Individual differences on an auditory-visual speech perception test for people with hearing loss

TATSUO NAKAGAWA*

College of Education and Human Sciences, Yokohama National University, Yokohama, Japan

Individual differences in auditory-visual speech perception in people with hearing loss were investigated using syllables, words, and sentences. The stimuli were presented in auditory-only, visual-only, and auditory-visual conditions for both congruent and incongruent conditions. In the congruent condition auditory speech stimuli were presented with their identical visual cues, and in the incongruent condition auditory stimuli were presented with conflicting visual cues. Nine young adults with varying degrees of hearing loss, fitted with hearing-aids or cochlear implants participated in the study. The relative increase in auditory-visual speech perception as measured by these tests resulting from the addition of visual cues to the auditory signal was calculated for each condition. The results showed that the subjects were better able to integrate both auditory and visual cues in the auditory-visual congruent condition. The auditory-visual gain in speech perception was less for the incongruent condition. The subjects showed significant individual differences in the amount of gain for different experimental conditions. These results suggest that auditory-visual integration of speech information does occur but that the degree of integration varies among the subjects. The speech stimuli showing the most auditory-visual integration are discussed in the text.

INTRODUCTION

It is well known that we depend on vision in addition to audition in daily speech communication, particularly in difficult listening situations such as in low signal levels and/or in high noise levels. The dependence on visual cues is greater for people with hearing loss. Several studies have shown that some people with hearing loss demonstrated visually biased responses to incongruent auditory-visual stimuli. These data were obtained with different auditory and visual syllables after cochlea implantation (Desai *et al.*, 2008; Rouger *et al.*, 2008). But recently other studies have failed to demonstrate this trend and concluded that the factors in the experimental design, such as subject's proficiency and informational content of the sensory channels, may have accounted for the different results (Schwartz, 2010; Huyse *et al.*, 2012). This study aims to clarify the characteristics of auditory-visual speech perception in college students with varying degrees of hearing loss and

*Corresponding author: t-nakagawa@ynu.ac.jp

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

different kinds of wearable listening devices. It is of particular interest to determine whether auditory visual integration is equivocally seen in different speech stimuli such as syllables, words and sentences, and if there is some difference in the perceived degree of integration between the stimuli, which stimuli are the most suitable for the clarification of auditory-visual integration

METHODS

Participants

Nine college students with hearing loss (7 women and 2 men, mean age=20.6 years, SD=0.5) participated in this study. All students communicated verbally with hearing people in daily life. Two of them had a unilateral cochlear implantation and the remaining 7 students had hearing aids bilaterally or monaurally. Table 1 provides a summary of the average hearing levels and the worn listening devices.

		S1	S2	S3	S4	S5	S6	S7	S8	S9
Right	**MHL**	91	89	55	109	75	50	80	100	109
Left		100	89	129	109	116	55	90	106	109
Right	**Devices**	HA	HA	HA	—	HA	HA	HA	HA	CI
Left		HA	HA	—	CI	—	HA	HA	HA	—

Table 1: Audiological characteristics of the participants. MHL: mean hearing level. Devices: worn listening devices, HA: hearing aid, CI: cochlear implant.

Stimuli

A female Japanese speaker was videotaped producing the speech stimuli. The twenty-one consonant-vowel syllables, twenty familiar words often used in daily school life, and twenty sentences which consist of 35 key words were used. The CV stimuli used in this study were as follows: /a, ki, shi, ta, ni, yo, ji, u, ku, su, ha, ba, ri, ba, o, te, mo, wa, to, ga, da/. A digital camera was set to record the speaker's face and shoulder. All recordings were made in a single walled sound treated room.

Original digitized videotaped stimuli were edited with the specially developed editing software. Auditory signals were digitized at a sampling rate of 24000 Hz and were equalized in level. The synchronization of audio-visual stimuli was measured. It was within 60 ms. For the auditory only (AO) condition, the visual image of the speaker was hidden by visual masking. For the visual only (VO) condition, the audio signal was turned off. The congruent auditory-visual (AV-C) stimuli consisted of digital audio-video files of the speaker saying and articulating the same speech stimuli. For incongruent auditory-visual (AV-I) condition, stimuli were created by combining audio files with non-corresponding video files and matching the onset times. More specifically, in the AV-I condition for syllable presentation, an auditory syllable was paired with a visual syllable whose vowel was the same as the original auditory syllable. In the AV-I condition for word presentation, an auditory word was

paired with a visual word whose first vowel was the same as the original auditory word. In the AV-I condition for sentence presentation, the entire length of an auditory sentence was paired with a visual sentence whose entire length was the same as the original auditory sentence as much as possible.

Procedure

For all nine students stimuli were presented in four conditions: AO, VO, AV-C, and AV-I. All stimuli were presented randomly to the participants. Before testing, practice sessions were used to familiarize the subjects with the procedure. Participants were instructed to listen and/or watch each stimulus and repeat what they judged to have been said. The auditory stimuli were presented at 65 dB SPL.

RESULTS AND DISCUSSION

Figures 1 to 3 show the individual percentage correct scores of the 9 participants for CV syllables, words, and sentences, respectively. Except for one participant (S8), all of the other participants showed higher scores in the AO condition than in the VO condition. These subjects showed even higher scores in the AV-C condition than in the AV-I condition. From these data the participants seemed to use auditory-visual information effectively in integrating auditory and visual stimuli in speech perception but the degree of auditory-visual integration was different between subjects.

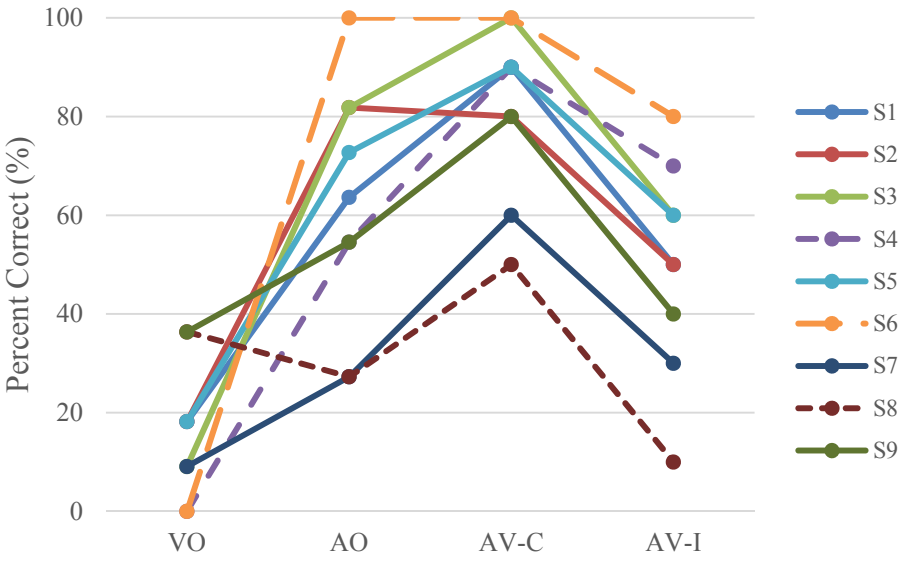

Fig. 1: Syllable identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C: congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

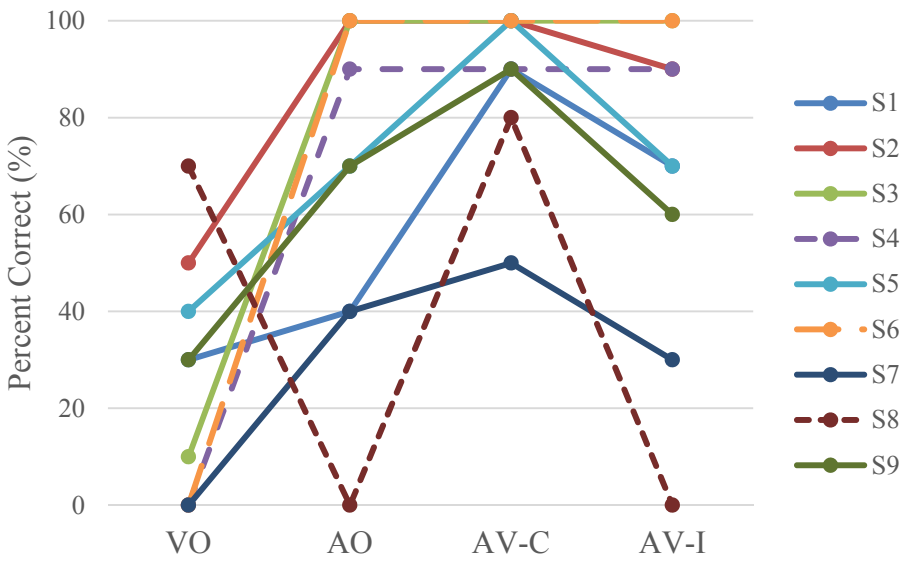

Fig. 2: Word identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C: congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

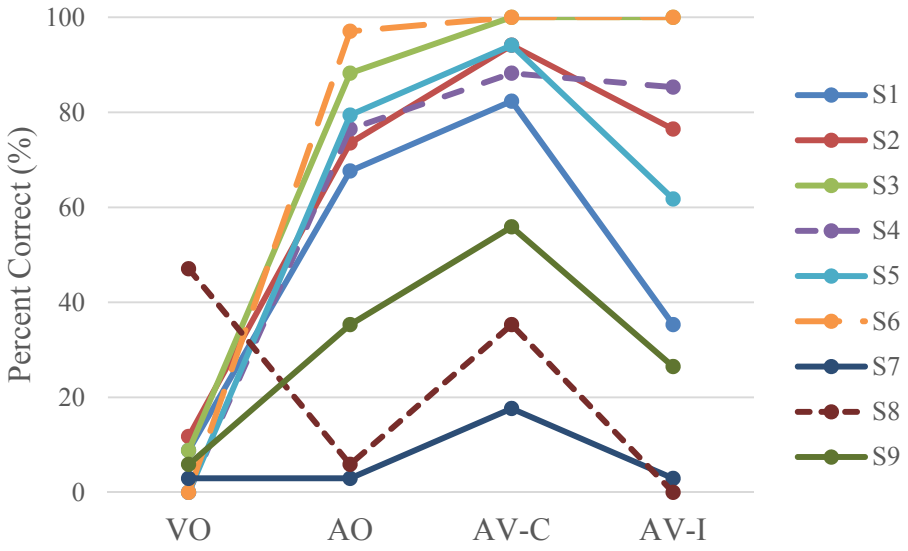

Fig. 3: Sentence identification score (%) for each condition. VO: visual only, AO: auditory only, AV-C congruent auditory-visual, AV-I: incongruent auditory-visual conditions.

Except for 2 participants (S4 and S6), all of the remaining participants showed increased percent correct scores for all speech stimuli in the AV-C condition than in the AV-I condition. For the two exceptional participants, percent correct scores for words and sentences were almost 100% and they appeared to have reached a ceiling effect. The context effects seem to be involved in the process of identification of the words and sentences.

On the other hand unlike other 8 participants, S8 got higher scores in the VO condition than in the AO condition and also got almost the same scores in the VO and AV-I conditions. When the scores between AV-C and AV-I conditions were compared, the former scores were higher than the later ones in syllables, words, and sentences. From these results S8 seems to be more dependent on vision in speech perception.

The relative benefit score (Grant and Seitz, 1998), defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct, was calculated for each subject. Figures 4 to 6 show the individual relative benefit score of 9 participants for CV syllables, words, and sentences, respectively. When plotting relative benefit scores vs. AO performance, the data seemed to be distributed steadily in all three kinds of speech stimuli. But in Figs. 4 and 5 celling effects were seen for the highest performances in the AO condition. Some subjects got the same highest scores in the AO and AV-C conditions, such that their relative benefit scores were zero. But in Fig. 6 when key words in sentences were calculated, the data were scattered evenly. It thus seems that sentences were the most suitable stimuli for clarifying the AV integration.

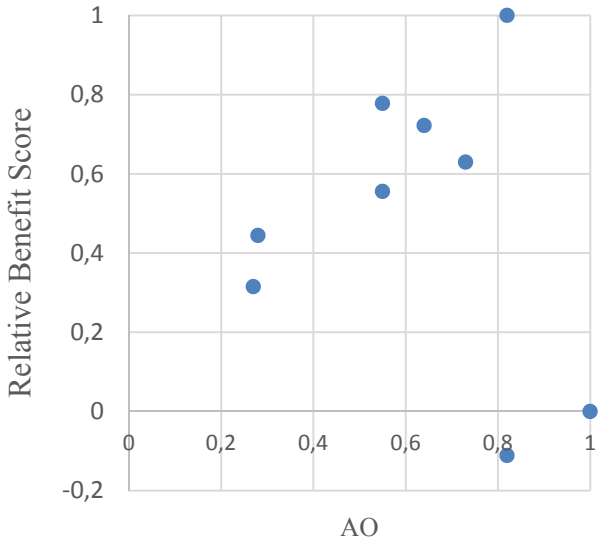

Fig. 4: Relative benefit score for syllables. The relative benefit score, defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

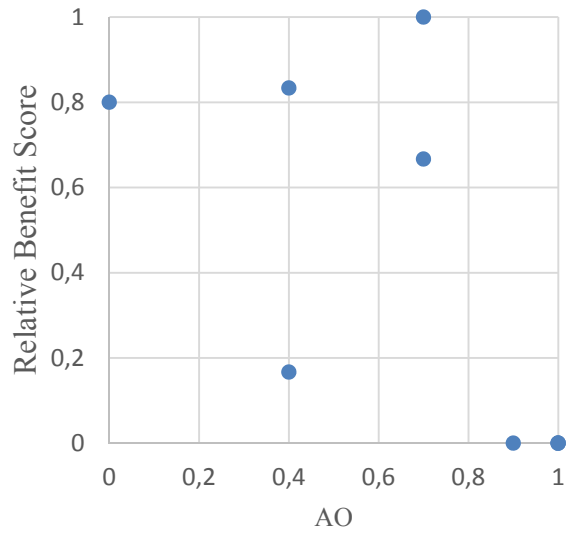

Fig. 5: Relative benefit score for words. The relative benefit score, defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

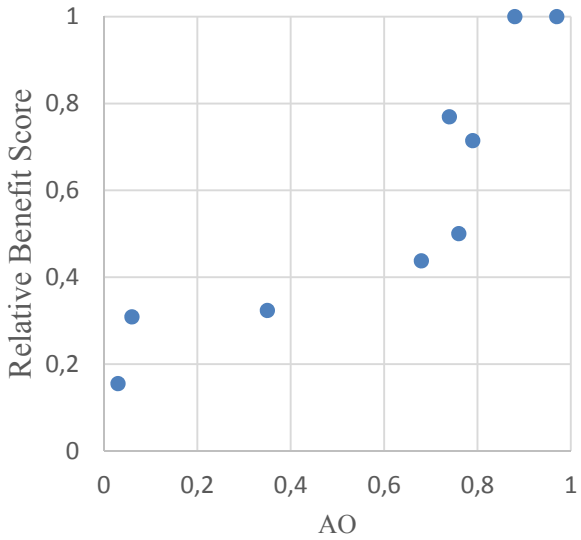

Fig. 6: Relative benefit score for sentences. The relative benefit score, defined as (AV-C－AO)/(1－AO) with AO and AV-C score expressed as percent correct was calculated for each subject. AO: auditory only, AV-C: congruent auditory-visual conditions.

ACKNOWLEDGEMENT

I wish to thank Harry Levitt and Sheila Moodie for helping me to prepare the manuscript.

REFERENCES

Desai, S., Stickney, G., and Zeng, F.G. (**2008**). "Auditory-visual speech perception in normal-hearing and cochlear-implant listeners," J. Acoust Soc Am, **123**, 428-440.

Grant, K. and Seitz, P. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am., **104**, 2438-2450.

Huyse, A., Berthommier, F., and Leybaert, J. (**2012**). "Degradation of labial information modifies audiovisual speech perception in cochlea-implanted children," Ear Hearing, **34**, 110-121.

Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (**2008**). "McGurk effects in cochlear-implanted deaf subjects," Brain Res., **1188**, 87-99.

Schwartz, J. (**2010**). "A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent," J. Acoust. Soc. Am., **127**, 1584-1594.