Compensatory responses to real-time perturbation of visual feedback during vowel production

Camille Vidou[1], Cristina Uribe[1], Tarik Boukhalfi[2], David Labbé[2] and Lucie Ménard[1]
[1] Phonetics Laboratory, Université du Québec à Montréal, Montréal, CANADA
[2] École de technologie supérieure, Montréal, CANADA

The current understanding of speech production and perception assumes that multisensory (auditory, visual, and proprioceptive) information affects the actions of orofacial articulators, and that these relationships are established early in life. Many studies have examined the role of visual cues in speech *perception* (Sumby and Pollack, 1954; McGurk and MacDonald, 1976; Sams *et al.*, 2005), but less is known about their role in speech *production*. However, several recent studies have provided indirect evidence of the role of visual cues in speech production, by examining kinematic and acoustic features of speech produced by congenitally blind adults. Those studies suggest that sighted adults and children produce expanded acoustic and articulatory vowel spaces as a result of larger lips movements, which are likely driven by vision (Ménard *et al.*, 2009; 2013). Direct evidence of visual influence on speech production in adults with unimpaired vision is rare. Speech convergence experiments have shown that individuals tend to align their produced speech to match those of a speaker, when auditory or visual stimuli are presented (Miller *et al.*, 2010;). Similarly, Gentilucci and Bernardis (2007) investigated how lip kinematics and voice spectra of participants uttering phoneme strings are affected by seeing a face or hearing a voice utter the same phonemes and they found that the participants' lip closures and lip apertures changed to match those of the speakers that they saw. Taken together, these findings suggest that there is a direct link between speech production (lip movement) and vision. However, it is not clear how much speakers rely on vision to control speech production. The current study aimed to examine the effect of visually perceived self-produced lip movements on the control of vowel production, using a real-time self-avatar.

The method was based on the real-time sensory perturbation paradigm (Houde and Jordan, 1998). A total of twenty-four French speaking adults (15 males) were recruited. They had unimpaired audition and vision and no history of any speech or language disorder. An animated virtual avatar and his environment were developed in Unity 3D and presented using an Oculus Rift CV1 virtual reality head-mounted display (HMD). An Optotrak Certus 3020 was used to track head and lips movements of the participants using 7 active markers (IREDs; see Figure 1, left panel): three on the virtual reality HMD to track head movement, two at the centers of the upper and lower lips and two on the lips corners. The animation was captured and mapped on the avatar in real-time (see Figure 1, right panel). Speakers were thus animating the avatar using their own lips and head movements. The task consisted of 55 repetitions of each of the three isolated vowels /i/, /a/, and /u/. For each vowel, four conditions were elicited as follows. During the baseline, 10 repetitions were produced without any perturbation. In the ramp condition, 20 repetitions were produced while gradually scaling down the lips movements produced by the avatar. For /a/, the vertical interlip distance was scaled down (perturbing the visual perception of lip opening) ; for /u/, the vertical and horizontal interlip distance was scaled up (perturbing the visual perception of lip rounding) and for /i/, the horizontal interlip distance was scaled down (perturbing the visual perception of lip retraction). The maximal perturbation was 50% of the total movement. In the hold

phase, 15 repetitions of the vowel were elicited with the maximal perturbation applied. Finally, in the postperturbation phase, the perturbation was removed and 10 repetitions of the vowel were produced with normal visual feedback. The x, y, and z spatial coordinates for each IRED were extracted using Matlab at vowel midpoints. The compensatory responses were compared across vowels, conditions, and speakers using linear mixed effects modelling.

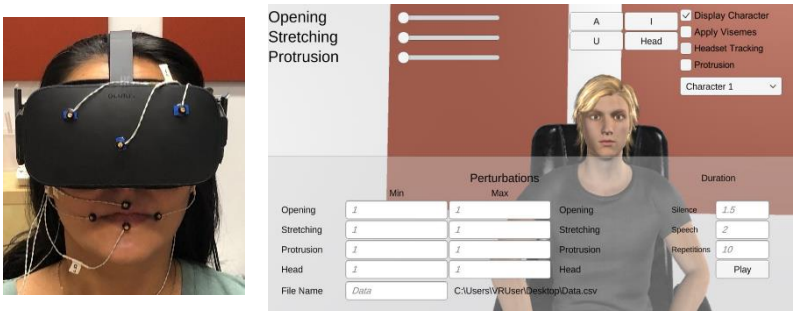

Figure 1: IREDs placement with the Oculus Rift HMD (left panel) and animated avatar with superimposed controlled parameters (right panel).

Results show that the speakers had various compensatory responses depending on the vowel they uttered. Scaling down the lip opening for /a/ significantly affected the speakers' produced lip opening. Similarly, perturbing the visual display of lip rounding yielded increased production of lips rounding. The results for /i/ varied across individuals. These findings suggest that self-produced lips movements have visual consequences that guide speech production. We discuss the links between visually perceived actions on orofacial articulators and self-produced actions, in light of multimodal theories of speech perception and production.

REFERENCES:

Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologica*, *45*(3): 608–15.

Houde, J. F. and Jordan, M. (1998). Sensorimotor adaptation in speech production, *Science*, 279 (5354), 1213-6.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America*, *126*, 1406–1414.

Ménard, L., Toupin, C., Baum, S., Drouin, S., Aubin, J., & Tiede, M. (2013). Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America*, 134 (4), 2975-2987.

Miller, R. M., Sanchez, K. and Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72 (6), 1614-1625.

Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*, 429–435.

Sumby, H. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.