

Binding and unbinding in audiovisual speech fusion: *Removing the McGurk effect by an incoherent preceding audiovisual context*

Olha Nahorna, Frédéric Berthommier and Jean-Luc Schwartz

GIPSA-Lab, UMR CNRS 5216 – Grenoble University, France

olha.nahorna, frederic.berthommier, jean-luc.schwartz@gipsa-lab.grenoble-inp.fr

ABSTRACT

The McGurk effect demonstrates the existence of a fusion process in audiovisual speech perception (McGurk & MacDonald, 1976). Recent experiments question the automaticity of this process (Alsus et al., 2005; Tiippana et al., 2004). We consider here the assumption that fusion is controlled by a conditional binding process, which can block fusion in the case of strong audiovisual inconsistencies. To test this hypothesis, we designed two experiments in which a consistent or inconsistent audiovisual context is placed before McGurk stimuli, and we show that inconsistent contexts can remove the McGurk effect.

Keywords: McGurk effect, binding, multisensory fusion, AV speech perception and scene analysis.

1. INTRODUCTION

In the first experiment, a congruent audiovisual “Ba” or incongruent “McGurk” target stimulus was preceded by coherent or incoherent audiovisual contexts (Figure 1). The coherent context consisted of a series of 5, 10, 15 or 20 audiovisual syllables randomly selected within {“pa”, “ta”, “va”, “fa”, “za”, “sa”, “ka”, “ra”, “la”, “ja”, “cha”, “ma”, “na”}. In the incoherent context, the auditory content was the same, but the visual content was replaced by sentences that matched in global duration. Subjects were presented with the sequences (context plus target) in random order, and their task was to detect online “ba” or “da” syllables (syllable monitoring task). 19 French subjects participated in Exp. 1.

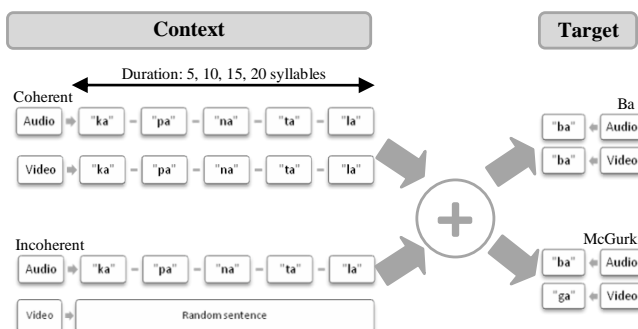


Figure 1 – Experimental paradigm

In the second experiment, a short 200-ms audiovisual “alert” cue was superimposed on the sequences of experiment 1, 280 ms before the target, to enhance the subject’s attention at the period of the McGurk stimulus. 20 French subjects participated to Exp.21.

2. RESULTS AND DISCUSSION

Experiment 1 results are displayed on Fig. 2 (displaying the ratio between “ba” responses and summed “ba” and “da” responses). While the coherent audiovisual context produced the classical McGurk effect (about 60% of “ba” responses vs. 40% of “da” responses, typical ratio for French subjects, see

Cathiard et al., 2001), the incoherent context completely erased the McGurk effect, leading to almost the same number of “ba” responses for “McGurk” targets than for “Ba” targets, that is more than 90%. The effect of context is highly significant ($F(1,18)=35.67, p<0.0001$), essentially due to McGurk stimuli, as shown by the significant interaction between stimulus and context ($F(1,18)=24.14, p<0.0001$). Furthermore, an ANCOVA (qualitative factor “subject”; quantitative factor “number of syllables in the context duration”) applied on responses to McGurk stimuli shows that there is a significant decrease in the percentage of “ba” responses (hence increase in the McGurk effect) with context duration in the coherent context ($F(1,56)=4.37, p<0.05$).

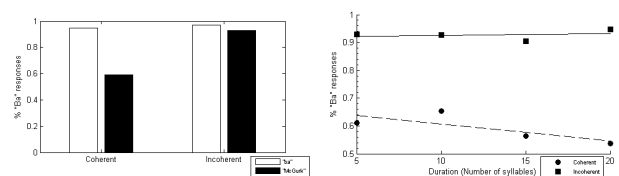


Figure 2 - Results of Experiment 1

Results for Experiment 2 were exactly the same, which shows that the alert cue did not change the previous results. Altogether, these results confirm the existence of a conditional binding process. Complete unbinding (almost complete removal of the McGurk effect) occurs even with the shortest incoherent context of 5 syllables, with a duration of about 4s, while binding seems slow, with a significant increase of the McGurk effect when the duration of the coherent context is increased from 5 to 20 syllables. The alert cue does not allow to re-establish binding in the incoherent context. Further experiments are presently under study to explore this unbinding mechanism, which had never been displayed previously, to our knowledge. They will be presented in the conference.

3. ACKNOWLEDGEMENTS

This work was supported by the French National Research Agency (ANR) through funding of the MULTISTAP project (MULTISTability and binding in Audition and sPeech: ANR-08-BLAN-0167 MULTISTAP)

4. BIBLIOGRAPHIE

- Alsus, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, 15, 839-843.
- Cathiard, M., Schwartz, J., & Abry, C. (2001). Asking a naive question about the McGurk Effect: why does audio [b] give more [d] percepts with visual [g] than with visual [d]? *AVSP-2001*, (pp. 138-142).
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 746-748.
- Tiippana, K., & Andersen, T. S. (2004). Visual attention modulates audiovisual speech perception. *Eur. J. Cogn. Psychol.* 16, 457-472.