Do speakers make use of the visual channel to improve their intelligibility in adverse conditions? A pilot study.

Maëva Garnier¹, Lucie Ménard², Gabrielle Richard²

¹ Speech and Cognition Department, GIPSA-Lab, UMR CNRS 5216 & Grenoble Universités, France
² Laboratoire de phonétique, Université du Québec à Montréal, Canada

maeva.garnier@gipsa-lab.grenoble-inp.fr, menard.lucie@uqam.ca

1. Introduction

On one hand, it is now well known that seeing speech improves its perception, especially when speech is degraded by a noisy background [1]. On the other hand, some studies have shown that speakers adapt their speech production in noisy conditions. This adaptation, also called the « Lombard effect », mainly consists in talking louder and at higher pitch [2]. However, it is also accompanied by other speech modifications, such as increased amplitude and speed of lip articulation [3]. This raises the question of whether this hyper-articulation observed in Lombard speech can be considered as a communicative strategy to improve visual intelligibility.

This study aims at bringing elements of answers, by examining whether, in noise:
- speakers enhance significantly more their visible articulatory movements when their speech partner can see them compared to when the partner can only hear them.
- all the articulatory movements are enhanced similarly, or if the most visible ones (lips, jaw) are more enhanced than the others (tongue).

2. Material and Method

A French Canadian speaker was recorded while speaking in a quiet environment and in a cocktail-party noise of 85 dB played over loudspeakers. Three conditions of interaction were examined: (S1) No Interaction: The speaker read sentences aloud. (S2) Audio Only (AO): The speaker gave instructions to the experimenter who was standing at a writing board placed 2m in front of him and the back to him. (S3) Audio Visual (AV): The experimenter was standing at the same place as before, this time facing the speaker. Seven target-words were selected: /pap/, /pip/, /pup/, /pɛp/, /pəp/, /map/, /tæp/, /næp/. They were produced in the carrying sentence « le mot ___ me plait » (I like the word ___) and repeated ten times in each condition. In the two interactive conditions (S2 and S3), the speaker chose freely the order of production of the 70 sentences, so that the experimenter could not predict the target-word. The audio signal was recorded synchronously with the 3D movements of the lips, the jaw and the tongue, using electromagnetic articulography (Carstens AG 200).

3. Discussion

The results confirmed that all the speech modifications from a quiet to a noisy situation (increase of voice intensity and fundamental frequency, amplified movements of the lips and the tongue) are significantly greater when the speaker interact with a speech partner (S2 and S3), compared to when he only reads sentences aloud (S1).

As expected, acoustic modifications measured from a quiet to a noisy situation (increase of voice intensity, fundamental frequency and first formant frequency) were found to be significantly greater in a condition of AO interaction relative to a condition of AV interaction. However, contrary to our hypothesis, articulatory modifications were not found to be greater in AV interaction. Very visible movements such as lip aperture, spreading, closure and protrusion, and jaw aperture had their amplitude and speed enhanced in noise. Nevertheless, this enhancement was significantly greater in AO interaction than in AV interaction.

Less visible tongue movements were also significantly modified in noise, but only in the situation of AO interaction. For all vowels, the tongue was lower and more forward in the noisy condition compared to the quiet one. Tongue displacements were significantly amplified, with increased speed for the tip of the tongue and reduced speed at its root.

4. Conclusions

The results obtained from this speaker do not support the hypothesis that speakers modulate their production of visible cues in adaptation to the perceptual modalities of interaction. Instead, these results support the idea that all articulatory movements, regardless of their visibility, are enhanced similarly when speaking in noisy conditions and that this enhancement is primarily related to the increase of intensity. To compensate for the perturbation of intelligibility – which is greater in AO interaction than in AV interaction–, increasing voice intensity appears to be the main strategy. As a finer strategy, speakers do not seem to pay on the visual channel to improve their intelligibility. The investigation of five other speakers will enable us to determine if these results can be widespread.

5. References