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

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

ARE THE MCGURK ILLUSIONS AFFECTED BY LEFT OR RIGHT PRESENTATION OF THE SPEAKER FACE ?

C. COLIN (1) & M. RADEAU (1) (2)

(1) Free University of Brussels, Belgium

(2) National Fund for Scientific Research, Brussels, Belgium

10.21437/Eurospeech.1999-380

ABSTRACT

When conflicting syllables are presented in the auditory and visual modalities, two kinds of illusions have been reported by McGurk and MacDonald (1976): combinations (cluster responses) and fusions (fused responses). In the present experiment, we examined the lateralization issue of these illusions. In one half of the session the center of the TV screen was displaced 5° to the right of straight ahead and in the other half, it was displaced 5° to the left. The sound, played at an average level of 40dB, always came from straight ahead. Neither for fusions nor for combinations did we found any difference in the percentage of illusion as a function of the hemifield in which the visual stimuli were presented. This result does not corroborate Diesch's (1995) finding. Using a somewhat different method, this author found a left hemifield advantage for fusions and a right hemifield advantage for combinations.

INTRODUCTION

Speech perception has often been considered as a purely auditory process until Sumby and Pollack [1] showed that lipreading improves the intelligibility of an auditory signal embedded in noise. However, lipreading also influences speech perception in normal listening conditions. Indeed, when confronted to discrepant auditory and visual speech information, participants often report hearing phonetic segments that are not presented either auditorily, or visually. Typical demonstrations are the fusions and combinations reported by McGurk and MacDonald [2]. For example, when an acoustic /ba/ is dubbed onto a visual /ga/, a "fused" response /da/ is likely to occur; with the opposite presentation, the subsequent perception is a combination response such as /bga/. Thus, the perception is not merely auditory or visual, but rather seems to be an integration of sound and sight.

Several studies raised the question of the cerebral lateralization of audiovisual speech. Although no complete agreement has been reached, most data from neurological studies are in favor of the left hemisphere. In particular, the skills and impairments of the neuropsychological cases reported by Campbell et al. [3] are consistent with the idea that audiovisual speech could be associated with speech processing (assumed to be performed by the left hemisphere) and not with faces processing (assumed to be performed by the right hemisphere). More recently, Campbell [4] showed in a activate the same areas: the auditory cortex (especially the Brodmann area) of the left hemisphere and the visual areas bilaterally. Using electromagnetic recordings (over the left hemisphere only), Sams et al. [5] located the integration of auditory and visual speech components in the left supratemporal auditory cortex. They concluded that visual information from articulatory movements may have an entry into the auditory cortex. However, a recent PET study of the McGurk effect [6] provided data strongly divergent from those previous results, since only right subcortical regions were found to be activated (insula, hippocampus and superior colliculus). Behavioral studies still provide a different pattern of results. Baynes et al. [7] showed a contribution of both hemispheres in audiovisual speech integration. However, when the two types of McGurk illusions (fusions and combinations) were studied separately, different results emerged [8]. So, Diesch, in a McGurk study conducted in German, displayed the visual stimuli in the left or in the right hemifield. On each trial, two speakers' faces were presented side by side on a TV screen but only one of them was uttering a syllable. The eccentricity of the center of the speaker's mouth on both the left and right sides of the screen was 3,5° of visual angle. Results showed a left hemifield advantage for fusions and a right hemifield advantage for combinations

The aim of the present work was to re-examine the cerebral lateralization issue of the McGurk effect, using a somewhat different presentation procedure than in Diesch' study. Instead of showing the two faces side by side, we presented only one face at a time on a whole screen. The screen was placed either on the participant's left side, or on their right side (with a 5° eccentricity between the center of the screen and the straight ahead position). In addition, whereas in Diesch's experiment the auditory signals were sent through two loudspeakers located on each side of the screen, here the sound always came from a loudspeaker positioned straight ahead.

In a previous study [9], conducted in French, we manipulated several variables that were likely to have an effect on the McGurk illusions: the item length (CV monosyllables vs VCV bisyllables), the speaking rate of the bisyllables (slow, medial or fast), the vocalic environment (/a/ vs /i/), the voiced-voiceless feature of the consonant (/b/-/g/ vs /p/-/k/), the speaker gender and the sound intensity level (70 dB vs 40 dB). In the present study, we used the same materials and manipulated the same variables except sound intensity. All stimuli were played at an average of 40dB, an intensity level which proved to give rise to a stronger McGurk effect than 70dB

METHOD

Participants

Thirty-two students (17 to 24 years) participated in the experiment as part of an introductory psychology course. They were all French speaking, without reported history of hearing disorder and with normal or corrected-to-normal vision. Sixteen participants were presented the tapes of a woman speaker and the 16 other those of a man speaker.

Materials

The materials consisted of eight CV monosyllables (ba, ga, pa, ka, bi, gi, pi, ki) and eight VCV bisyllables (aba, aga, apa, aka, ibi, igi, ipi, iki). They were articulated by two speakers, a woman and a man. Only the lower part of their face was filmed (from the top of the nose till the chin). The bisyllables were uttered according to three speaking rates: slow, medial and fast.

The stimuli were constructed on a Panasonic AG-A770 editing controller. All experimental items were incongruent. They were created by replacing the original audio signal of a given item by the audio signal corresponding to another item of about the same length in frames. The sound was synchronized with the last picture preceding mouth opening. For example, an auditory aba was dubbed onto a visual aga, an auditory ipi onto a visual iki, … Eight kinds of stimuli were edited this way for each item length and, in the case of the bisyllables, for each speaking rate.

In order to test the intelligibility of the auditory stimuli, control items consisting of the same auditory syllables as the experimental items were recorded on a still face. Four tapes were edited, two for each speaker and for each kind of items (experimental vs control). Each tape consisted of four blocks, each corresponding to one of the three speaking rates of the bisyllables or to the monosyllables. A block was made up of 24 trials (eight different stimuli repeated three times each) presented in random order. Each block also contained two or three catch trials consisting in funny faces (the speakers putting out the tongue, for example) that the participants had to detect in order to check whether they looked at the screen.

Procedure

The participants seated in front of a table, at 75 cm from a Panasonic color screen (width: 33 cm; height: 25 cm). They had the head on a chin rest during the whole experiment and were asked to keep their eyes fixating a red strap of paper, stuck vertically on the middle of the screen and horizontally in such a way that it corresponded to the center of the table and felt on the speaker's mouth corner. For one half of the participants, the two tapes corresponding to one speaker were presented first in the right hemifield while for the other half, they were presented first in the left hemifield. Between two presentations, the experimenter changed the screen location and the red strap position. The eccentricity between the center of the table and the center of the screen was 5°. The sound was played, at an average level of 40dB, through a loudspeaker located on the top of the screen.

The participants had to choose between several [illegible] and /other/) the answer corresponding to what they had heard. Two items were separated by a 3 s ISI which consisted of a black screen period, during which the participants had to write their answer. The session began by a 16 trials training block and lasted about one hour.

RESULTS

For each item, the participants were likely to give an answer corresponding to the auditory information, or to the visual information, or an audiovisual response (that is a combination or a fusion). We considered as illusory responses, the audiovisual responses and also the visual responses. These responses were included because in more than 90% of the cases, there was no confusion concerning the voiced-voiceless feature of the reported consonant (for example, /b/ was almost never confused with /p/, so was it for /g/ and /k/). Since voicing is conveyed by the auditory modality, not by the visual one, the visual responses could thus not be explained in terms of pure lipreading but involved necessarily the integration of the auditory information. Although control trials gave rise to correct auditory identification in more than 90% of the cases, errors on these trials were subtracted from the total number of illusory responses for each participant.

The results are presented in Table 1, as a function of Type of consonant, Type of vowel, Item length, Speaker gender and Hemifield of presentation.

	Combinations					**Fusions**				
	Man		Woman			Man		Woman		
	B	M	B	M	**Me**	B	M	B	M	**Me**
Right hemifield										
V+	44	25	69	41	**45**	30	33	46	12	**30**
V-	59	54	68	65	**61**	30	31	330	7	**25**
/a/	48	44	713	53	**54**	22	19	44	19	**26**
/i/	55	36	66	52	**52**	37	44	36	0	**29**
Me	**51**	**40**	**68**	**53**	**53**	**30**	**32**	**40**	**10**	**28**
Left hemifield										
V+	45	18	61	45	**42**	28	36	48	12	**31**
V-	58	63	62	74	**64**	26	33	360	9	**26**
/a/	45	38	66	65	**53**	18	21	45	19	**26**
/i/	58	43	58	54	**53**	37	38	40	1	**29**
M	**51**	**40**	**62**	**59**	**53**	**27**	**32**	**42**	**10**	**28**

Table 1: Percentages of combinations and fusions as a function of Speaker gender, Item length (B=bisyllables; M= monosyllables; **Me** = mean), Hemifield, Consonant (V+ = voiced; V- = voiceless) and Vowel (/a/; /i/).

Eight separated Anovas were performed on each item length (monosyllables vs bisyllables), each type of illusions (combinations vs fusions) and each speaker (the man being systematically faster than the woman, the results of the two speakers were not comparable). Anovas were carried out with Screen position (left or right), Type of consonant (voiced vs voiceless), Type of vowel (/i/ vs

/a/) and, for the bisyllables, Speaking rate (slow, medial, fast) as within-participants variables.

Screen position was never significant and did not interact with any other factor. For both illusions, both item lengths and both speakers, the illusion percentages were equivalent whatever the screen was located at the participant's right or left. We will thus no longer mention this factor in our results report.

Bisyllables

Combinations

For the man speaker, Consonant and Vowel effects were significant. Voiceless consonants elicited 14% combinations more than voiced ones ($F(1,15)=12,39$, $p<.01$) and /i/ produced 10% illusions more than /a/ ($F(1,15)=5,99$, $p<.05$). The interactions between Consonant and Speaking rate ($F(2,30)=4,21$, $p<.05$) and between Vowel and Speaking rate ($F(2,30)=4,79$, $p<.05$) were also significant, as well as the triple interaction ($F(2,30)=3,55$, $p<.05$). The results of all significant contrasts were very congruent. We always observed an advantage of the slow speaking rate over the other rates, of voiceless consonants over voiced ones and of the vowel /i/ over /a/.

Only Speaking rate was significant for the woman speaker ($F(2,30)=11,15$, $p<.01$), the fast rate giving rise to more illusions than the slow (20%; $F(1,15)=18,09$, $p<.01$) and than the medial ones (16%; $F(1,15)=9,61$, $p<.01$). The only significant interaction was between Vowel and Speaking rate ($F(2,30)=7,79$, $p<.01$). When the vowel was /i/, there were more combinations with the fast speaking rate than with the other two rates.

Fusions

The vowel /i/ produced significantly more fusions (17%) than the vowel /a/ ($F(1,15)=9,12$, $p<.01$) for the man speaker. Speaking rate was also significant ($F(2,30)=14,75$, $p<.01$), the slow and medial rates eliciting, respectively, 11% and 17% fusions more than the fast one (slow vs fast: $F(1,15)=18,71$, $p<.01$; medial vs fast: $F(1,15)=31,47$, $p<.01$). All double interactions were significant (Vowel x Consonant: ($F(1,15)=12,29$, $p<.01$); Speaking rate x Consonant ($F(2,30)=3,55$, $p<.05$); Speaking rate x Vowel ($F(2,30)=3,65$, $p<.05$). For all significant contrasts, the vowel /i/ gave rise to more illusions than the vowel /a/ and the slow and medial speaking rates produced more illusions than the fast rate. No contrast involving the Consonant factor was significant.

Concerning the woman speaker, voiced consonants produced 12% fusions more than voiceless ones ($F(1,15)=11,81$, $p<.01$). The interaction between Vowel and Consonant was significant ($F(1,15)=32,03$, $p<.01$), as well as the interaction between Vowel and Speaking rate ($F(2,30)=4,15$, $p<.05$). The contrasts involving the Consonant effect confirmed the tendency found for the main effect (more fusions with voiced than with voiceless consonants). For Speaking rate, we found more illusions with the slow and fast rates relative to the medial one. Finally, for vowels, there was a tendency to observe more illusions with /a/ than with /i/.

Monosyllables

Combinations

The Consonant effect was the only significant factor for the man speaker ($F(1,15)=14,06$, $p<.01$), voiceless consonants producing 37% illusions more than voiced ones. The Consonant effect was also significant for the woman speaker ($F(1,15)=25,82$, $p<.01$) and went in the same direction as for the man speaker (26% combinations more with voiceless consonants). The only significant interaction was that between Vowel and Consonant ($F(1,15)=41,19$, $p<.01$). Illusions were more numerous with voiceless consonants when the vowel was /i/. There was no clear tendency for the vowel

Fusions

The Vowel effect was significant for the man speaker ($F(1,15)=6,00$, $p<.05$), /i/ giving rise again to 21% fusions more than /a/. The interaction between Vowel and Consonant was also significant ($F(1,15)=5,44$, $p<.05$) and was due to a greater percentage of illusions with /i/ and with voiced consonants. For the woman speaker, the Vowel effect was also significant ($F(1,15)=6,56$, $p<.05$) but exhibited a pattern opposite to that found for the man speaker (18% fusions more with /a/).

Comparison between bisyllables and monosyllables

The results of the monosyllables were compared to the mean results of the three speaking rates of the bisyllables in four Anovas (two speakers and two kinds of illusions) with three within-participants factors (Consonant, Vowel and Item length). Item length was significant only for the fusions of the woman speaker, bisyllables giving rise to 31% illusions more than monosyllables ($F(1,15)=16,53$, $p<.01$). The interaction between Item length and Vowel was also significant ($F(1,15)=5,34$, $p<.05$) as well as the triple interaction ($F(1,15)=10,21$, $p<.01$). In all significant contrasts, we observed more fusions with the bisyllables than with the monosyllables.

CONCLUSION

The main aspect of our results is that we failed to replicate Diesch's data. We observed a similar number of combinations and of fusions in both visual hemifields.

However, as regards the other factors under study, our results confirm those obtained in our previous experiments [9]. Playing the auditory stimuli at a rather low intensity level (40dB) led to a substantial number of illusions (53% of combinations and 28% of fusions). In our former studies, we observed an increase of the illusions as the intensity level of the auditory stimuli decreased from 70dB to 40dB (41% vs 49% for combinations and 3% vs 19% for fusions). Moreover, the asymmetry between the number of combinations and of fusions found in those experiments was replicated here as well as the higher number of combinations with voiceless consonants and of fusions with voiced ones. The vowel /i/ still elicited more illusions for the man speaker, whereas for the woman, the effect was non significant or went in the other direction. As observed in our other studies, item length did not play a very important role (the bisyllables produced more illusions than the monosyllables in only one condition). The

experiments (in which the slow speaking rate elicited the more illusions). In the case of combinations, there was an advantage of the fast speaking rate for the woman speaker, whereas for the man, the slow speaking rate produced more illusions. For fusions in the case of the woman speaker, we found a small advantage of the slow and fast speaking rates, but for the man, there was an advantage of the slow and medial rates.

Hypotheses enabling us to understand the impact of intensity level, vocalic environment, type of consonant, item length and speaking rate on both McGurk illusions have already been drawn and discussed in a previous paper [9]. On the whole, these results can be accounted for in terms of perceptual saliency. The less salient are the auditory stimuli, the more numerous are the illusions. The auditory saliency of our stimuli was lessened when the stimuli were played at 40dB rather than at 70dB and when the vowel /i/ was used instead of /a/ (indeed, /i/ is of lower intrinsic intensity than /a/). Voiceless consonants are auditorily more salient than voiced ones (greater perceptual weight of the burst and longer acoustic closure). This could explain why voiceless consonants produced more combinations than voiced ones (more /pk/ than /bg/ responses) and why less salient voiced consonants could be "more" attracted by the visual component in case of fusions.

The quantitative asymmetry between McGurk fusions and combinations together with the differential effect of type of consonant on both illusions might suggest that different mechanisms underlie the two effects. This hypothesis has been put forward by Diesch [8]. He interpreted the left hemisphere superiority for the combinations in terms of higher load imposed on phonetic coding which calls on the left hemisphere functions. The right hemisphere advantage exhibited for fusions was explained in terms of visuo-spatial analysis which is a right hemisphere specialized process. Indeed, the visual consonants giving rise to fusions are velar and thus visually less salient than those which result in combinations (bilabials). Our results indicate however that both hemispheres would contribute in the same way to the McGurk fusions and combinations. Nevertheless, the fact that we did not use exactly the same methodology than in Diesch' study should be taken into account. For the auditory signal, Diesch used two loudspeakers located on either side of the screen, while we used only one loudspeaker positioned straight ahead. In his study, the speakers faces were presented side by side with a 3,5° eccentricity between the center of their mouth and the participant' straight ahead position. In our case, the screen was moved from left to right (or inversely) after the first half of the experiment and the eccentricity was of 5°. The absence of hemifield effect in the present study could certainly not allow us to reject definitively the assumption that the audiovisual integration processes underlying fusions and combinations are lateralized to the right and to the left hemisphere, respectively. This hypothesis could be further investigated with other methods, such as brain imagery techniques. Brain imagery might enable to shed light on the neural substrates of audiovisual speech and could also provide some information about the temporal course of both McGurk illusions.

REFERENCES

1. Sumby, W.H. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.
2. McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
3. Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions. Four cases studies and their implications. *Neuropsychologia*, 28, 787-802.
4. Campbell, R., & De Haan, E. H. F. (1998). Repetition priming for face speech images: Speech-reading primes face identification. *British Journal of Psychology*, 89, 309-323.
5. Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: Different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, 28, 21-29.
6. Okada, K., Kawashima, R., Fukuda, H., Mori, K., Imaizumi, S., Kiritani, S. & Ogawa, A. (1998). A PET study of the McGurk effect. *NeuroImage*, 7(4), Parts 2 of 3 parts, S163.
7. Baynes, K., Funnell, M. G., & Fowler, C. A. (1994). Hemispheric contributions to the integration of visual and auditory information in speech perception. *Perception and Psychophysics*, 55, 633-641.
8. Diesch, E. (1995). Left and right hemifield advantages of fusions and combinations in audiovisual speech perception. *Quarterly Journal of Experimental Psychology*, 48A, 320-333.
9. Colin, C., Radeau, M. & Deltenre, P. (1998). Intermodal interactions in speech: A French study. *Proceedings of Auditory-Visual Speech Processing*, 55-60.

Acknowledgments: this work has been supported by grants A.R.C. (96/01-203) and F.R.F.C. (8.4519.96). We thank W. Serniclaes and D. Demolin for their fruitful advice as well as F. Goossens and the C.A.V of the Free University of Brussels for their help in editing the stimuli.