

Vision of Tongue in Augmented Speech: Contribution to Speech Comprehension and Visual Tracking Strategies

**Pierre Badin, Frédéric Elisei, Lin Huang, Yuliya Tarabalka &
Gérard Bailly**

GIPSA-lab, BP 46 38402 Saint Martin d'Hères
France

`Pierre.Badin@gipsa-lab.inpg.fr`

A large number of studies has established that the vision of typically visible articulators (lips, jaw, face, tongue tip, teeth) eases speech understanding by humans, and significantly increases the detection and identification performance of words in noise. However, everything cannot necessarily be “read” without ambiguity from the mere vision of face. For instance, the lip shapes for vowels [u] and [y] are nearly identical whereas the associated tongue shapes are very different.

In a recent study (Tarabalka, Badin, Elisei & Bailly, 2007), we have attempted to determine if direct and full vision of the tongue can enhance speech comprehension, based on the augmented-speech capabilities of our talking head (in this study, a cutaway profile view is used). Using a motion capture paradigm based on ElectroMagnetic Articulography, we have elaborated a set of audiovisual Vowel-Consonant-Vowel stimuli. These stimuli have been presented to a group of listeners in a series of audiovisual perception experiments using various presentation conditions: (1) audio signal + cutaway view along the sagittal plane without tongue, (2) audio signal + cutaway view with tongue, (3) audio signal + complete face with skin texture. Each condition was played at four different Signal to Noise Ratios (SNRs) of white noise added to the sound. The analysis of the results has shown some implicit learning effects of tongue reading, a preference for a more ecological rendering of the complete face compared to a cutaway presentation, a predominance of lip reading over tongue reading (except for cases where – the audio signal being so much degraded or absent – tongue reading is taking over). Though the subjects are objectively provided with more articulatory or phonemic information when the tongue is displayed simultaneously with the lips, the strategies of visual tracking of both tongue and lips may not be natural or easy. Therefore, we have replicated the previous experiment, adding presentation conditions where lips are not displayed. During the perception test, we use an eye tracking system to determine the trajectories of eye gaze of the subjects when they are watching the talking head. We will analyse the visual tracking strategies and attempt to establish correlations between the degree of success in tongue reading and these visual strategies.

REFERENCES

- Tarabalka, Y., Badin, P., Elisei, F. & Bailly, G., 2007. Can you “read tongue movements”? Evaluation of the contribution of tongue display to speech understanding. In *1^{ère} Conférence internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situation de handicaps (ASSISTH'2007)* (N. Vigouroux, P. Gorce & J.-L. Nespoulous, editors), vol., pp. 187-193. Toulouse, France, Editions Cepaduès, Toulouse.