Proceedings of 2016 IEEE
International Conference on Mechatronics and Automation
August 7 - 10, Harbin, China

Study of Audiovisual Asynchrony Signal Processing: Robot Recognition System of Different Ages

Yanna Ren

The Graduate School of Natural Science and Technology
Okayama University
3-1-1 Tsushima-naka, Okayama, 700-8530, Japan
pejj003a@s.okayama-u.ac.jp

Weiping Yang

Department of Psychology, Faculty of Education ,
Hubei University
No.11, Friendship Street, Wuchang District, Hubei, 430062, China
swywp@163.com

Qiong Wu, Fengxia Wu, Satoshi Takahashi, Ejima Yoshimichi, Jinglong Wu

The Graduate School of Natural Science and Technology
Okayama University
3-1-1 Tsushima-naka, Okayama, 700-8530, Japan
wu@mech.okayama-u.ac.jp

***Abstract* - The current studies indicated that sound can improve the visual perception. However, whether the gain is influenced by temporal disparity between visual and auditory stimuli, and whether the gain is the same for older and younger adults remains unclear. To clarify this mechanism, ten older adults and 10 younger adults were invited to confirm and visual/auditory discrimination task. The visual target task, visual irrelevant task, auditory target task, and auditory irrelevant task randomly presented on the right or left hemi-space of the central fixation point and the subjects were instructed to respond to target stimuli rapidly and accurately. The results showed that in the younger adults, the gain was greatest in the simultaneous audiovisual condition. With the enlargement of temporal disparity between visual stimulus and auditory stimulus, the gain decreased significantly, and the alterative slope was different between auditory proceeded visual stimulus and visual proceeded auditory stimulus. In the older adults, the variation tendency was similar to younger adults, but in all conditions, the response of older adults was significantly slower than that of younger adults. Our results suggested that the visual gains were influenced by temporal disparity between auditory and visual stimuli, and further suggested that the gain mechanism between auditory proceeded visual stimuli and visual proceeded auditory stimuli was different, which indicated that the development of intelligent robot should consider the condition of the target subject to switch to the adapt procedure.**

Index Terms - Visual Perception, Response Enhancement, Audiovisual Synchrony, Older Adults, Immersive Perception

I. INTRODUCTION

Artificial intelligence (AI) is arguably the most exciting field in robotics. However, the intelligent of robot is limited, and all its action is basing on the stored programs. The computer compares this information to stored data and decides what the information signifies. The computer runs through various possible actions and predicts which action will be most successful based on the collected information. Therefore, the stored data is the key to realize the level of intelligent (Figure 1). In order to simulate the procedure of human being, it is necessary to clarify how the brain of human being processes the multiple information signals from the environment.

Auditory information and visual information is the basic factors of various received information during communicating with human being. It is important to clarify the mechanism of audiovisual interaction for the development of intelligent robot technology. Audition is important in orienting. In the real life, we mainly depend upon sound localization for orientation toward significant distal events that occur outside the field of view, during vision occlusion or darkness. Recent studies have investigated the possibility that sound also improves visual perception[1]. The mechanism for the enhancement of visual processing by an auditory input has been extensively investigated in animal work. This has demonstrated the existence of special rules underlying this integration at the cellular level. However, few studies have investigated in humans the behavioral effects of these rules.

The previous studies include the finding that two visual stimuli moving toward each other are seems to “bounce” if an auditory stimulus is presented simultaneous to the contact of the two visual stimuli[2]. Besides, presence of multiple auditory stimuli increased the number of perceived visual stimuli[3]. Gregg H. Recanzone conducted designed four tasks, including single visual stimulus, bimodal change stimuli, target constant w/Distracter change, and target change W/Distracter constant, to investigate the auditory influences on visual temporal rate perception. The results showed that the auditory system has a pronounced influence on visual temporal rate perception, and further suggested that the influence was independent of the spatial location, spectral bandwidth, and intensity of the auditory stimulus, however, strongly dependent on the disparity in temporal rate between the two stimulus modalities[4]. Bolognini and his colleagues valued whether such temporal principle observed in animals also apply to human using modality-specific selective attention paradigm, and their results revealed that enhancement of visual response was found when auditory and visual stimuli presented simultaneously, while disappeared at larger SOA (from 100 ms to 500 ms)[1].

978-1-5090-2396-7/16/$31.00 ©2016 IEEE

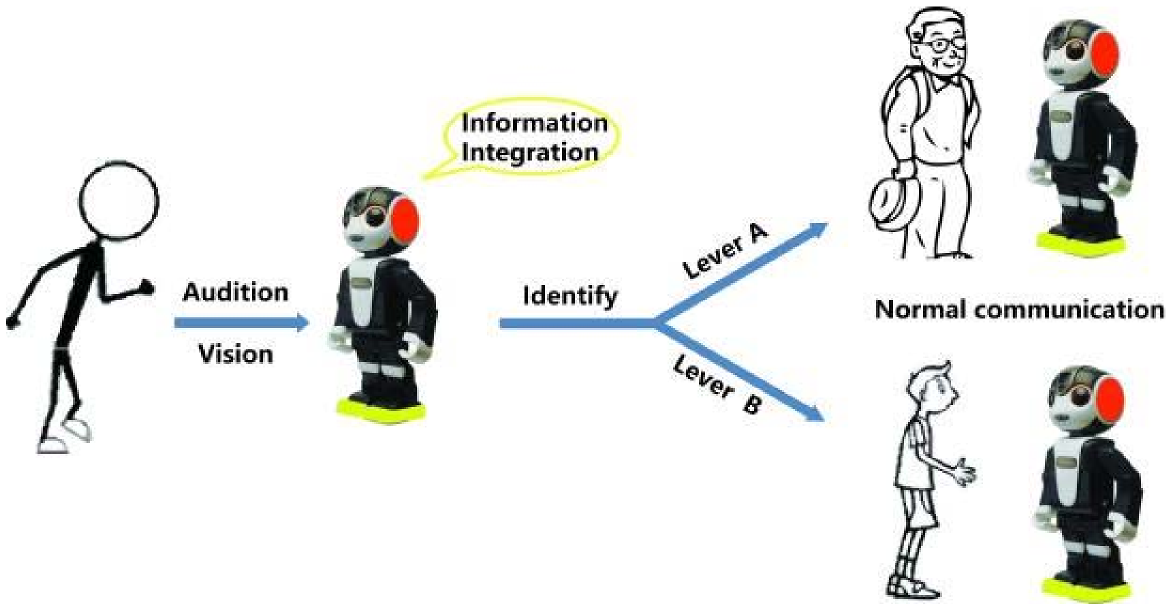

Figure 1 According to the auditory and visual information that the robot received from communication, he stars different stored procedure to serve the target subject much better.

Additionally, the temporal ventriloquism effect for audiovisual integration suggest that when the auditory stimulus and visual stimulus presented asynchronously, to combine the two signals together, the visual stimulus will be realignment again[5]. However, whether the mechanism of auditory stimulus proceeded visual stimulus and visual stimulus proceeded auditory stimulus is same or not remains unclear.

In order to clarify this mechanism, we designed an auditory/visual stimuli discrimination task. Two subject groups participated in our experiment, and we rescored the response time (RT), hit rate and false alarm for all tasks. By comparing visual enhancement in each condition, we assessed whether the gain is influenced by temporal disparity between visual and auditory stimuli, and whether the gain is the same for older and younger adults, and when the output of the auditory and visual information will be appropriate to the immersive perception.

II. METHODS

A Subjects

Ten healthy younger volunteers (21 – 24 years old, mean age 23.6 years) and 10 old volunteers (ages 60 - 75 years, mean age 65.4 years) participated in this study. All of the participants had normal or corrected-to-normal vision and normal hearing capabilities. Participants were excluded if their mini-mental state examination (MMSE) scores were greater than 2.5 standard deviations (SD) from the mean for their age and education [6]. Besides, participants who reported a history of cognitive disorder were also excluded. The participants provided written informed consent for their participation in this study, which was previously approved by the ethics committee of Okayama University.

B Stimuli and task

Stimulus presentation and response collection were accomplished using Presentation software (Neurobehavioral Systems Inc., Albany, California, USA). A stream comprised of unimodal visual stimuli, unimodal auditory stimuli and bimodal audiovisual stimuli were presented randomly either to the left or the right hemi-space. Each type of stimulus contained two subtypes of target and task-irrelevant stimuli.

The visual target stimulus was a checkerboard image with 2 dots contained within the checkerboard (5.2 × 5.2 cm, subtending a visual angle of 5-degrees), and task-irrelevant visual stimulus was the checkerboard. All the visual tasks presented on a black background on a 21-inch computer monitor positioned 60 cm in front of the participant's eyes (Fig 2A). These visual stimuli were randomly presented to the lower left or the lower right quadrant of the screen (at a 12-degree visual angle to the left or the right of the center and a 5-degree angle below the central fixation) [7, 8]. The auditory target stimulus was a white noise and task-irrelevant auditory stimulus was a 1 000 Hz sinusoidal tone with linear rise and fall times of 5 ms with amplitude of 60 dB. All of the auditory stimuli were presented to either the left or the right ear through earphones. Bimodal audiovisual target stimuli included a visual stimulus and an auditory stimulus. The auditory stimuli either preceded or followed the vision's onset by stimulus onset asynchronies drawn randomly. Additionally, task-irrelevant stimuli composed 20% of the total stimuli; these stimuli were similar to the visual and auditory target stimuli. The task-irrelevant audiovisual stimulus was composed of the task-irrelevant visual and auditory stimuli. The duration of each component of each stimulus was 150 ms. The inter stimulus interval varied randomly between 1 300 and 1 800 ms. Participants performed five experimental blocks, and each block lasted approximately 8 mins, and begin with a fixation cross for 3 000 ms. In each blocks, 30 unimodal visual stimuli, 30 unimodal auditory stimuli and 210 audiovisual stimuli (30 AV, 30 V50A, 30 V100A, 30V150A, 30A50V, 30 A100V, 30 A150V) were presented at a random with a randomly varied inter-stimulus interval (ISI) between 1 300 and 1 800 ms. Each subject completed 1350 trails (1080 target trails and 270 task-irrelevant trails), and was instructed to identify whether the targets appeared on the left hemi-space (press the left button of the mouse) or the right hemi-space (press the right button of the mouse) as rapidly and accurately as possible.

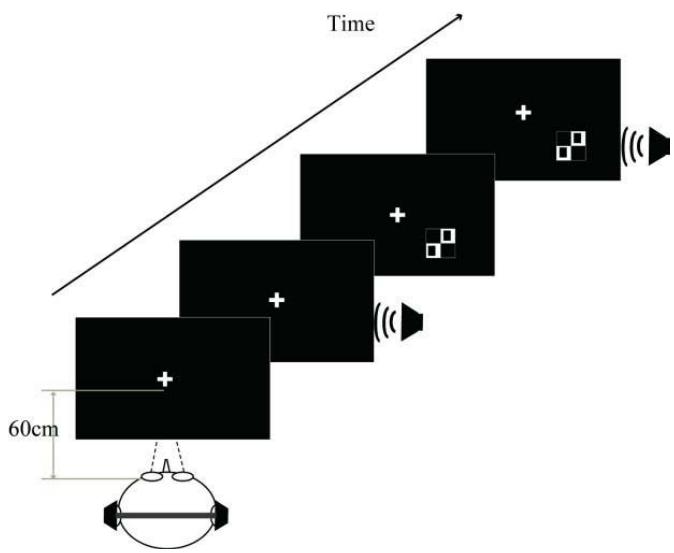

Figure 2. Go/No-Go task. As an example of the task sequence, the task was beginning with a fixation cross for 3 000 ms, and then the auditory target stimulus was present (white noise tone 1 000Hz, 60 dB) for 150 ms. Following the target auditory stimulus presentation, the auditory stimulus is disappeared, and an inter-stimulus interval (ISI: 1 300-1 800ms) was provided for the subject to describe which sides the stimulus came from (right or left). After discriminating, irrelevant visual stimulus was presented.

C Analysis

Hit rates and response times of targets were computed separately for each stimulus type. Hit rate is percentage of correct respond (the response time fall into the period of average time ± 2.5 SD) to the total target stimuli, and response time data were analyzed for correct responses. Hit rates and response times were subjected to repeated-measures analyses of variance (ANOVAs).

To evaluate the visual enhancement of the multisensory condition, response enhancement proportionality was used to analyze mean response time (RT). Visual enhancement = $(TR_V - RT_{AV})/RT_V*100\%$[9, 10]. TRV is the mean response time to visual-only condition, and RTAV is the mean response time to all bimodal conditions (AV, A50V, A100V, V50A, V100A).

III. RESULTS

A Hit rate (HR)

All hit rate was calculated, and the data was submitting to ANOVA to analysis. In addition to the response of old adults to visual stimulus, for all other conditions of both younger and elderly groups, mean performance hit rates were greater than 94%. 2 (two groups) * 9 (conditions) repeat-measures ANOVA was performed on the hit rate. The results showed that there was no significant main effect of group [$F(1, 18) = 0.062$, $P = 0.807$], however, there are significant main effect of stimuli conditions [$F(8, 144) = 9.616$, $P < 0.001$], and interaction between group and condition [$F(8, 144) = 0.02$, $P = 0.042$]. Additionally, within each age group, further post-hoc comparisons found that there is significant different between condition V and all other bimodal condition AV ($P < 0.05$), but have no significant difference with condition A ($P > 0.05$) (Figure 3).

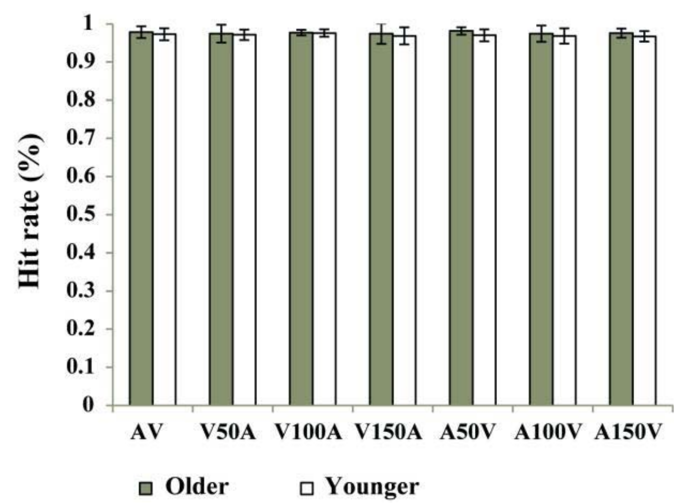

Figure 3 Hit rates for all unimodal and bimodal conditions. The gray column is the hit rate for the older adults, and the white column is the hit rate for the younger adults.

B Mean response time (RT)

All the response time were also calculated and submit to ANOWA analysis. 2 (two groups) * 9 (conditions) repeat-measures ANOVA was performed on the response time. The results showed that there is significant main effect for both condition [$F(8, 144) =39.296$, $P < 0.001$] and group [$F(8, 144) = 23.581$, $P < 0.001$], which exhibiting faster response speed when auditory stimuli and visual stimuli were presented synchronously than asynchronously or independently, and a faster response speed for younger adults in all conditions. However, there is no significant interaction between group and condition [F (1, 18) = 1.241, P = 0.304]. Additionally, within each age group, further post-hoc comparisons found that when temporal disparity was the same, regardless of whether the visual or auditory stimulus appeared first (e.g.: A50V and V50A), there were no significant difference between different temporal order of A and V ($p > 0.05$) and that the response time for older adults under each condition was significantly longer than that for younger adults ($p \leq 0.001$) (Figure 4).

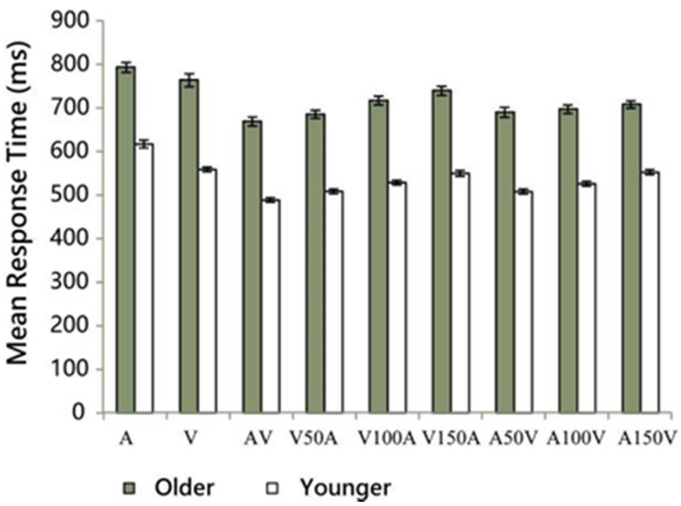

Figure 4 Mean response time for all unimodal and bimodal conditions. The gray column is the hit rate for the older adults, and the white column is the hit rate for the younger adults.

C Visual response enhancement

To assess the facilitation effect of auditory stimulus to visual stimulus, we calculated the visual response enhancement. 2 (two groups) *7 (conditions) repeat-measures ANOVA was performed on the response enhancement. The results showed that there was a significant main effect of condition [F (6, 108) = 35.065, P < 0.001], however, there are no significant main effect of group [F (1, 18) = 0.191, P = 0.668] and interaction between group and condition [F (6, 108) = 2.748, P = 0.08] (Figure 5). Additionally, within each age group, further post-hoc comparisons found that in both older and younger groups, the response enhancement decreased greatly with enlargement of temporal disparity regardless of in visual proceeded auditory stimuli or auditory proceeded visual stimuli. However, the variation was greater in auditory proceeded visual condition than visual proceeded auditory conditions with the expansion of temporal disparity (Table 1).

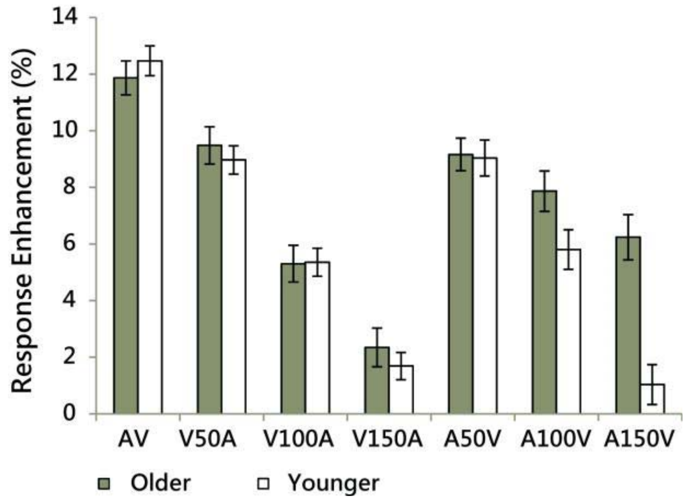

Figure 5 The visual response enhancement for bimodal conditions. The gray column is the hit rate for the older adults, and the white column is the hit rate for the younger adults.

Table 1 The visual response enhancement in all bimodal conditions

	Older		Younger	
	VRE(%)	SE(%)	VRE(%)	SE(%)
AV	11.86	1.19	12.47	0.53
V50A	9.48	0.95	8.97	0.50
V100A	5.30	0.53	5.35	0.50
V150A	2.35	0.23	1.69	0.48
A50V	9.16	0.92	9.03	0.63
A100V	7.87	0.79	5.80	0.70
A150V	6.24	0.62	1.04	0.70

IV. DISCUSSIONS

The research showed that the response of older adults in all condition was slower than that of younger adults. Comparing with the younger adults, the older adults have a much higher threshold to percept auditory stimulus and visual stimulus [11, 12]. and a cognitive functional decline[13]. Therefore, it is reasonable that the response time was slower in elderly adults. Besides, several investigations indicated that attentional cognitive function was declined with aging[14-16], and both of selective attention and divided attention can modulate auditory and visual perception greatly[17, 18]. Therefore, the most possible reasons that older adults showed a slower response may be result from lower threshold to auditory or visual stimuli[11, 13, 19, 20], a cognitive functional decline [13-16], or decline of attention ability[17, 18] for older adults.

The response enhancement decreased greatly with enlargement of temporal disparity regardless of in visual proceeded auditory stimuli or auditory proceeded visual stimuli. This result was constant with the previous study[1], which showed that when auditory stimulus presented with visual stimulus simultaneously, the visual detection sensitively increased greatly comparing with visual-only condition. However, when the temporal disparity was 100 ms, 200 ms, 300ms, 400 ms, and 500 ms, there was no significant increase for the visual stimulus detection. However, the variation was greater in auditory proceeded visual condition than visual proceeded auditory conditions with the expansion of temporal disparity. This result suggested that the neuronal mechanisms of visual response enhancement by auditory inputs are different between visual proceeded auditory condition and auditory proceeded visual condition. In visual proceeded auditory condition, it is maybe mainly result from the cue-target facilitation processing theory, which is explained by that the exogenous shift of attention in auditory modality leads to a corresponding shift of attention in visual modality[21]. At this condition, as an alerting signal, the auditory stimulus can active our brain before the visual target arrival. While in the auditory proceeded visual stimulus, the response enhance effect maybe mainly depended on the audiovisual integration effect. Studies on multisensory audiovisual integration revealed that response to audiovisual stimuli were faster and more accuracy than to unimodal auditory or visual stimulus [1, 22]. However, further studies are need to uncover the neuronal mechanism.

V. CONCLUSIONS

The current study confirmed that the visual gains was influenced by temporal disparity between auditory and visual stimuli, and further suggested that the gain mechanism between auditory proceeded visual stimuli and visual proceeded auditory stimuli was different. Therefore, for the development of intelligent robot, we much consider the condition of the target subject to set up different procedure model.

VI. LIMITATION OF THE CURRENT STUDY

The limitations of this study are the followings: firstly, only temporal disparity 0 ms and ±150 ms were studied, and the performed conditions were not enough to investigate the effect of temporal disparity on visual enhancement systematically. Secondly, there were only 10 older adults and 10 younger adults participant in the current research, and for the future study, the number of the participants should be increased. Thirdly, in the present study, we just conducted the behavioral experiment. The neuroimaging will be much better to uncover the neuronal mechanism. Basing on these results, in the further studies, we would choose more conditions from 0 ms - ±100 ms and ±100 ms - ±500 ms. Furthermore, it might be explained using simultaneous brain imaging and electrophysiological evidence.

ACKNOWLEDGE

This study was supported by JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE (JSPS) KAKENHI grant numbers 25249026, 25303013 and 16K18052, a Grant-in-Sid for Strategic Research Promotion from Okayama University, and the National Science Foundation of Hubei University (098379, WY).

REFERENCE

[1] N. Bolognini, F. Frassinetti, A. Serino, and E. Ladavas, ""Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs," *Experimental Brain Research,* vol. 160, pp. 273-282, Jan 2005.

[2] K. Watanabe and S. Shimojo, "When sound affects vision: effects of auditory grouping on visual motion perception," *Psychological Science,* vol. 12, pp. 109-116, 2001.

[3] L. Shams, Y. Kamitani, and S. Shimojo, "Visual illusion induced by sound," *Cognitive Brain Research,* vol. 14, pp. 147-152, 2002.

[4] G. H. Recanzone, "Auditory influences on visual temporal rate perception," *Journal of neurophysiology,* vol. 89, pp. 1078-1093, 2003.

[5] C. Spence and S. Squire, "Multisensory integration: maintaining the perception of synchrony," *Current Biology,* vol. 13, pp. R519-R521, 2003.

[6] G. Bravo and R. Hebert, "Age- and education-specific reference values for the Mini-Mental and modified Mini-Mental State Examinations derived from a non-demented elderly population," *Int J Geriatr Psychiatry,* vol. 12, pp. 1008-18, Oct 1997.

[7] D. Talsma and M. G. Woldorff, "Selective Attention and Multisensory Integration: Multiple Phases of Effects on the Evoked Brain Activity," *Journal of Cognitive Neuroscience,* vol. 17, pp. 1098-1114, 2005.

[8] S. He, P. Cavanagh, and J. Intriligator, "Attentional resolution and the locus of visual awareness," *Nature,* vol. 383, pp. 334-337, 1996.

[9] N. Tye-Murray, M. Sommers, B. Spehar, J. Myerson, and S. Hale, "Aging, audiovisual integration, and the principle of inverse effectiveness," *Ear and hearing,* vol. 31, p. 636, 2010.

[10] R. A. Stevenson, D. Ghose, J. K. Fister, D. K. Sarko, N. A. Altieri, A. R. Nidiffer*, et al.*, "Identifying and quantifying multisensory integration: a tutorial review," *Brain topography,* vol. 27, pp. 707-730, 2014.

[11] X. Z. Liu and D. Yan, "Ageing and hearing loss," *J Pathol,* vol. 211, pp. 188-97, Jan 2007.

[12] P. D. Spear, "Neural bases of visual deficits during aging," *Vision Res,* vol. 33, pp. 2589-609, Dec 1993.

[13] K. Schomann, "The Handbook of Aging and Cognition, 3rd edition," *Zeitschrift Fur Erziehungswissenschaft,* vol. 12, pp. 573-577, Sep 2009.

[14] A. Kok, "Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (ERP)," *Biological Psychology,* vol. 54, pp. 107-143, Oct 2000.

[15] D. J. Plude, J. T. Enns, and D. Brodeur, "The development of selective attention: a life-span overview," *Acta Psychol (Amst),* vol. 86, pp. 227-72, Aug 1994.

[16] C. Quigley, S. K. Andersen, L. Schulze, M. Grunwald, and M. M. Muller, "Feature-selective attention: evidence for a decline in old age," *Neurosci Lett,* vol. 474, pp. 5-8, Apr 19 2010.

[17] J. L. Mozolic, C. E. Hugenschmidt, A. M. Peiffer, and P. J. Laurienti, "Modality-specific selective attention attenuates multisensory integration," *Exp Brain Res,* vol. 184, pp. 39-52, Jan

2008.
[18] R. A. Stevenson, J. K. Fister, Z. P. Barnett, A. R. Nidiffer, and M. T. Wallace, "Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance," *Experimental Brain Research,* vol. 219, pp. 121-137, May 2012.
[19] W. P. Yang, B. Q. Chu, J. J. Yang, Y. H. Yu, J. L. Wu, and S. Y. Yu, "Elevated audiovisual temporal interaction in patients with migraine without aura," *Journal of Headache and Pain,* vol. 15, Jun 24 2014.
[20] J. L. Wu, J. J. Yang, Y. H. Yu, Q. Li, N. Nakamura, Y. Shen, *et al.*, "Delayed Audiovisual Integration of Patients with Mild Cognitive Impairment and Alzheimer's Disease Compared with Normal Aged Controls," *Journal of Alzheimers Disease,* vol. 32, pp. 317-328, 2012.
[21] E. Macaluso, C. Frith, and J. Driver, "Multisensory integration and crossmodal attention effects in the human brain - Response," *Science,* vol. 292, Jun 8 2001.
[22] M. A. Meredith, J. W. Nemitz, and B. E. Stein, "Determinants of Multisensory Integration in Superior Colliculus Neurons .1. Temporal Factors," *Journal of Neuroscience,* vol. 7, pp. 3215-3229, Oct 1987.