

# Word Recognition in Audiovisual Speech: Preliminary Results

**Mathilde Fort<sup>1</sup>, Sonia Kandel<sup>1,2</sup>, Elsa Spinelli<sup>1,2</sup> &  
Christophe Savariaux<sup>3</sup>**

<sup>1</sup> Laboratoire de Psychologie et NeuroCognition (CNRS UMR 5105) – Université  
Pierre Mendès France

<sup>2</sup> Institut Universitaire de France

<sup>3</sup> GIPSA-lab, Dpt. Parole et Cognition (CNRS UMR 5216)

Mathilde.Fort@hotmail.fr;

{Sonia.Kandel;Elsa.Spinelli}@upmf-grenoble.fr;

Christophe.Savariaux@gipsa-lab.inpg.fr

## INTRODUCTION

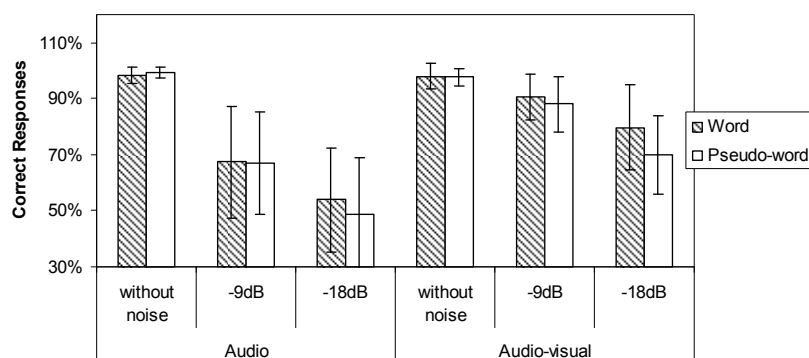
Most of the time, speech perception takes place in an audiovisual environment. Several studies have shown that the visual information on the labial gestures of the speaker enhance phoneme detection in noisy environments (Sumbly & Pollack, 1954 for English; **Benoît**, Mohamadi & Kandel, 1994 for French) as well as in perceptual conflict situations like the McGurk effect (McGurk & MacDonald, 1976). The visual information sometimes appears before the acoustic information, so we can detect a phoneme before having heard it (Noiray, Ménard, Cathiard, Abry, Aubin, & Savariaux, 2006; Munhall & Tokhura, 1998). The goal of our study was to provide evidence that the visual cues on the articulatory gestures of the speaker also activate lexical representations during word recognition. Surprisingly, research in the field of word recognition only studied lexical access in an auditory context (Ganong, 1980 ; Cutler, Mehler, Norris & Segui, 1987 ; Frauenfelder, Segui & Dijkstra, 1990). To our knowledge, only two studies investigated this issue. Both studies used the McGurk effect, which places the individual in a situation of perceptual conflict. Their results are contradictory. Sams, Manninen, Surakka, Helin and Kättö (1998) were unable to show that visual information contributed to lexical access whereas Brancazio (2003) found the opposite pattern of results. In our study, we used a more ecological experimental paradigm. Our participants had to do a word recognition task in silent and noisy environments, which are both situations found in everyday life. We hypothesized that visual information would contribute to word recognition, especially in noisy conditions. We conducted a phoneme detection task with words and pseudo-words. The idea was that visual information would activate lexical representations and that this top down lexical information would facilitate phoneme detection in words with respect to pseudo-words.

## METHOD

Fifty-nine native French speakers participated in the experiment. They all had normal or corrected-to-normal vision and reported no auditory disorders. The stimuli were bi-syllabic words and pseudo-words. They were registered in a sound proof room by a male native French speaker. The participants had to detect a phoneme that always appeared in the second syllable. They had to press the space bar as soon as they perceived the target phoneme. They went through an auditory only (A) and audiovisual (AV) presentations. For 18 participants, the stimuli were presented without noise. For 41 participants the stimuli were presented in with a white noise of -9dB and -18dB S/N ratio.

## RESULTS

The results indicate that when there is no noise, the percentage of correct phoneme detection is equivalent in words and pseudo-words and in both A and AV presentations,  $F_1(1, 17) < 1$  (see Figure 1). In noisy conditions, the scores are higher in the AV than in A conditions, as in Benoît *et al.* (1994),  $F_1(1, 40) = 147.03$   $p < .001$  ;  $F_2(1, 39) = 42.18$ ,  $p < .001$ .



**Figure 1** – Percentage of correct responses in the Audio only and Audiovisual presentations, in the conditions without noise, and with noise at -9 dB and -18 dB SN levels.

The results also yield that the percentage of correct response is higher at -9 dB than at -18 dB,  $F_1(1, 40) = 86.07$ ,  $p < .0001$ ;  $F_2(1, 39) = 43.91$ ,  $p < .001$ . What is more interesting for the purpose of our study is that the scores are significantly higher for words than pseudo-words,  $F_1(1, 40) = 8.82$ ,  $p < .005$ ;  $F_2(1, 39) = 4.11$ ,  $p < .05$ . The significant interaction between the noise level and the lexical status indicates that the contribution of visual information in the lexical access process increases as the noise level increases,  $F_1(1, 39) = 6.09$ ,  $p = .02$ ;  $F_2(1, 39) = 4.28$ ,  $p = .04$ . The word superiority effect is higher at -18 dB (word = 79.8 %, pseudo-word = 70.0 %,  $F_1(1, 40) = 14.59$ ,  $p < .001$ ;  $F_2(1, 39) = 8.79$ ,  $p < .01$ ) than at -9 dB (word = 90.7 %, pseudo-word = 88.0 %,  $F_1(1, 40) = 6.09$ ,  $p = .02$ ;  $F_2(1, 39) = 4.28$ ,  $p = .04$ ).

## DISCUSSION

The aim of this study was to show that the visual information provided by the speaker contributes to the process of lexical access during word recognition. We observed no lexical effect in the situation without noise. In contrast, when the acoustic information was partially masked by noise, the AV condition facilitated phoneme detection. Furthermore, the scores were higher for words than pseudo-words. This suggests that in noisy situations, phoneme detection is not only enhanced by visual information—as Benoît *et al.* (1994) have shown—but that this information contributes to the process of lexical access during word recognition. These results are in line with Brancazio's (2004) study. Further research must be done to investigate the time course of this process, and in particular, whether lexical access takes place before or after the integration of auditory and visual information.

## BIBLIOGRAPHY

- Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of speech. *Journal of Speech and Hearing Research*, 37, 1195-1203.
- Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 445-463.
- Noiray, A., Ménard, L., Cathiard, M.-A., Abry, C., Aubin, J., & Savariaux, C. (2006). Extending the Movement Expansion Model (MEM) for rounding from French to English. *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brazil.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141-177.
- Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception & Performance*, 16(1), 77-91.
- Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, 6 (1), 110-125.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Munhall, K.G., & Tokhura, Y. (1998). Audiovisual gating and the time course of speech perception. *Journal of the Acoustical Society of America*, 104, 530-539.
- Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). Mc Gurk effect in Finnish syllables, isolated words and words in sentence: Effects of word meaning and sentence context. *Speech Communication*, 26, 75-87.
- Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.