

Locating the Sources for Cross-modal Interactions and Decision Making during Judging the Visual-affected Auditory Intensity Change

Xuan Li, Xiaoli Ge, Junfeng Sun, *Member, IEEE* and Shanbao Tong^{*}, *Senior Member, IEEE*

Abstract—Audiovisual interaction has been one of the most important topics in cognitive neurosciences. Visual stimuli could significantly impact the auditory perception, and vice versa. Nevertheless, how much the change in visual stimuli would influence the perception of auditory change remains to be investigated. In this paper, we designed an audiovisual experiment in which subjects were required to judge whether there is a change in the intensities of two sounds with 150 ms interval, while there are two simultaneously presented size-changed visual stimuli. Behavioral results demonstrated that incongruent audiovisual change could result in the illusory perception of the change in sound intensity. For the correctly judged trials, source analysis showed two characteristic windows post the first auditory stimulus, i.e., (i) the 160-200 ms window including the auditory P200 and visual N100 wave, which was related to audiovisual interaction and working memory of the first stimulus with localized sources in insula and agranular retrolimbic area; and (ii) the 300-400 ms window for P300 with sources in premotor cortex and caudate nucleus, which were related to later audiovisual interaction, change discrimination and working memory. These preliminary results implied two stages in the audiovisual change perception task, with the involvement of insula, agranular retrolimbic, premotor cortex and caudate nucleus.

I. INTRODUCTION

HUMAN beings get information from the outside world through different modalities of sensory [1], which would usually integrate and/or interact in the central neural systems. As two most important sensory systems, the visual and auditory neural pathways have been the most investigated topics in cognitive neurosciences [2], [3]. In many cases, the audiovisual integration could enhance the perceptions [4], [5]. Nevertheless, when visual and auditory inputs are incongruent, the cognitive process would be more complicated, for example, (i) an auditory input could be altered by an incongruent visual stimulus presented simultaneously as in McGurk effect [7], or mis-located to an visual stimulus as in ventriloquism effect [6], (ii) the perception of the motion, rate, or temporal resolution of a visual stimulus could also be influenced by accompanying incongruent sound signals [7], [8], [9].

Funding for this study was provided by the Key Project of Chinese Ministry of Education (No.109059), National Natural Science Foundation of China (No.60901025).

X. Li, J. Sun, and S. Tong (e-mail: stong@sytu.edu.cn) are with the Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200030, China.

X. Ge is with the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

^{*} To whom correspondence should be addressed.

So far, most audiovisual interaction studies have been mainly based on a single visual and/or a single auditory stimulus. It has been showed that after an early audiovisual interaction at about 55 ms post-stimulus, later audiovisual interactions involve a significant cognitive process at nearly 200 ms [10]. However, although cognitive process for the perception of the auditory change has been investigated [11], [12], there are few reports about the cross-modal interactions on the perception of changes in visual or auditory stimuli. For example, how much would the visual change influence the perception of the change in simultaneously presented auditory stimuli? And one more interesting question is which areas in the brain are involved in the perception and judgment of the audiovisual changes?

In order to answer these questions, we designed an audiovisual cognitive experiment in which auditory and visual stimuli are incongruent and subjects were required to determine whether there's an intensity change in two successive sounds, with or without changed visual stimuli simultaneously presented. Multichannel event-related potentials (ERP) were recorded to investigate the underlying cognitive process. By running t-test of ERP difference wave analysis, we identify the temporal windows corresponding to the audiovisual interactions in judging the auditory intensity change. Further, source analysis of ERP difference wave within these windows by dipole localization method (DLM) will be used to locate the intracranial generators of brain activities for such cross-modal interactions.

II. MATERIALS AND METHODS

A. Subjects

Twelve volunteers from Shanghai Jiao Tong University (right-handed; age: 19-24 years; 5 males) were paid to participate in the experiment after they gave written informed consents. All subjects reported normal hearing and normal or corrected-to-normal vision. None of them had a history of neurological or psychological diseases.

B. Experimental procedures and data acquisition

Different from previous audiovisual experiments with a single visual and/or a single auditory stimulus, we study the cognitive process for the perception of the change in auditory stimuli accompanying with visual stimuli. Each trial consists of two visual stimuli, or two auditory stimuli, or two audiovisual stimuli with 150 ms interval. The visual stimulus is a 17 ms small (2°, V1) or large (4°, V2) circle filled in

white color in a black background, and the auditory stimulus is a 17 ms, 3.5 kHz sound at 65 dB (A).

Totally, we had three different stimulation trials, i.e., V1-V2, A-A, and V1A-V2A with equal probabilities. In order to get the task-related activity, we included blank trials in each block. Each experiment included 10 blocks. In each block, 42 trials (V1-V2: 12; A-A: 12; V1A-V2A: 12, blank: 6) were presented randomly. The stimuli and protocols are illustrated in Fig. 1.

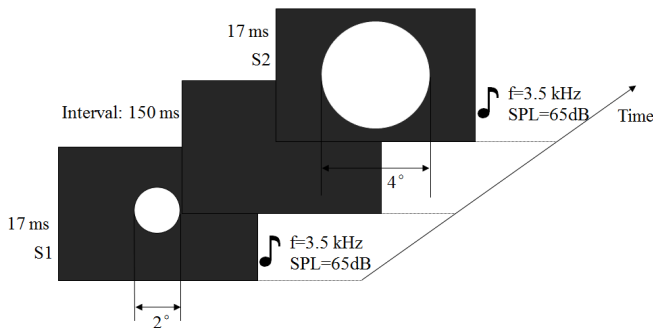


Fig. 1. Experimental stimuli and protocols with V1A-V2A as an example. The visual stimuli were (V1: 2°, V2: 4°), and the auditory stimuli in S1 and S2 were 3.5 kHz sinusoidal tones (A: 65 dB) with the same duration of 17 ms. The interval was 150 ms. Visual stimuli were presented in synchronization with the auditory stimuli.

In all trials, visual stimuli were presented on a 32-inch LCD display (model: KLV-32J400A, SONY, Shanghai) 120 cm in front of the subjects while auditory stimuli were from the two speakers (Model: R1600T08; Edifier; China) at both sides of the display. Subjects were asked to keep fixation at the center of the display and pay attention to both visual and auditory inputs. The task in each trial was to judge whether the intensities of the two auditory stimuli were the same or not. If the subject perceived the 2nd sound was at different intensity, then s/he need click the right button of the mouse, otherwise, click the left button instead. Thirty two channels of EEG signals were continuously recorded throughout the experiment (BrainAmp; Brain Products; Germany). Both auditory and visual stimuli were delivered with the software E-Prime (version 2.0; Psychology Software Tools, Inc, Pittsburgh, PA) which also recorded the response time (RT) and accuracy (ACC) automatically.

C. Source analysis

The EEG signals were offline pre-processed by Brain Vision Analyzer Software (version 2.0; Brain Products GmbH, Germany) first, then the trials for the same stimuli were grand-averaged to obtain the ERP signals. In order to study the activation during audiovisual interaction, we obtain the difference wave by subtracting the ERPs for A-A, V1-V2 from the cross-modal ERP and resting state (i.e., (V1A-V2A) + (No-Stimulus) - (A-A) - (V1-V2)). A running t-test comparing the amplitudes of ERP difference wave with baseline during blank trials was conducted to determine the time windows that were related to audiovisual interactions at

early stage of cognition and later change discrimination and decision making. Then we used DLM in BESA (version 5.2.2, MEGIS Software GmbH, Munich, Germany) to locate the brain regions that were responsible for these cognitive processes. In order to use DLM, we need first decide the number of dipoles responsible for these activations and then choose a volume conductor model to fit the dipoles with Downhill-Simplex optimization algorithm. In this study, the number of the dipoles was determined by the minimal number of the principal components in ERP difference wave that accounted for more than 90% of power, and the residual variance after fitting should be less than 15%. A four-shell (brain, cerebrospinal fluid, bone and skin) spherical volume conductor model (radius=90 mm) was chosen to locate the dipoles. The scalp, skull and cerebrospinal fluid thickness were 6 mm, 7 mm and 1 mm, respectively.

III. RESULTS

A. Behavioral results

The accuracy of judging the auditory intensity change for the trials of V1A-V2A was significantly lower than that of A-A (0.84 ± 0.12 vs. 0.92 ± 0.09 ; $p < 0.007$), while RT for V1A-V2A was significantly longer than that under auditory stimuli (A-A) only (629 ± 77 ms vs. 562 ± 56 ms; $p = 0.000$), which showed that incongruent visual change did influence the judgment of the intensity change in the simultaneously presented sounds.

B. ERP results

As shown in the ERP waves (Fig.2), the amplitudes for different stimuli varied in several time windows. To further determine the temporal characteristics of audiovisual change perception, we did running t-test comparing the amplitude of ERP difference wave with baseline (blank trial) at each sample and found two post-stimulus windows with significant difference as shown Fig.2, i.e., (i) W1: the 1st window of 160-200 ms which is overlapping the window 150-195 ms related to the early audiovisual interaction [10]; (ii) W2: the 2nd window of 300-400 ms which is corresponding to the P3 component contributed by the late process of the first stimuli and early cross-modal interactions of the second ones.

C. Source analysis

In both activated time windows, we did source analysis of the ERP difference wave to locate the activated brain areas.

In the window of 160-200 ms, principal component analysis (PCA) resulted in two components accounting for 98.9% of power for the ERP difference wave. DLM reliably found two dipoles: one was located in agranular retrolimbic area (Talairach coordinates: $x = 15$, $y = -33$, $z = -7$; \approx Brodmann area 30), shown as the red dipole in Fig. 3 (a), and the other was in the insula (Talairach coordinates: $x = 38$, $y = -15$, $z = 15$; Brodmann area 13), shown as the blue dipole in Fig.3 (a).

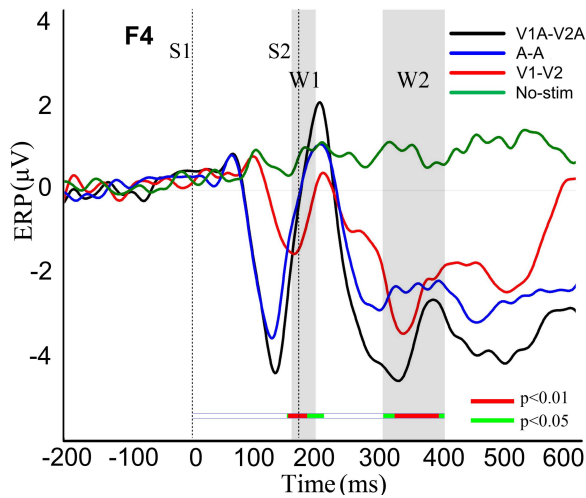


Fig. 2. ERP results for the F4 channel averaged by twelve subjects. The grey bars showed the two activated windows W1 and W2 according to ERP difference wave across all channel by running t-test. The horizontal color bars indicated the significant change in ERP difference wave by running t-test (red, $p < 0.01$; green, $p < 0.05$).

After fitting these two dipoles, the residual variance of the ERP difference wave was 7.1%.

In the window of 300–400 ms post the first stimuli, PCA also suggested two primary components which accounted for 97.4% of power in ERP difference wave. The first reliable dipole found by DLM was caudate nucleus (red dipole in Fig. 3 (b); Talairach coordinates: $x=9, y=6, z=16$). However, the second dipole is a little unstable by DLM. After 50 times of fitting, the second dipole was located either in the premotor cortex (72%; blue dipole in Fig. 3 (b); Talairach coordinates: $x=2, y=35, z=35$; \approx Brodmann area 6) or in insula (28%; Talairach coordinates: $x=44, y=-23, z=19$; Brodmann area 13). The corresponding residual variances of two freely fitted dipoles by DLM were respectively 13.9% and 14.1%. According to the possibility of occurrence and the residual variance, it is likely that the premotor cortex rather than insula was the activated region during 300–400 ms.

(a) + 160-200 ms

(b) + 300-400 ms

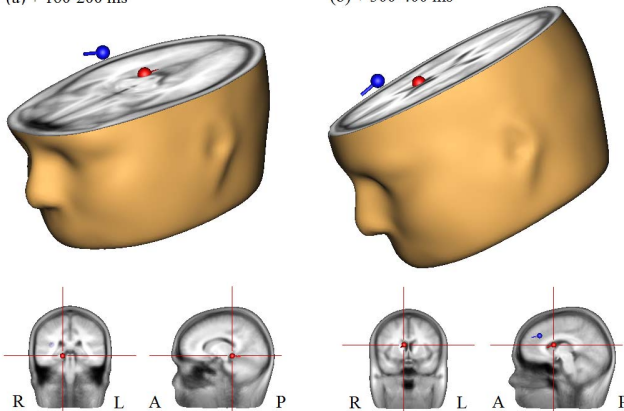


Fig. 3. Source analysis results according to ERP difference wave averaged by twelve subjects. (a) DLM found two dipoles at the time window 160–200 ms, the red dipole is located in the agranular retrolimbic area (Brodmann area 30) while the blue dipole is found in insula (Brodmann area 13). (b) DLM found two dipoles at the time window 300–400 ms, the red and blue dipoles are separately located in caudate nucleus and premotor cortex (Brodmann area 6).

Thus, the behavior results confirmed that the size-changed visual stimuli did result in illusory perception of the change in the simultaneously presented sounds and may increase the RT even the subjects made correct judgments. While the electrophysiological results showed distributed brain regions involved in the whole cognitive process. The influence of incongruent visual stimuli on perception of the change in auditory intensity involves the activations in insula and the agranular retrolimbic area, whereas the late discrimination and working memory were also affected by such audiovisual incongruency, with activation in premotor cortex and caudate nucleus.

IV. DISCUSSION

A. 160-200 ms post stimulus

In the 160–200 ms time window, insula (Brodmann area 13) and agranular retrolimbic area (Brodmann area 30) were activated during the audiovisual interaction and related working memory for the first stimuli in our experiment. Insula was thought to be closely linked to attention and emotion. It was also thought to take part in several auditory processes, e.g., allocating auditory attention and integrating audiovisual information [13], especially when the auditory and visual stimuli were presented in temporal synchrony [10]. Thus, we thought insula might take part in the audiovisual interaction for the first stimuli in this discrimination experiment. A recent study on motion perception also found the bilateral insula activation in the visual-induced illusion [14]. The involvement of insula in our experiment on incongruent audiovisual task also confirmed the role of insula in the visual-induced auditory illusion.

The activated Brodmann area 30 within this window might be related to working memory [15]. Previous studies on working memory showed its latency at about 300 ms post stimulus [16]. Such an earlier activation of Brodmann area 30 in our experiment might be due to the involvement of the agranular retrolimbic area in starting the working memory of the first stimulus.

B. 300-400 ms post stimulus

The 300–400 ms window is thought to be related to the audiovisual interaction of the second stimuli and the auditory change discrimination.

The source analysis could reliably locate the first dipole at caudate nucleus, which is always involved in working memory [17]. Previous study demonstrated that caudate nucleus also took part in the audiovisual motion perception [18]. Therefore, the activation of caudate nucleus implied its involvement in the change discrimination task, including the audiovisual interaction of the second audiovisual stimuli and the working memory in comparing S1 and S2. The second dipole had two possible locations, i.e., premotor cortex (possibility 72%) and insula (possibility 28%) according to DLM, which might be due to the low signal-noise ratio of ERP difference wave in the window.

Nevertheless, Bordmann area 6 was located as the source in most times (36/50 times), which implied it was more likely to be the second dipole for discriminating the sound intensity change. Since this area is related to planning, its activation was likely to be related to the decision making in this task. The finding is close to previous study that the 280-430 ms window was responsible for decision making by using S1-S2 paradigm [12].

Based on these source analyses, we may infer that the agranular retrolimbic area and insula are involved in the audiovisual interaction and working memory during the cognition of the first stimuli in our experiment. While after the onset of the second stimuli, the discrimination of auditory change is mediated by the premotor cortex. Meanwhile, caudate nucleus may participate in the audiovisual interaction for the second stimulus and the working memory in judging the auditory change.

It should be noted that DLM has been widely used in event-related potential (ERP) experiments [19] to locate the sources of brain activities, with great potential clinical applications. However, the results of source analysis by DLM are sensitive to many factors (e.g., the volume conductor model and the signal-noise ratio in ERP signals) and may lead to different outputs. Therefore, further neuroimage methods are needed to confirm and verify the findings in this study.

V. CONCLUSION

In summary, we designed a sound intensity change discrimination task to study the audiovisual interactions under incongruent stimuli, which is different from previous work with a single auditory and/or a single visual stimulus. Both behavioral and neurophysiological data showed the disturbance of simultaneously presented visual stimuli on the discrimination of the intensity change in sounds. Source analysis further indicated that insula and agranular retrolimbic area were involved in audiovisual interaction and starting related memory after the onset of the first stimuli. While after the onset of the second stimuli, premotor cortex played an important role in the discrimination of the auditory intensity change and caudate nucleus participated in the audiovisual interaction and working memory.

REFERENCES

- [1] M.O. Ernst and H.H. Bühlhoff, "Merging the senses into a robust percept," *TRENDS in cognitive Sciences*, vol. 8, no. 4, pp. 162-169, 2004.
- [2] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, and J.J. Foxe, "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.
- [3] S. Shimojo and L. Shams, "Sensory modalities are not separate modalities: plasticity and interactions," *Current opinion in neurobiology*, vol. 11, no. 4, pp. 505-509, 2001.
- [4] D.A. Bulkin and J.M. Groh, "Seeing sounds: visual and auditory interactions in the brain," *Current opinion in neurobiology*, vol. 16, no. 4, pp. 415-419, 2006.
- [5] J. Navarra and S. Soto-Faraco, "Hearing lips in a second language: visual articulatory information enables the perception of second language sounds," *Psychological research*, vol. 71, no. 1, pp. 4-12, 2007.
- [6] T.M. Woods and G.H. Recanzone, "Visually induced plasticity of auditory spatial perception in macaques," *Current Biology*, vol. 14, no. 17, pp. 1559-1564, 2004.
- [7] R. Sekuler, A.B. Sekuler, and R. Lau, "Sound alters visual motion perception," *Nature*, vol. 385, no. 6614, pp. 308-308, 1997.
- [8] L. Shams, Y. Kamitani, and S. Shimojo, "Visual illusion induced by sound," *Cognitive Brain Research*, vol. 14, no. 1, pp. 147-152, 2002.
- [9] C. Scheier, R. Nijhawan, and S. Shimojo, "Sound alters visual temporal resolution," *Investigative Ophthalmology & Visual Science*, vol. 40, no. Suppl 4, pp. 4169, 1999.
- [10] J. Vidal, M.H. Giard, S. Roux, C. Barthelemy, and N. Bruneau, "Cross-modal processing of auditory-visual stimuli in a no-task paradigm: A topographic event-related potential study," *Clinical Neurophysiology*, vol. 119, no. 4, pp. 763-771, 2008.
- [11] C. Alain, S.R. Arnott, S. Hevenor, S. Graham, and C.L. Grady, "What and "where" in the human auditory system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 21, pp. 12301, 2001.
- [12] J. Kaiser, T. Lennert, and W. Lutzenberger, "Dynamics of oscillatory activity during auditory decision making," *Cerebral Cortex*, vol. 17, no. 10, pp. 2258, 2007.
- [13] D.E. Bamiou, F.E. Musiek, and L.M. Luxon, "The insula (Island of Reil) and its role in auditory processing: literature review," *Brain research reviews*, vol. 42, no. 2, pp. 143-154, 2003.
- [14] A. Alink, W. Singer, and L. Muckli, "Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex," *Journal of Neuroscience*, vol. 28, no. 11, pp. 2690, 2008.
- [15] L. Davachi, J.P. Mitchell, and A.D. Wagner, "Multiple routes to memory: distinct medial temporal lobe processes build item and source memories," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 4, pp. 2157, 2003.
- [16] L. Nielsen-Bohlman and R.T. Knight, "Electrophysiological dissociation of rapid memory mechanisms in humans," *NeuroReport*, vol. 5, no. 12, pp. 1517, 1994.
- [17] R. Levy, H.R. Friedman, L. Davachi, and P.S. Goldman-Rakic, "Differential activation of the caudate nucleus in primates performing spatial and nonspatial working memory tasks," *Journal of Neuroscience*, vol. 17, no. 10, pp. 3870, 1997.
- [18] O. Baumann and M.W. Greenlee, "Neural correlates of coherent audiovisual motion perception," *Cerebral Cortex*, vol. 17, no. 6, pp. 1433, 2007.
- [19] J. Dien, K.M. Spencer, and E. Donchin, "Localization of the event-related potential novelty response as defined by principal components analysis," *Cognitive Brain Research*, vol. 17, no. 3, pp. 637-650, 2003.