Inter-language differences in the McGurk effects for Dutch and Cantonese listeners

de Gelder, B.; Bertelson, P.; Vroomen, J.; Chen, H.C.

Published in:
Eurospeech 1995

Publication date:
1995

De Gelder, B., Bertelson, P., Vroomen, J., & Chen, H. C. (1995, september). Inter-language differences in the McGurk effects for Dutch and Cantonese listeners. Proceedings of the Fourth European Conference on Speech Communication and Technoloy, Madrid, pp. 1699-1702

INTER-LANGUAGE DIFFERENCES IN THE MCGURK EFFECT FOR DUTCH AND CANTONESE LISTENERS

[1,2]Beatrice de Gelder, [1,2]Paul Bertelson, [1]Jean Vroomen and [3]Hsuan Chin Chen
[1]Tilburg University, [2]Université Libre de Bruxelles, and [3]Chinese University of HongKong

ABSTRACT

A group of Dutch and Cantonese listeners were compared on a audio-visual speech perception task. Using video techniques, lipmovements of syllables were dubbed on a speech signal such that the heard and seen place of articulation did not match [4]. The Cantonese participants were more influenced by vision than the Dutch. We suggest that the phonological repertoire has an influence on audio-visual speech perception.

1. INTRODUCTION

Speech is perceived in the auditory as well as in the visual modality. While the auditory modality is by far the dominant and the most explored one, lipreading is a powerful source of information for understanding speech in noise as well as in normal hearing circumstances. The most convincing demonstration of the importance of the visual modality is given by the McGurk-illusion [4]: when a visual 'ga' is dubbed on an auditory 'ba', subjects often report hearing 'da'. The McGurk-effect is a perceptual illusion: The locus of the blends or fusions is in speech perception and not at a strategic postperceptual decision level. As a perceptual phenomenon, the McGurk-illusion is subject to the actual phonetic information and the phonological representations involved in the processing of the two modalities. It is now well established that the phonology of the native language tailors the speech processing architecture of the listener. The differences one expects to find in the perception of a given set of stimuli in two different linguistic groups is thus a matter of language-specific processes.

As such, there appears to be little reason for expecting linguistic or, a fortiori, cultural differences in the occurrence of the McGurk-effect. Of course, because of differences in phonological repertoire between languages, one should expect a certain amount of language-specificity of actual blends and fusions in a given linguistic population. Another way of arriving at the same prediction is by stressing the independence of processes and representations underlying the perception of speech, including visual speech, and the functional architecture involved in the processing of faces. To the extent that speech perception and face perception are two functionally independent processes, observed differences between linguistic groups on audio-visual speech processing may have two very different origins. The effects of possible cross-linguistic differences must sharply be distinguished from the effects of cross-cultural differences in the perception of faces and facial expressions. This means to say that prima facie observed differences in cross-

linguistic speech processing cannot be explained by reference to cross-cultural differences in face perception. Of course, if the absence of a visual effect must be traced to partial inattention of the study participants for the visual information, as mentioned in [5], it becomes somewhat difficult to interpret the data.

Recent papers [2, 3, 5, 6, 7] have addressed the issue of language- and culture- specificity of the McGurk-effect. In [6] it was reported that in a McGurk-type of situation Japanese listeners showed very little effect of the visual speech information when no noise was present. This results was confirmed in [5]. Besides perceptual judgements, they also obtained incompatibility ratings from their subjects. The data showed that the low McGurk-like effect is inversely related to the subjects' incompatibility ratings. On the above notion of the McGurk-effect as a perceptual illusion and on the strength of our distinction between effects due to the phonological repertoire and possible effects due to overall differences in language and culture, we would predict to observe the former, but not the latter when testing two different native speaker groups with the same materials. The present study addresses that issue by making a comparison in audio-visual speech perception between a group of native Dutch and native Cantonese speakers.

2. METHOD

Subjects. Two groups of subjects were tested: a group of 18 native Dutch speakers and a group of 18 native speakers of Cantonese with no knowledge of Dutch. All subjects were college students. Testing took place in small groups subjects.

Materials and procedure. Subjects watched a video recording of a female speaker. They were asked to repeat what she said. The speaker had been recorded on U-matic tape while pronouncing a series of VCV syllables. Each syllable consisted of one of the four plosive stops / p, b, t, d/ or a nasal / m, n/ in between the vowel / a/ (e.g., / aba or / ana/). There were three presentation conditions: an audio-visual, an auditory-only, and a visual-only. In the audio-visual presentation, dubbing operations were performed on the recordings so as to produce a new videofilm comprising six different auditory-visual combinations: auditory / p, b, t, d, m, n/ were combined with visual / t, d, p, b, n, m/ , respectively. Thus, the visual place of articulation feature never matched the auditory place feature. The dubbing was carried out so as to ensure that there was auditory-visual coincidence of the release of the consonant utterance. For the auditory-only condition, the original auditory signal was dubbed onto a video signal from the speaker while sitting quietly. For the visual-only condition, the auditory channel was deleted from the recording, so the subject had to rely entirely on lipreading. Each presentation condition comprised of three replications of the six possible stimuli. There was a 5-sec gap of blank film between the successive trials. To counterbalance presentation order, each condition was divided into two blocks of nine trials each. The presentation order of these blocks was always audio-visual, auditory-only, visual-only, visual-only, auditory-only, audio-visual. Stimuli were presented on a 19-inch TV screen. The subjects were instructed to watch the speaker and repeat what she had said. References to modality were strictly avoided. The subjects' response was written down by the experimenter. During the presentation the experimenter monitored subjects in order to make sure that the screen was being watched.

3. RESULTS

In the audio-visual condition there were three possible scoring: fusions, blends, or auditory responses. A fused response is one where visual information of the place of articulation is combined with the auditory information into a single syllable (e.g., ba-auditory/ da-lips into a / da/ response), a blend is a response were the visual place information is added to the auditory information into a two-phonemes composite (/ bda/), and an auditory response is one where vision did not have an influence (/ ba/). When an auditory bilabial was paired with a visual lingual (e.g., auditory / ba/ with visual / da/), Cantonese subjects reported 26% blend responses (/ bda/), 49% fused responses (/ da/), and 24% auditory responses (/ ba/). Dutch subjects reported only 9% blends [$t(34) = 3.04$, $p < .005$], 56% fusions (n.s), and 35% (n.s.) auditory responses. There were no significant difference between the two groups when a visual bilabial was paired with an auditory lingual (e.g., ba-visual and da-auditory). Cantonese subjects: 69% blends, 17% fusions, 14% auditory responses; Dutch subjects: 66% blends, 22% fusions, and 12% auditory responses. The cross-linguistic difference in susceptibility for the McGurk illusion is thus such that Cantonese subjects report more blends than Dutch when a visual lingual is combined with an auditory bilabial.

4. DISCUSSION

Systematic studies of lipreading over the last fifteen years have established that lipreading is a modality of spoken language processing. This means to say that the ability to process visual speech is based on linguistic representations and processes that are likely to relate to the same abstract competence for language as auditory speech does. Functionally and neuropsychologically, heard and seen speech appear increasingly similar [1]. There is thus at present little theoretical basis for expecting that lipreading would occur less in some linguistic communities than in others. Such a claim would amount to saying that some linguistic communities process auditory speech in a different way than others, a claim that is obscured to say the least. For the same reason there is no basis for a prediction that conflict between the auditory and the visual modality would not occur in speakers of some languages. Of course, there is room for language-specific ways of processing visual speech, given the differences in phonological repertoire of languages. The results of the present study illustrate both aspects. The two groups are comparable in auditory as well as in visual speech processing and both show blends as well as fusions. The most striking difference between the groups concerns the number of blends in the Cantonese group in one condition. Given the overall similarity between the groups, there is reason to believe that this difference follows from phonological differences between the two languages and the consequences these have for subjective processing stategies. One factor that can not be ingored in this context concerns the impact of orthographic strategies. In reporting blends the Cantonese subjets produce consonant clusters which do not exist in Cantonese. Further research needs to explore whether Cantonese subjects in writing down their answers rely on an alphabetic strategy to resolve the audiovisual conflict in some cases.

REFERENCES

[1]: Campbell, R., de Gelder, B., & de Haan, E. (submitted). Lateralisation of lipreading: a second look.

[2]: de Gelder, B., & Vroomen, J. (1992). Auditory and visual speech perception in alphabetic and nonalphabetic ChineseDutch bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 413-426). North Holland: Elsevier Science Publishers.

[3]: Massaro, D.W., Cohen, M.M., Gesi, A., Heredia, R., & Tsuzaki, M. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics, **21***, 445-478.

[4]: McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, **264***, 746-748.

[5]: Sekiyama, K. (1994). Differences in auditoryvisual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan*, 15(3), 143-158.

[6]: Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90*, 1797-1805.

[7]: Sekiyama, K., & Tohkura, Y. (1993). Interlanguage differences in the influence of visual cues in speech perception. *Journal of Phonetics,* 21, 427-444.