

## THE MCGURK EFFECT IN JAPANESE AND AMERICAN PERCEIVERS

K. Sekiyama\*, L. D. Braid\*\*\*, K. Nishino\*, M. Hayashi\*\*\*, and M. M. Tuyo\*\*

\* Kanazawa University, Kanazawa, Japan

\*\* Massachusetts Institute of Technology, Cambridge, MA, USA

\*\*\* Osaka City University, Osaka, Japan

### ABSTRACT

This study examined our previous finding that Japanese perceivers are less subject to the McGurk effect than Americans. Stimuli were created using eight speakers. Although the results replicated the group difference, Japanese subjects showed a strong McGurk effect in some cases. The strong visual effect in the Japanese was related to auditory ambiguity and the visual robustness of the stimuli whereas the McGurk effect in Americans was stable in various cases.

### PURPOSE

The McGurk effect demonstrates that visual lip movements influence perception of auditory speech even when the two sources of information are in conflict [1]. For example, when auditory /ma/ is dubbed onto visual lip movements of /ta/, this auditory-visual speech will be often perceived as "na." Whereas we found this effect to be robust for American perceivers, it was much weaker for Japanese perceivers when stimuli were Japanese syllables [2, 3]. This finding suggests that linguistic and/or cultural framework affects the manner of integration. This study examined this inter-language difference further with new stimuli, because our previous finding was based on stimuli from only one Japanese speaker. Experiment 1 was to see if the group difference is replicated for various Japanese speakers. In Experiment 2, the examination was done with more forceful visual stimuli of both Japanese and American speakers.

### METHOD

#### Subjects

The subjects were native speakers of Japanese and American English with normal hearing and normal or corrected vision. In Experiment 1, the subjects were 24 students at Kanazawa University

and 24 students at Massachusetts Institute of Technology (MIT). In Experiment 2, the subjects were 16 students at Osaka City University and 16 newly recruited students at MIT. All the subjects were age under 30 and had no experience of living in a foreign country.

#### Stimuli

Eight syllables were used: /ba/, /pa/, /ma/, /da/, /ta/, /na/, /ga/, and /ka/.

In Experiment 1, the speakers were four native speakers of Japanese. Recorded sound and videotaped lip movements from one speaker were combined and sixteen pairs were created so that each syllable had an auditory-visual discrepant pair (as shown in Table 1) as well as an auditory-visual identical one (audio /pa/, video /pa/).

Table 1. Combinations of auditory (A) and visual (V) syllables in auditory-visual discrepant pairs. R shows a typical response when the A and V syllables are perceptually integrated.

	A	V	R
Labial	b	g	d
	p	k	t
	m	n	n
Nonlabial	d	b	bd
	t	p	pt
	n	m	mn
	g	b	bg
	k	p	pk

In Experiment 2, among 30 who were videotaped, we chose two Japanese and two American speakers whose utterances were the clearest to lipread. The auditory-visual stimuli were created using the same syllable combinations as in Experiment 1. Considering that the duration of Japanese vowels is much shorter than that of English ones, the Japanese speakers were instructed to

pronounce vowels longer than usual so that their duration is comparable with those of English stimuli. This resulted in slower articulations than usual, which seemed to make the visual stimuli easier to lipread.

#### Procedure

In both experiments, the stimuli were presented in three conditions: auditory-visual (AV), visual (V), and auditory (A)

conditions. The subjects were asked to write what they heard (AV and A conditions), or what they thought the speaker was saying (V condition), as an open choice response.

#### RESULTS

Figure 1 shows the magnitude of the McGurk effect (visual effect) for each speaker. The visual effect here refers to the influence of discrepant visual cues. It

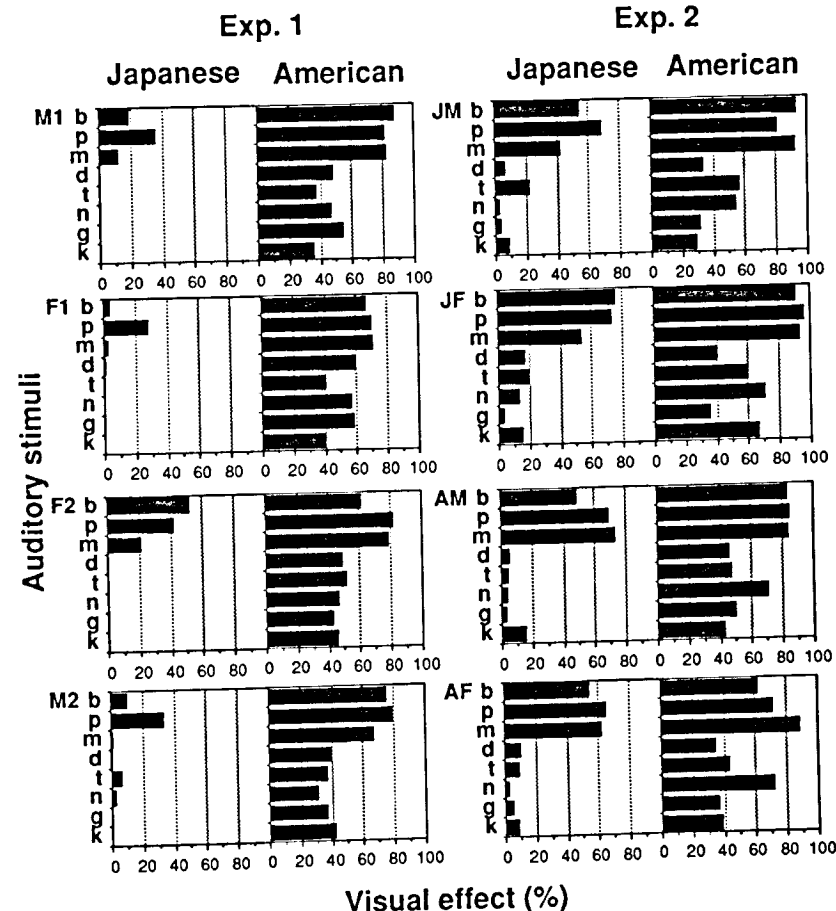


Figure 1. The magnitude of the McGurk (visual) effect in the Japanese and American subjects for each auditory syllable of each speaker. In Experiment 1, all the speakers, were Japanese, two males (M1, M2), and two females (F1, F2). In Experiment 2, there were two Japanese (JM: male, JF: female) and two American (AM: male, AF: female) speakers. The results are shown only for auditory-visual discrepant pairs.

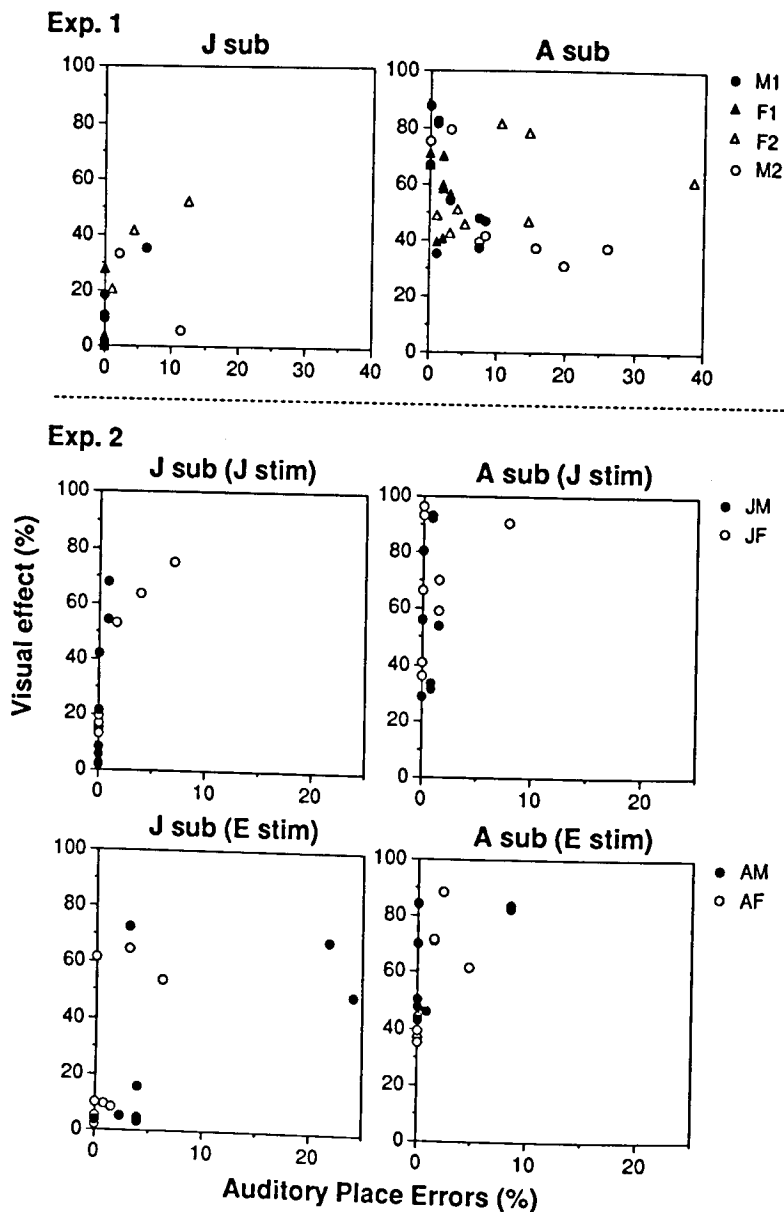


Figure 2. The relationship between the magnitude of the McGurk effect (visual effect) and auditory ambiguity (auditory place errors, i.e., errors in terms of place of articulation when presented with no visual cues). Each point corresponds to one of the eight auditory syllables pronounced by one speaker.

was defined as percent of auditory errors in terms of *place of articulation* (labial vs. nonlabial) caused by discrepant visual cues. For example, if auditory /ba/ is perceived as "da" in 60 % of the cases when combined with visual /ga/ and this "da" response occurs in 5% to the cases when auditory /ba/ is presented with no visual cues, then the visual effect is  $60 - 5 = 55\%$ .

The results of Experiment 1 replicated our previous finding, showing a much weaker McGurk effect for the Japanese subjects than for the American subjects. Agreeing with our previous results, the McGurk effect was stronger when auditory stimuli were labial.

In Experiment 2, although the group difference was replicated, the Japanese subjects showed a stronger McGurk effect than before for auditory labials, together with large individual differences. Unlike our previous study [3], there were no difference between the Japanese and English stimuli.

## DISCUSSION

### Effect of speaking rate

Although the two experiments replicated the result that the McGurk effect is stronger in the American subjects than in the Japanese subjects, the McGurk effect for the Japanese subjects perceiving Japanese speech was stronger in Experiment 2 than in Experiment 1 and our previous study. We believe this difference to be due to the forcefulness of visual stimuli in Experiment 2, where we choose speakers easy to lipread as well as instructed the Japanese speakers to *lengthen* the vowels. We did not instruct the Japanese speakers to do so in Experiment 1 or our earlier study [3]. We instructed the American speaker in our earlier study to *shorten* the vowels. These facts suggest that the slower articulations in Experiment 2 led to more forceful visual stimuli, which increased the magnitude of the visual effect for the Japanese subjects.

### Effect of auditory ambiguity

Figure 2 shows the relationship between the magnitude of the McGurk effect and auditory ambiguity. We

hypothesized that if an auditory token has ambiguous quality, it is very susceptible to the visual effect, whereas an auditory token of unequivocal quality will not be influenced by visual cues so easily. In Figure 2, auditory ambiguity is indicated by auditory place errors.

The results supported this auditory ambiguity hypothesis only for the Japanese subjects perceiving Japanese speech. Look at the upper two panels in the left column. In these cases, it is found that the magnitude of the visual effect is limited when the percent of auditory place errors is zero: The visual effect was less than 30% in Experiment 1 and less than 40% in Experiment 2. When there are some auditory place errors, the magnitude of the visual effect is an increasing function of the ambiguity. Thus, if the percent of auditory place errors could be smaller than zero, then the two indexes might show a linear correlation.

In contrast, the results for the American subjects show strong visual effects even when the percent of auditory place errors was zero.

These results show that the magnitude of the McGurk effect in the Japanese subjects tends to vary depending on quality of auditory and visual speech, while the McGurk effect in American subjects are stable under various conditions.

## ACKNOWLEDGMENT

This study was supported by grants from HFSP, NIH, and the Japanese Ministry of Education.

## REFERENCES

- [1] McGurk, H. and MacDonald, J. (1976) "Hearing lips and seeing voices," *Nature*, Vol. 264, Pp. 746-748.
- [2] Sekiyama, K. and Tohkura, Y. (1991) "McGurk effect in non-English listeners," *Journal of the Acoustical Society of America*, Vol. 90, Pp. 1797-1805.
- [3] Sekiyama, K. (1994) "Differences in auditory-visual speech perception between Japanese and Americans," *Journal of the Acoustical Society of Japan (E)*, Vol. 15, Pp. 143-158.