ISCA Archive
http://www.isca-speech.org/archive

AVSP 2010 - International Conference
on Audio-Visual Speech Processing
Hakone, Kanagawa, Japan
September 30-October 3, 2010

An ERP examination of audiovisual speech perception in Japanese younger and older adults

Satoko Hisanaga [1], *Kaoru Sekiyama* [2], *Tomohiko Igasaki* [3] *and Nobuki Murayama* [3]

[1] Graduate School of Social and Cultural Sciences, Kumamoto University, Japan
[2] Faculty of Letters, Kumamoto University, Japan
[3] Graduate School of Science and Technology, Kumamoto University, Japan
[1] satoko.hisanaga@gmail.com, [2] sekiyama@kumamoto-u.ac.jp

Abstract

We studied differences between Japanese younger (YA) and older adults (OA) by recording event-related brain potentials (ERP). Participants were asked to identify audio only (AO) and congruent audiovisual (AV) syllables as /ba/ or /ga/). We found age-related ERP changes (N1, P2, and N2 latencies) in Japanese audiovisual speech perception. Whereas the visual influence was sustained (maintained from P2 to N2) in the OA group, the influence was transient (limited only to N1) in the YA group. It seemed that the OA group's slower auditory processing is compensated by visual information.
Index Terms: audiovisual speech perception, aging, ERP

1. Introduction

Information processing for bimodal congruent conditions is often faster than for unimodal conditions, for example, in speech and emotion [1] perception. In speech perception, congruent AV conditions result in shortened reaction times compared to the audio-only (AO) condition in [2-4]. While the bimodal acceleration in speech perception has been reported in native speakers of English, French, and Finish, our previous results showed that this is not the case for native speakers of Japanese who showed prolonged reaction times for the AV condition compared to the AO condition [5]. Although the bimodal slowing down in Japanese may look peculiar, it is in accordance with reported interlanguage differences: for example, Japanese speakers identify syllables faster in audio-only condition than in visual-only condition, whereas English speakers' reaction times were the same for the two conditions [6].

Recently, ERP has been applied to analyze various processes of speech perception in the brain. The previous studies on audiovisual speech perception indicated that N1 latency of the AV condition was shorter than tha of the AO condition [2,7,8], and in some cases, also in P2 latency [8]. Our previous study indicated that the visual influence, that is, the reduction in latency due to additional visual cues, was maintained from N1 to P2 in the native speaker of English (thus replicating the results by [8]), but it was limited only to N1 in the native speaker of Japanese [5].

Our present aim was to study age-related changes of ERPs in the visual influence in audiovisual speech perception for Japanese speakers. In aging, hearing gradually declines [9,10]. Thus, even if younger Japanese speakers do not use visual information as much as the English counterparts do, older Japanese would use visual information more than younger Japanese [11]. We examined this hypothesis by using ERP.

2. Experiment 1: Behavioral study

Experiment 1 examined simple reaction time and choice reaction time as behavioral data.

2.1. Simple reaction time experiment

2.1.1. Participants

Eleven YA (mean age 21.3 years, range 20-27, 8 females and 3 males) and 9 OA (mean age 66.2 years, range 60-73, 6 females and 3 males) participated in this experiment. All participants in this study were right handed, monolingual Japanese speakers, and had no documented hearing disorders or brain dysfunction. They all had normal or corrected to normal vision and normal hearing. Their hearing thresholds were less than 25dB HL (in average of thresholds for 1,000, 2,000, and 4,000 Hz). All participants gave an informed consent to participate in this study.

Audiometric thresholds are shown for pure-tone sensitivity in Figure 1. The OA group's hearing declined compared to the YA group ($p<.05$).

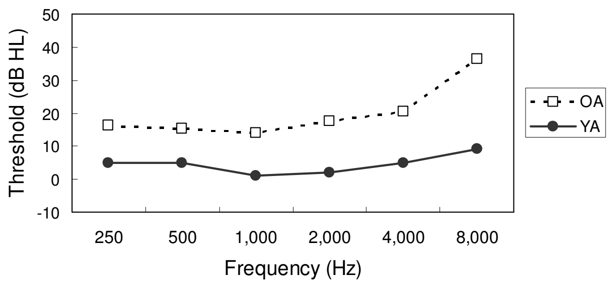


Figure 1: Mean audiometric thresholds for the right ear.

2.1.2. Stimuli

Stimuli were prepared for the auditory and visual conditions. In the auditory condition, the stimulus was white noise (auditory target). The duration of auditory stimulus was 50 ms. In the visual condition, the stimulus was a white dot (5 mm in radius) presented at the center of the display (visual target). The duration of visual stimulus was 100 ms.

After a fixation point (3 mm in radius) was randomly presented for 800 ms, 900 ms, 1,000 ms, 1,100 ms, and 1,200 ms (a total of 5 steps), auditory or visual target was introduced.

2.1.3. *Procedure*

The auditory stimulus was presented through a loudspeaker (AIWA SC-B10) at 65 dB SPL. The visual stimulus was presented at the center of a 15-inch SONY SDM-S51 monitor and the loudspeaker was placed above the monitor. Participants sat approximately 90 cm from the monitor. One experimental block consisted of 20 trials and there were two conditions (auditory and visual conditions), resulting in a total of 40 trials. The presentation order was counterbalanced across participants. The participants were instructed to look at the monitor, press the button as fast as possible when they detected each target. Reaction time was measured as the time from the target onset to the button press.

2.1.4. *Result and Discussion*

The simple reaction times were analyzed by an ANOVA [Between-subjects: age group (2) × Within-subjects: condition (2)]. The two-way ANOVA revealed a significant interaction of age group × condition [F (1, 18) = 5.09, p<.05]. The mean reaction time for the audio stimuli was significantly longer in the OA group compared with the YA group [F (1, 36) = 9.12, p<.005]. On the other hand, the reaction time of the visual condition did not differ between the YA and OA groups. In addition, the reaction time of the YA group was significantly shortened in the auditory condition compared with the visual condition [F (1, 18) = 14.06, p<.005]. These results indicate that the time to auditory detection vary depending on age. From the equivalent reaction times in the visual condition, the slower auditory performance in the OA group should not be due to slowed motor responses, but to delayed auditory detection, as indicated in Figure 2.

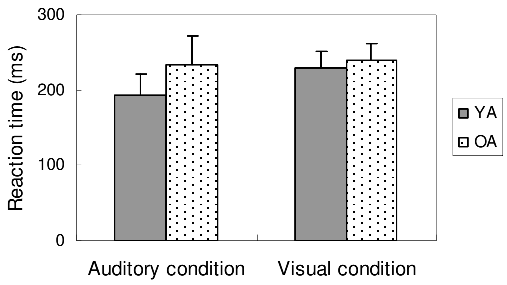


Figure 2: Simple reaction time
Mean simple reaction times (ms) for the YA and OA groups obtained under auditory and visual conditions.
(0ms = Target onset)

2.2. Choice reaction time experiment

2.2.1. *Participants*

Twenty three YA (mean age 20.9 years, range 20-26, 16 females and 7 males) and 14 OA (mean age 65.6 years, range 60-73, 10 females and 4 males) participated in this experiment.

2.2.2. *Stimuli*

Stimuli were the same speech stimuli as those used in [5]. They were prepared by using two female talkers (a native Japanese and a native English talkers) who uttered syllables /ba/ and /ga/. The movie clips were edited for the AV and AO conditions. Both talkers were included in a given block for each condition. The AO stimulus consisted of auditory speech of /ba/ or /ga/ and a visual fixation point. The AV stimulus consisted of matching auditory and visual speech.

2.2.3. *Procedure*

The AO and AV stimuli were presented in the same way as in the simple reaction-time experiment. One experimental block consisted of 40 trials (each of /ba/ and /ga/ presented 20 times per block), and there were two conditions (AO and AV conditions), resulting in a total of 80 trials. The presentation order was counterbalanced across participants. In both conditions, there were two response alternatives "ba" and "ga". The participants were instructed to look at the display, listen to the sound, make a decision whether it is "ba" or "ga", and press one of two buttons. Reaction time was measured as the time from the audio onset to the button press (see Figure 3). The onset of the next stimulus was 1,500 ms after the button press.

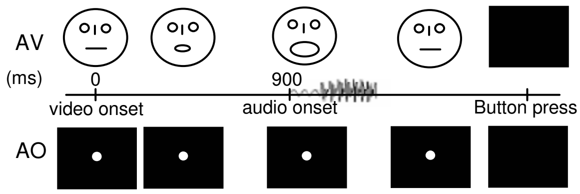
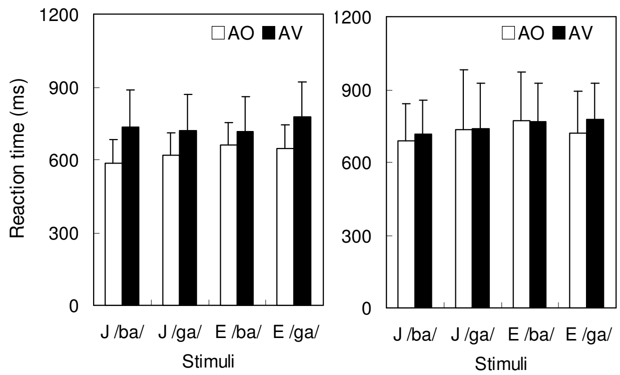


Figure 3: A trial flow of movie
The talker's face appeared 900ms before the audio onset.

2.2.4. *Results*

The choice reaction times were analyzed by an ANOVA [Between-subjects: age group (2) × Within-subjects: condition (2) × talker (2) × syllable (2)]. There was a significant interaction of age group × condition [F (1, 35) = 6.68, p<.05]. Analyses of simple main effect of age showed that the reaction time of the AO condition was significantly longer in the OA group compared with the YA group [F (1, 70) = 4.48, p<.05], but the simple main effect was not significant in the AV condition. Mean reaction time in the AO condition was 627.8 ms for the YA group, and 730.4 ms for the OA group (*audio onset = 0ms*). Analyses of simple main effect of condition showed that the reaction time for the AO condition was significantly faster than that for the AV condition in the YA group [F (1, 35) = 22.72, p<.001], but the simple main effect was not significant in the OA group .



(a) YA group (b) OA group

Figure 4: Choice reaction time
Mean choice reaction time for the YA group (a) and the OA group (b) for each stimulus. The error bars show standard deviation. (0ms = Audio onset)

2.3. Discussion

We studied differences between YA and OA group in simple and choice reaction times. The choice reaction times showed that the OA group was slower than the YA group only in the AO condition and that the two groups were equivalent in the AV condition. Together with the significantly slower AO performance of the OA group in the simple reaction time, age-related change is clear as delayed auditory processing both in auditory non-speech detection and auditory speech perception.

Visual influence on the speed of speech processing differed between the two age groups. The additional visual information slowed down the speech perception processes for the YA group, but such tendency was not shown in the OA group. Thus, it seems that the congruent visual cues interfere with auditory speech identification in Japanese younger adults, but such interference disappears as they age, perhaps due to a need for compensating declined hearing.

3. Experiment 2: ERP study

3.1. Method

3.1.1. Participants

Thirteen YA (mean age 21.4 years, range 20-26, 8 females and 5 males) and 10 OA (mean age 65.4 years, range 60-73, 8 females and 2 males) participated in this study.

3.1.2. Stimuli

Stimuli were identical to those used in the choice reaction -time experiment.

3.1.3. Procedure

The procedure was similar to that of the choice reaction - time experiment, but no overt responses were requested in Experiment 2 in order to rule out motor-related potentials during ERP measurement. A trial consisted of a 2,000 ms stimulus period and 1,500 ms blank interval. The participants were instructed to look at the monitor, listen to the sound, and make a decision whether it was “ba” or “ga” Electroencephalographic (EEG) recordings were conducted during each experimental block. Experiment 2 consisted of 10 AO blocks and 10 AV blocks. Each block had 40 trials (each of /ba/ and /ga/ was presented 20 times per talker).The total number of trials was 800. The AO and AV blocks were alternated, and the presentation order was counterbalanced across participants. The experiment took about 5 hours per participant, and EEG recordings were divided into a few days.

3.1.4. EEG Recording

Neurofax EEG-1100 (Nihon Kohden, Tokyo) and an electro-cap (International, Inc. Eaton, Ohio USA: 10/20 system) were used. EEG data were acquired (sampling rate,500 Hz) from 19 channels of the electro-cap (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). Recording electrodes were referred to the linked earlobes (A1+A2) and ground electrodes were placed on Fpz and nasion. Trigger signals (30 ms) had been inserted on an audio track so that each of them was synchronized with the onset of the speech stimuli. The trigger signals were recorded on the 20th channel of the EEG, and the speech signals on the 21st channel to make sure the synchronization.

3.1.5. Analysis

Average ERPs were processed with the trigger of audio onset. ERPs were calculated separately for the stimulus type.

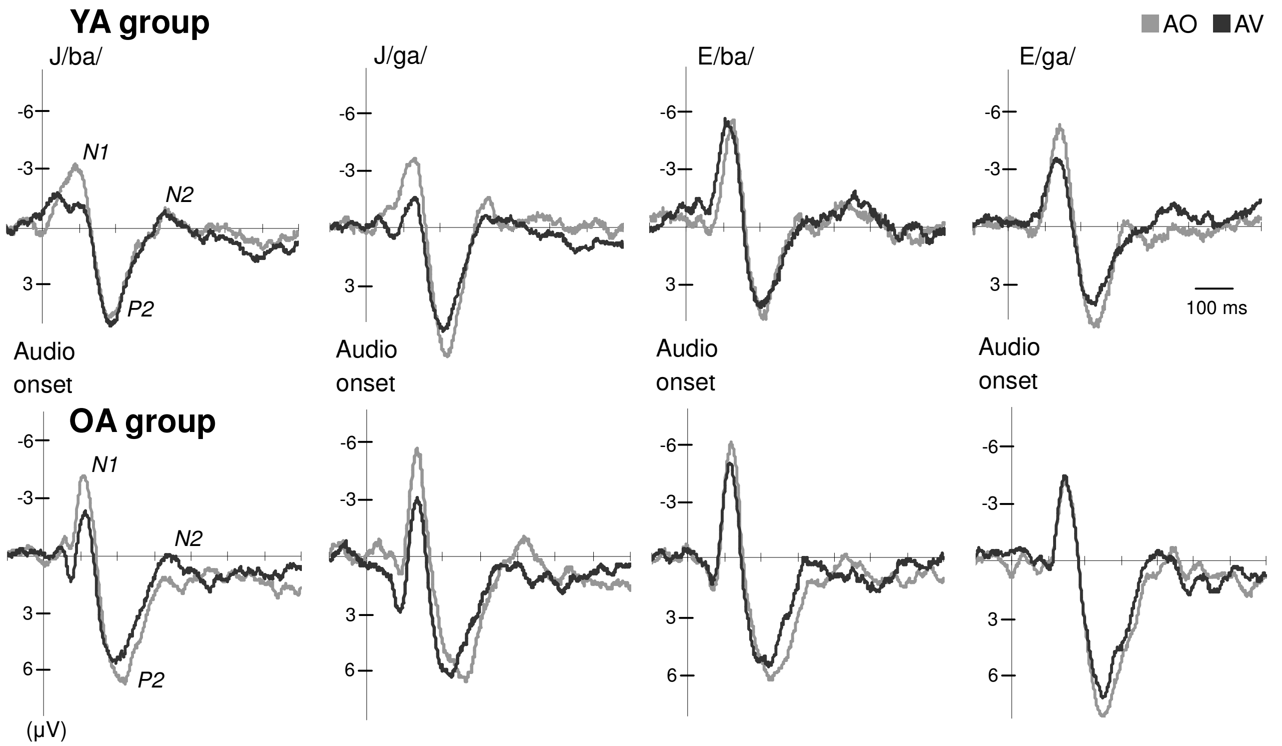


Figure 5: Grand-average
Average ERPs were depicted for stimuli type each age group.

Baseline was corrected on 200ms before an audio onset per participant. The early components N1 and P2 were only analyzed at the maximum electrode (Cz). Additionally, the N2 was analyzed at the same channel in this study.

3.2. Results

The ERP latencies were analyzed by an ANOVA [Between-subjects: age group (2) × Within-subjects: condition (2) × talker (2) × syllable (2)] for each of N1, P2, and N2.. Table1 shows mean latencies in each age group.

3.2.1. N1 component

The ANOVA showed a significant interaction of age group × condition [F (1, 21) = 7.09, p<.05]. Analyses of simple main effect of condition showed that the N1 latency for the AV condition was significantly shorter than for the AO condition in the YA group [F (1, 21) = 19.97, p<.001], but the simple main effect was not significant in the OA group. In addition, the simple main effect analysis of age showed that the N1 peak latency for both the AO and AV conditions were significantly faster in the YA group compared with the OA group [AO condition; F (1, 42) = 4.86, p<.05, AV condition; F (1, 42) = 18.91, p<.001]. These results indicate that the latency reduction at N1 due to the additional visual information was significant only in the YA group, and that the OA group was generally slower than the YA group in N1 latency irrespective of the conditions.

3.2.2. P2 component

The ANOVA found a significant interaction of age group × condition [F (1, 21) = 5.27, p<.05]. Analyses of simple main effect of condition showed that the P2 peak latency for the AV condition was significantly shorter than for the AO condition in the OA group [F (1, 21) = 12.25, p<.005], but the simple main effect was not significant in the YA group. In addition, the simple main effect analysis of age showed that the P2 peak latency for both the AO and AV conditions were significantly faster in the YA group compared with the OA group [AO condition; F (1, 42) = 21.27, p<.001, AV condition; F (1, 42) = 6.99, p<.05]. These results indicate that the latency reduction at P2 due to the additional visual information was significant only in the OA group, and that the OA group was generally slower than the YA group in P2 latency irrespective of the conditions.

3.2.3. N2 component

The ANOVA detected a significant interaction of age group × condition [F (1, 21) = 28.12, p<.001]. Analyses of simple main effect of condition showed that the N2 peak latency for the AO condition was significantly shorter than for the AV condition in the YA group [F (1, 21) = 19.09, p<.001], whereas the latency for the AO condition was significantly longer than for the AV condition in the OA group [F (1, 21) = 9.80, p<.01]. Analyses of simple main effect of age showed that the N2 peak latency of the AO condition was significantly faster in the YA group compared to the OA group [F (1, 42) = 26.04, p<.001], but the simple main effect was not significant in the AV condition. These results indicate that the relationship in N2 latencies between the AO and AV conditions was opposite for the YA and OA groups: with the additional visual information, there was a latency reduction in the OA group, but a latency increase in the YA group. The results also indicate that the OA group was slower than the YA group in the AO condition, but the two groups were equivalent in the AV condition.

Table 1: Peak latencies (ms) for N1, P2, and N2 at Cz

		Conditioin		
Component	Group	AO	AV	p<.05
N1	YA	122.46	112.31	*
	OA	131.25	129.65	
P2	YA	215.77	214.96	
	OA	240.00	228.85	*
N2	YA	356.58	389.04	*
	OA	407.50	384.25	*

3.3. Discussion

In the present study, we examined ERP elicited by audiovisual speech stimuli. Our results revealed age-related ERP changes in audiovisual speech perception for Japanese speakers.

First, we observed a significant shortening of the N1 peak latency in the AV condition compared with AO condition for the YA group, but no shortening was observed for the OA group. Second, in the P2 peak latency, the OA group was faster for the AV condition compared to the AO condition, but the AO-AV difference was not significant for the YA group. Moreover, the N2 latency exhibited an opposite tendency between the two age groups. Whereas the additional visual information reduced the N2 latency for the OA group, it prolonged the N2 latency for the YA group.

In summary, the visual influence differently appeared for the two age groups. The visual promoting effect was found for the OA group in P2 and N2. For the YA group, a visual promoting effect was observed only in N1, and there was an interfering effect in N2.

4. General discussion

We studied both behavioral (Experiment 1) and ERP data (Experiment 2). Based on the simple reaction times for the unimodal auditory and visual stimuli, delayed auditory responses for the OA group were revealed in Experiment 1.

The choice reaction times in AO/AV syllable identification showed that the visual influence is interfering for the YA group, indicating a tendency unique for the Japanese young adults [5]. However, the visual influence was not obvious for the OA group as far as reaction times were concerned, with the same speed for the AO and AV syllables.

In the ERP experiment, there was a promoting visual influence for the OA group in the form of the latency reduction in the AV versus AO condition. The OA group showed such a latency reduction at P2 and N2. In contrast, the YA group did not show such a promoting effect, but an interfering effect at N2.

Testing younger adults, previous research has shown that the N1 latency for audiovisual stimuli is shorter than for the audio-only stimuli [2,5,7,8] and also in the P2 latency [8]. It is not clear why the OA group did not elicit a difference between the AO and AV conditions in N1 latency. The previous study [7] has reported that speech and non-speech AV stimuli which contained visual anticipatory motion resulted in shorter N1 compared to AO stimuli. When the visual motion starts abruptly with the auditory stimulus (as in paper-tearing behavior), the effect disappeared. Thus, in younger adults, the latency reduction at N1 for AV versus AO stimuli presumably depends on whether or not visual information contains anticipatory motion.

In the present study, why the OA group did not show the latency reduction at N1 for the AV stimuli with anticipatory visual motion? At least for static visual stimuli, their processing time for detection was equivalent with the YA group as shown in the simple reaction times. In visual motion processing, older adults may need more time than younger adults because motion processing requires some memory load to integrate information over time. The equivalent N1 latencies for the AV and AO conditions in OA group may represent their delayed unimodal processing of both auditory and visual dynamic information.

Interestingly, we found that the visual influence on N2 latency was opposite between age groups. The N2 latency is known to be longer for difficult discrimination tasks [12,13]. Assuming this, the present results on N2 latency suggest that the AV condition was more difficult than the AO condition for the YA group, and the AO condition was more difficult than the AV condition for the OA group. This is partly consistent with the choice reaction time results in which the YA group needed more time for the AV than for the AO stimuli. Although the OA group did not show a difference between AV and AO conditions in reaction times, ERP data showed the clear difference in the N2 latency.

We found that the YA group depended on the visual information little, but the OA group greatly depended on the visual information. In the older people, hearing gradually declines [9,10, Figure 1]. Such delayed auditory processing is greatly compensated for by visual information, as the latency reduction at P2 and N2 showed. Although the visual motion processing may be also delayed in older adults, it will presumably catch up the auditory processing at some point before P2. An ERP study reported AV integration of phonetic-level at 155 ms from the audio onset [4]. Further analyses will clarify the relationship between the present and previous studies.

We conclude that ERP is powerful to detect age-related changes, that is, more use of visual information in auditory-visual speech processing in older adults.

5. Acknowledgements

This study was supported by a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science (21243040) to KS.

6. References

[1] Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F., "Audiio-visual integration of emotion expression", *Brain Research*, *1242*, 126-135, 2008.

[2] Besle, J., Fort, A., Delpuech, C., & Giard, M-H., "Bimodal speech: early suppressive visual effects in human auditory cortex", *European Journal of Neuroscience*, *20*, 2225-2234, 2004.

[3] Winneke, A. H., & Phillips, N., "Older and younger adults use fewer neural resources during audiovisual than during auditory speech perception", *Proceedings of the 2009 Conference on Audio Visual Speech Processing*, Norwich, 123-126.

[4] Kluncharev, V., Möttönen, R., & Sams, M., "Electrophysiological indicators of phonetic and non- phonetic multisensory integrations during audiovisual speech perception", *Cognitive Brain Research*, *18*, 65-75, 2003.

[5] Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N., ".Audiovisual speech perception in Japanese and Englilsh: Interlanguage differences examined by event-related potentials", *Proceedings of the 2009 Conference on Audio Visual Speech Processing*, Norwich, 38-42.

[6] Sekiyama, K., & Burnham, D., "Impact of language on development of auditory-visual speech perception", *Developmental Science*, *11*, 306-320, 2008.

[7] Stekelenburg, J. J., & Vroomen, J., "Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events", *Journal of Cognitive Neuroscience*, *19*, 1-10, 2007.

[8] van Wassenhove, V., Grant, K. W., & Poeppel, D., "Visual speech speeds up the neural processing of auditory speech", *PNAS, 25*, 1181-1186, 2005.

[9] Glorig, A., & Davis, H., "Age, noise and hearing loss", *Ann Otol Rhinol Larygol*, *70*, 556-571, 1961.

[10] Nixon, J, C., Grolig, A., & High, W, S., "Changes in air and bone conduction thresholds as a function of age", *J Laryngol Otol*, *76*, 288-298, 1962.

[11] Sekiyama, K., & Sakamoto, S., "Hearing deterioration indicated by reaction tine: auditory-visual speech perception in older adults", *Transactions on Technical Committee of Psychological and Physiological Acoustics*, *The Acoustical Society of Japan*, *37*, 799-804, 2007.

[12] Ritter, W., Simson, R.,Vaughan Jr. HG., & Friedman, D., "A Brain Event Related to the Making of a Sensory Discrimination", *Science*, *203*, 1358-1361, 1979.

[13] Towey, J., Rist, F., Hakeren, G., Ruchkin, D. S., & Sutton, S., "N250 latency and decision time", *Bulletin of the Psychonomic Society*, *15*, 365-368, 1980.