead, electrode cap, signal amplifier, stimulation presentation computer, data recording computer and ERP synchronization software. The details are shown in figure 2:

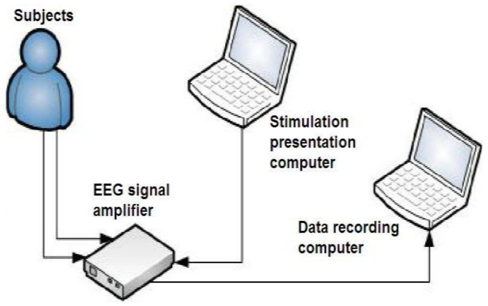


Figure 2. Experimental system equipment

In the experiment, 64-lead EEG equipment of NEUROSCAN is used and the storing data is collected. Quik-Cap of NEUROSCAN is employed for the electrode cap, and the caps of all electrodes are marked. The location of electrodes is simple and fast. The data collection is realized by AC. The reference electrodes are placed at the nose tip. Bilateral mastoid is the recorded electrode. The data sampling rate is set as 1000Hz. Before performing HTT analysis treatment to EEG, it is necessary to pre-treat the recorded and stored EEG data. The general methods include several procedures, namely, eliminating electro-oculogram (EOG), digital filtering, dividing EEG into epochs, baseline correction, superposition average and group average [13].

III. Data Treatment Method

In the process of the EEG signal processing with traditional HHT, the problems such as end effect [14] and mode mixing [15] may be caused, so very great affect will be brought to the experiment results. Therefore, based on a great number of researches which are implemented for the existing solutions, this paper puts forth an end extension algorithm of similar waveform weighted average to restrain the end effect. Meanwhile, EEMD is used to replace EMD to eliminate the mode mixing. The combination of these two methods is named as I-EEMD (Improved-EEMD). The relevant methods are described in details below.

A. Extension Algorithm of Similar Waveform Weighted Average

So far, EMD method has been widely applied in several fields of signal analysis. Although this method has the advantage which is not possessed by other methods, the problem of end effect will bring great obstacles to the application in practical uses. For the problem of end effect, the researchers have brought forward some solutions, such as mirror extension method [16], envelope extension method [17], cycle extension method [18] and even continuation [19] etc. These methods can reduce the influence of end effect in some extent. However, EEG signal is a typical nonlinear and non-stationary signal and it has high requirements for the detail feature of the signal [20] during analysis and treatment. Therefore, these methods still need to be improved. For the end extension, the continued signal must be maintained with the variation trend inside the original signal. After analyzing all kinds of end extension methods, this paper puts forth a method of similar waveform weighted matching to extend the ends.

Definitions $S_1(t)$, $S_2(t)$ are set as two signals on the same time axis, $P_1(t_1, S_1(t_1))$ and $P_2(t_2, S_2(t_2))$ are two points on $S_1(t)$ and $S_2(t)$ respectively. The condition of $t_1 \neq t_2$ is satisfied but $S_1(t_1) = S_2(t_2)$. Here the condition of $t_1 < t_2$ is set. The signal $S_1(t)$ is moved right with the length of $(t_2 - t_1)$ horizontally along the time axis t, to make the points P_1 and P_2 coincide. Along the coincident point P_1 takes the wave form section with the length of L on the left (or right), and the waveform matching degree m of the signals $S_1(t)$ and $S_2(t)$ for the point P_1 (or P_2) can be defined as:

$$m = \frac{\sum_{i=1}^{L}[S_2(i) - S_1(i)]^2}{L}. \qquad (1)$$

Apparently, the more $S_1(t)$ and $S_2(t)$ are matching, the less the *m* value will be.

According to the signal analysis theory we know that, the similar waveform will appear in the same signal repeatedly, so that we can choose a number of matching waves similar to the waveform at the end. Moreover, weighted averaging is performed to them, then, the obtained average wave is used to extend the signal ends. The extension for the signal ends generally includes two ends on the left and right. In the following, the left end of the signal is taken as the example. The original signal is set as $x(t)$, the leftmost end of $x(t)$ is $x(t_0)$, the rightmost end is $x(t^{'})$, and the signal contains *n* samplings points.

Starting from the left end $x(t_0)$ of the signal, part of the curved section of $x(t)$ is taken from the right, and this curved section is set as $w(t)$, which needs to only contain an extreme point (either the maximum value or the minimum value) and a zero crossing point. The length of $w(t)$ is l .The right end of the curved section $w(t)$ can be set as one zero crossing point, and it is recorded as $x(t_1)$. The intermediate point $x(t_{m1})$ in the horizontal axis of $w(t)$ is taken, of which $t_{m1} = (t_0 + t_1)/2$. Taking $x(t_{m1})$ as the reference point, the sub-wave $w(t)$ is moved to the right horizontally along the time axis t. When some point $x(t)$ on the signal $x(t)$ coincides with $x(t_{m1})$, the sub-wave with the same length of $w(t)$ and the point $x(t_i)$ as the central point is taken and recorded as $w_i(t)$. The wave form matching degree m_i of $w_i(t)$ and $w(t)$ is calculated, and the wave form matching degree m_i as well as a small section of data wave (the wave form length of this section is set as $0.1\,l$) in the front of $w_i(t)$ are stored. Move it to the right horizontally with the same process, and successively record these adjacent data waves on the left with the length of $0.1l$ as $v_1(t)$, $v_2(t)$... $v_k(t)$. Finally, a data pair collection comprised of the wave form matching degree and corresponding sub-waves in the adjacent part on the left of the matching waves is obtained:

$$[V,m]=\left\{(v(t),m)\middle|(v_1(t),m_1),(v_2(t),m_2)\cdots(v_k(t),m_k)\right\} \quad (2)$$

If the collection $[V,m]$ is null, it indicated that the wave form of the original signal is extremely irregular. It is not suitable to adopt the theory of similar wave form, and the extension is not performed to it. The extreme value point method is used to solve it. If the collection $[V,m]$ is not null, all the obtained value of wave form matching degree is ranked in sequence from small one to large one. Obtain $[V',m']$ and the first j data pair of $[V',m']$ is taken out, of which $j=[\sqrt[3]{k}]$. The weighted average v_p of all sub-waves in these j data pairs is calculated, and then v_p is used to extend the left end point of $x(t)$ of the signal.

The end extension algorithm for similar wave form weighted matching is as follows.

Input: signal $x(t)$.
Output: matching wave of weighted average v_p
Steps: (1) For $t=t_0$ to t'; (2) Calculate the waveform matching degree m_i according to the formula (1), and take part of sub-waves v_i on the left of the matching wave w_i. L(v_i)=0.1*L(w_i); (3) End for; (4) Rank the collection [m_i', v_i'] in sequence from small one to large one according to the value of m_i, and obtain the new collection [m_i', v_i'] with the length of k; (5) Take the first j data pairs of [m_i', v_i'] of which $j=[\sqrt[3]{k}]$; (6) Calculate the weighted average wave v_p. (7) Use v_p to extend the left end of signal $x(t)$.

B. Eliminating Mode Mixing Problem

The problem of mode mixing is often coming out in the process of EEG signal decomposition with EMD method, and its reason is relatively complex. Not only the factors of EEG itself can cause it, such like the frequency components, sampling frequency and so on, but also the algorithms and screening process of EMD. Once mode mixing problem is appearing, the obtained IMF components would lose the physical meanings they should have, and would bring negative influence to the correct analysis of the signals.

N.E. Huang did a lot of research on the EMD of white noise [21], and he found that the energy spectrum of white noise is uniform over the frequency band, and its scale performance in time-frequency domain is evenly distributed. At the same time, a French scientist, Flandrin, after doing a lot of EMD decompositions to white noise and in the base on statistics, also found that all the various frequency components it contains can be isolated regularly. That is to say, for white noise, EMD method acts as a binary filter, and each IMF component from decomposition has a characteristic of similar band pass in the power spectrum [22]. Based on the characteristics of white noise in EMD method and in order to better solve mode mixing problem, Z.Wu and NEHuang have proposed a noise-assisted empirical mode decomposition method on the basis of original EMD decomposition. And this new method is called as "EEMD", which means ensemble empirical mode decomposition [23].

The specific steps of EEMD algorithm are as follows:

1) Add a normal distributed white noise x(t) to the original signal s(t), and then obtain an overall S(t):

$$S(t)=s(t)+x(t) \quad (3)$$

2) Use standard EMD method to decompose S(t), which is the signal with white noise, and then decompose it into a plurality of IMF components c_i, and a surplus component of r_n:

$$S(t)=\sum_{j=1}^{n}c_i+r_n \quad (4)$$

3) Repeat step 1), 2) and add the different white noises to the to-be-analyzed signals:

$$S_i(t)=s(t)+x_i(t). \quad (5)$$

4) Decompose the superposed signals from the previous step in EMD method, and then obtain:

$$S_i(t)=\sum_{j=1}^{n}c_{ij}+r_{in}. \quad (6)$$

5) The added several white noises are random and irrelevant, and their statistical mean value must be zero. And do the overall average for each component to offset the influence of Gaussian white noise, and then obtain the final decomposition results:

$$c_j=\frac{1}{N}\sum_{i=1}^{n}c_{ij}. \quad (7)$$

In the formula, N means the number of added white noises.

IV. I-EEMD Processing and Analysis of Experimental Data

A. I-EEMD Processing and Analysis of Experimental Data

In the following, the evidence of existing coherence of audio-visual modals would be discussed from the perspective of EEG signals processing. So, it is needed to extract the main components of audio-visual evoked potentials and analyze them. We choose C3 and C4 leads, which related to the whole-brain information processing to analyze. The C3 lead is taken for example, and its ERP waveforms of single visual stimulus, single auditory stimulus, consistent visual & auditory data and inconsistent visual & auditory data are obtained shown as Fig. 3:

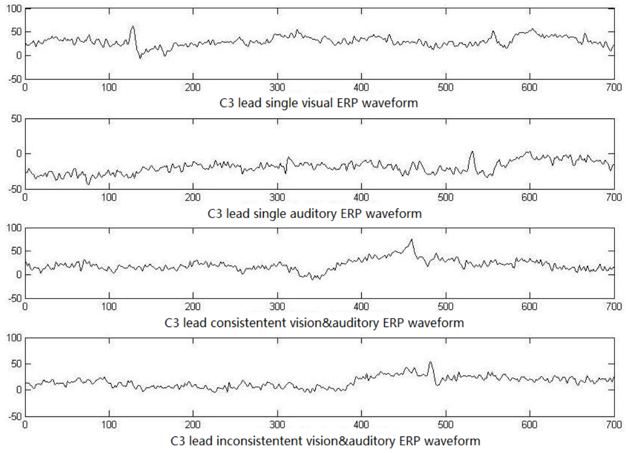
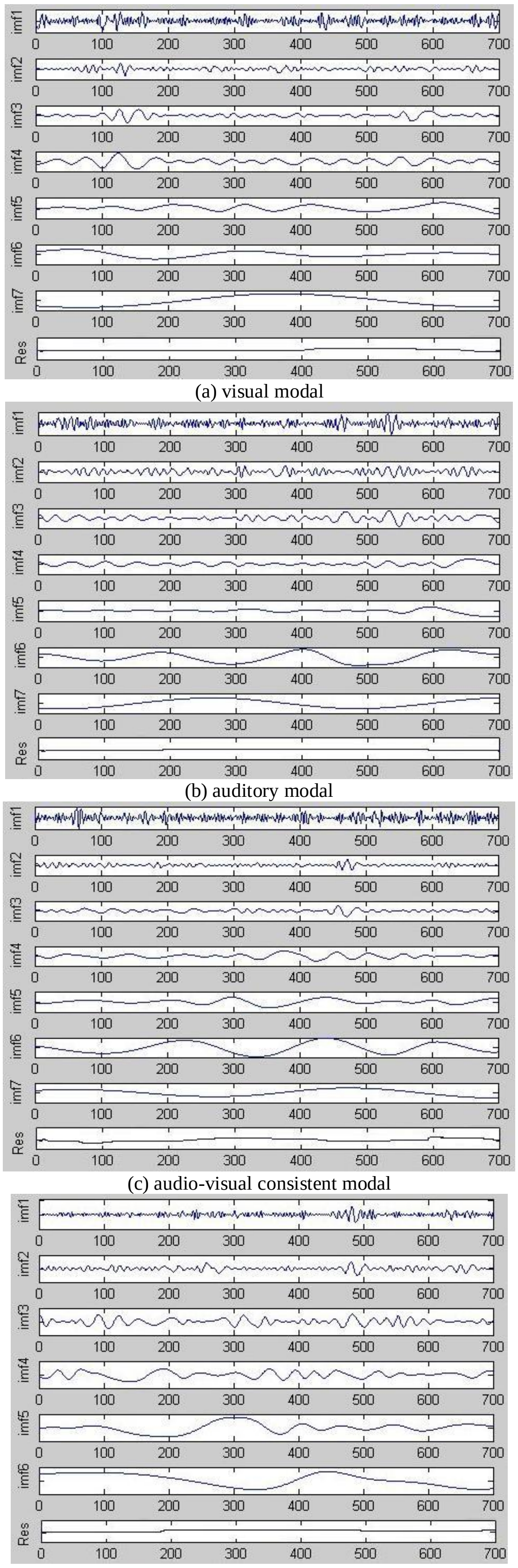


Figure 3. C3 Lead ERP Waveform

In Fig. 3 it shows four kinds of ERP data waveforms on the length of one epoch. Among them the selected stimulating text for the visual stimulation is the screen text letter "ba"; the selected stimulating material for the auditory stimulus is the sound "ba" from the audio; the stimulating material for the audio-visual consistent is the letter "ba" and sound "ba"; and the selected stimulating material for audio-visual inconsistent is the letter "ba" and sound "ga".

After decomposing the above four kinds of ERP data in the I-EEMD method, all the IMF components are obtained as Fig. 4 shown. Every component is distributed according to the frequency from high to low. For the point of decomposition effect, each component is relatively stable at the end, and there was also no flying wing phenomenon, and producing no significant mode mixing problem. Each component has a complete physical meaning, and through these components it could examine the advantages and disadvantages of decomposition effects. And the fact that Res, the surplus component is close to zero, once again proves the validity of the proposed method in this paper.And from that figure it can be seen, through the I-EEMD decomposition the VEP data of C3 lead decomposed into 7 IMF components and a surplus component. Among these seven IMF components, there may be some pseudo-components, which could be screened out by the method of correlation coefficients calculation. Through the I-EEMD decomposition the AEP data of the C3 leads turned into seven IMF components and a surlpus component. The audio-visual consistent data of C3 leads through I-EEMD decomposition turned into seven IMF components and a surplus component. The audio-visual inconsistent data of C3 lead through I-EEMD decomposition turned in six IMF components and a surplus component.

Among the seven IMF components of the VEP data through I-EEMD decomposition, there usually are some pseudo-components, which should be screened out and not taked into consideration. Through the relavent theory of signals, the varity of IMF components could be judged. The dicision threhold value of the pseudo-component is one-tenth of the biggest correlation coefficient [24]. Caculate the correlation coefficients between these seven IMF componnets and original signal, and they are shown in Table 1.



(a) visual modal
(b) auditory modal
(c) audio-visual consistent modal
(d) audio-visual inconsistent modal

Figure 4. I-EEMD Decomposition of ERP Data

TABLE I. THE CORRELATION COEFFICIENT BETWEEN IMF COMPONENTS AND ORIGINAL SIGNAL

IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	Res
0.0153	0.9435	0.5221	0.0275	0.7433	0.7028	0.0649	0.0032

TABLE II. THE CORRELATION COEFFICIENT BETWEEN IMF COMPONENTS AND ORIGINAL SIGNAL

IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	Res
0.0211	0.6541	0.9022	0.5728	0.0325	0.0411	0.0353	0.0049

As Table 1 shown, the correlation coefficients of IMF1, IMF4, IMF7 with the original signal are relatively low, which are 0.0153, 0.0275 and 0.0649. From that we could say that these three components are pseudo-components obtained from the decomposition. And the correlation coefficient between surplus component and the original signal is 0.0032. So these four decomposition components don’t have real physical meanings and don’t deserve deeper analysis. The correlation coefficents of IMF2, IMF3, IMF5, IMF6 with the original signal are relatively high, so they are the effective components from decomposition.

Similarly, do the same process to the IMF components from AEP data and obtain Table 2

It could be seen from Table 2 that the correlation coefficients of IMF1, IMF5, IMF6 and IMF7 with original signal are relatively low, which are 0.0211, 0.0325, 0.0411 and 0.0353. From that we could say that these four components are pseudo-components coming from decomposition and don’t have real physical meanings. The correlation coefficients of IMF2, IMF3, IMF4 with orignal signal are relatively high, which means they are the effecitve components from decomposition.

B. Analysis of Experiment Result

It could analyze the coherence of audio-visual modals by comparing the correlation coefficients of ERP signal in single visual or auditory modal with the ERP signal in audio-visual modals. Due to the sound is “ga” when audio-visual are inconsistent, it should choose the sound as “ga” when considering compare the ERP data in single auditory modal and audio-visual inconsistent modal. The other situations should take the sound “ba”. The comparison situation of correlation coefficient calculating from the above experiment data is shown in Table 3:

TABLE III. COMPARISON OF THE CORRELATION COEFFICIENT

Correlation coefficient value	Single visual ERP data	Single auditory ERP data
Audio-visual consistent ERP data	0.5331	0.4519
Audio-visual inconsistent ERP data	0.2379	0.2022

It could be seen from Table 3 that the signal correlation coefficient of ERP data between single visual modal and audio-visual consistent modal is 0.5331, while the signal correlation coefficient of ERP data between it and audio-visual inconsistent is 0.2379; the signal correlation coefficient between the ERP data of single auditory stimuli and that of audio-visual consistent is 0.4519, while the signal correlation coefficient of ERP data between it and audio-visual inconsistent is 0.2022. So from that we could say, when the information in audio-visual modal is consistent, it could improve the information in single modal; while, when the information in audio-visual modal is not consistent, it could have inhibitory effect on single-modal state information.

In addition, could also find some evidence to support the above points when considering the main compositions of ERP signal in single audio or visual stimulus and the correlation of ERP signal in audio-visual modals mixing stimuli.

From the principle of EMD decomposition, we could know that the IMF components got from decomposition have complete physical meaning. So it could inspect the coherence of audio-visual modals through the main compositions of single visual stimulus, single auditory stimulus and the correlation coefficient of the ERP data of audio-visual consistent and inconsistent.

To compare the valid components of single visual evoked potentials and single auditory evoked potentials, and compare the correlation coefficients of ERP data of audio-visual consistent and inconsistent, we could get the data in Table 4:

TABLE IV. VISUAL COMPOSITION'S COMPARISON OF THE CORRELATION COEFFICIENT

Correlation coefficient	IMF2	IMF3	IMF5	IMF6
Audio-visual consistent data	0.5111	0.3853	0.5037	0.4202
Audio-visual consistent data	0.2195	0.1528	0.3001	0.2673

From Table 4 we could see that the correlation coefficients between the main compositions of single visual stimulus evoked potentials and the ERP signal in audio-visual consistent are obviously greater than the correlation coefficients between it and ERP signal in audio-visual inconsistent. And that also shows that when the audio-visual information is consistent, it could help people to prompt information-controlling power of outside world. Then let’s look at the corresponding data in audio modal. Because in experiment under the situation of audio-visual inconsistent, we select sound “ga” and letter “ba” as the to-be-analyzed data. In order to get a better comparability of the experiment results, here we choose the sound “ga” in the auditory modal to compare with the data in audio-visual cross-modal data, and get the results shown as Table 5:

TABLE V. AUDIO COMPOSITION'S COMPARISON OF THE CORRELATION COEFFICIENT

Valid auditory components	IMF2	IMF3	IMF4
Correlation coefficient of audio-visual consistent	0.6232	0.7869	0.5466
Correlation coefficient of audio-visual inconsistent	0.2752	0.3456	0.0387

From Table 5 it could be seen that the comparison is similar to what in the visual modal, which is to say that, the correlation coefficients between the main compositions of single auditory evoked potentials and the ERP signal in audio-visual consistent modal are apparently greater than what of ERP signal in audio-visual inconsistent modal. And what could also show that when the information in audio-visual is consistent, the EEG signal will be stronger than single modal auditory information in the brain.

V. CONCLUSIONS

This paper discusses the theory evidence of audio-visual modals' coherence from the perspective of EEG signal processing. And the experiment is designed based on the EEG in audio-visual cross-modal and then collect, process and analyze the experiment data. During the process of EEG signal, it uses EEMD, which combined the similar waveform average end continuation algorithms, which is called in paper as I-EEMD, considering the need to restrain the effects of mode mixing and end point effect. And try to describe the collected ERP data through two perspectives based on the theory of signal coherence. First, investigated the correlation of the data in single visual modal, single auditory modal and the data in audio-visual cross-modal. From the calculated correlation coefficient it could be found that, when audio-visual is consistent, the correlation coefficient of it with any single modal is relatively great. Second, form the main valid compositions in single visual or audio modal, the comparison situation of the calculated correlation coefficient is similar. From these two points we could find that, when the informations in audio & visual modals are consistent, it could help the brain promptly handle outside environment, and that is to say, it could improve each other's information under the condition of audio-visual consistent; when the informations in these two modals are not consistent, it could restrain each other's information and the brain could get a combined result after integrated, which is consistent with the famous McGulk effect.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation for Young Scientists of Shanxi Province, China (GrantNo.2013021016-3).

REFERENCES

[1] Guo, Jianzeng and Guo, Aike. Cross modal interactions between olfactory and visual learning in Drosophila, Science, vol. 309, no. 5732, pp. 307–310, 2005.

[2] Suminski, Aaron J and Tkach, Dennis C and Hatsopoulos, Nicholas G. Exploiting multiple sensory modalities in brainmachine interfaces, *Neural Networks*, vol. 22, no. 9, pp. 1224–1234, 2009.

[3] Ohshiro, Tomokazu and Angelaki, Dora E and DeAngelis, Gregory C. A normalization model of multisensory integration. *Nature Neuroscience*, vol. 14, no. 5, pp. 775–782, 2011.

[4] Nishibori, Kento and Takeuchi, Yoshinori and Matsumoto, Tetsuya and Kudo, Hiroaki and Ohnishi, Noboru. "Finding the correspondence of audio-visual events by object manipulation. *in IEEJ Transactions on Electronics, Information and Systems*, 2008, pp. 242–252.

[5] Mitchell, T. AI and the Impending Revolution in Brain Sciences. *in Eighteenth National Conference on Artificial Intelligence*, July 28-August 1, 2002, Menlo Park, United States.

[6] Ethofer, Thomas and Pourtois, Gilles and Wildgruber, Dirk. Investigating audiovisual integration of emotional signals in the human brain" *in Progress in Brain Research*, vol. 156, pp. 345–361, 2006.

[7] Gonzalo-Fonrodona, Isabel. Functional gradients through the cortex. multisensory integration and scaling laws in brain dynamics. *in Neurocomputing*, vol. 72, no. 4C6, pp. 831-838, 2009.

[8] Liu, Baolin and Meng, Xianyao and Wang, Zhongning and Wu, Guangning. An ERP study on whether semantic integration exists in processing ecologically unrelated audio-visual information. *in Neuroscience Letters*, vol. 505, no. 2, pp. 119-123, 2011.

[9] Lee, TienWen and Wu, YuTe and Yu, Younger WY and Chen, Ming-Chao and Chen, Tai-Jui. The implication of functional connectivity strength in predicting treatment response of major depressive disorder: A resting EEG study. *In Psychiatry Research-Neuroimaging*, vol. 194, no. 3, pp. 372-377, 2011.

[10] Blankertz, Benjamin and Tomioka, Ryota and Lemm, Steven and Kawanabe, Motoaki and Muller, K-R. Optimizing spatial filters for robust EEG single-trial analysis. in IEEE Signal Processing Magazine, vol. 25, no. 1, pp. 41-56, 2008.

[11] Molholm, Sophie and Ritter, Walter and Murray, Micah M and Javitt, Daniel C and Schroeder, Charles E and Foxe, John J. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive brain research*, vol. 14, no. 1, pp. 115-128, 2002.

[12] Yuan, Ling and Yang, Banghua and Ma, Shiwei. Discrimination of movement imagery EEG based on HHT and SVM. "Chinese Journal of Science *Instrument*, vol. 31, no. 3, pp. 649-654, 2010.

[13] Coyle, Damien and McGinnity, T Martin and Prasad, Girijesh. Improving the separability of multiple EEG features for a BCI by neural-time-series-prediction preprocessing. *Biomedical Signal Processing and Control*, vol. 5, no. 3, pp. 649-654, 2010.

[14] He, Zhi and Shen, Yi and Wang, Qiang. Boundary extension for Hilbert-Huang transform inspired by gray prediction model." *Signal Processing*, vol. 92, no. 3, pp. 685-697, 2012.

[15] Lee, Ray F and Xue, Rong. A transmit/receive volume strip array and its mode mixing theory in MRI. Magnetic Resonance Imaging, vol. 25, no. 9, pp. 1312-1332, 2007.

[16] Jin Ping, ZHAO and Da ji, Huang. Mirror extending and circular spline function for empirical mode decomposition method. Journal of Zhejiang University, vol. 2, no. 3, pp. 247-252, 2001.

[17] Gai, Q. Research and Application to the Theory of Local Wave Time-Frequency Analysis Method. *PHD, Dalian University of Technology, Dalian*, 2001.

[18] Hamilton, James Douglas. Time Series Analysis; *Cambridge Univ Press*, vol. 2, 2001.

[19] Qiao Shijie. The Symmetric Extension Method for Wavelet Transform Image Coding." *Journal of Image and Graphics*, vol. 5, no. 9, pp. 725-729, 2005.

[20] Pigorini, Andrea and Casali, Adenauer G and Casarotto, Silvia and Ferrarelli, Fabio and Baselli, Giuseppe and Mariotti, Maurizio and Massimini, Marcello and Rosanova,

Mario. Time-frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert-Huang transform. *Journal of Neuroscience Methods,* vol. 192, no. 2, pp. 236-245, 2011.
[21] Huang, Norden E. Review of empirical mode decomposition. In Proceedings of SPIE - The International Society for Optical Engineering, Orlando, FL, United States, March 26; *SPIE: Bellingham WA, United States*, 2001.
[22] Bao, Fei and Wang, Xinlong and Tao, Zhiyong and Wang, Qingfu and Du, Shuanping. "EMD-based extraction of modulated cavitation noise." *Mechanical Systems and Signal Processing,* vol. 24, no. 7, pp. 2124-2136, 2010.
[23] Wu, Zhaohua and Huang, Norden E. Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1-41, 2009.
[24] Yu, Dejie and Cheng, Junsheng and Yang, Yu. Application of Improved Hilbert-Huang Transform Method in Gear Fault Diagnosis." *Journal of Aerospace Power*, vol. 41, no. 6, pp. 1899-1903, 2009.

Xiaojun Zhu was born in Jiangsu, China, in 1977. He received the Master degree in Computer Science in 2001, from Taiyuan University of Technology, Taiyuan, China. He received the Doctor's degree in 2012, from Taiyuan University of Technology, Taiyuan, China. His research interests include Intelligent Information processing, cloud computing, and audio-visual computing.

Jingxian Hu is currently a Graduate student and working towards his M.S. degree at Taiyuan University of Technology, China. Her current research interest includes wireless sensor network and Intelligent Information processing.

Xiao Ma is currently a Graduate student and working towards his M.S. degree at Taiyuan University of Technology, China. Hercurrent research interest includes cloud computing, and audiovisual computing.