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

Camille Vidou, Cristina Uribe, Tarik Boukhalfi, David Labbé, Lucie Ménard
Laboratoire de phonétique, Université du Québec à Montréal, Canada
vidou.camille@gmail.com

BACKGROUND

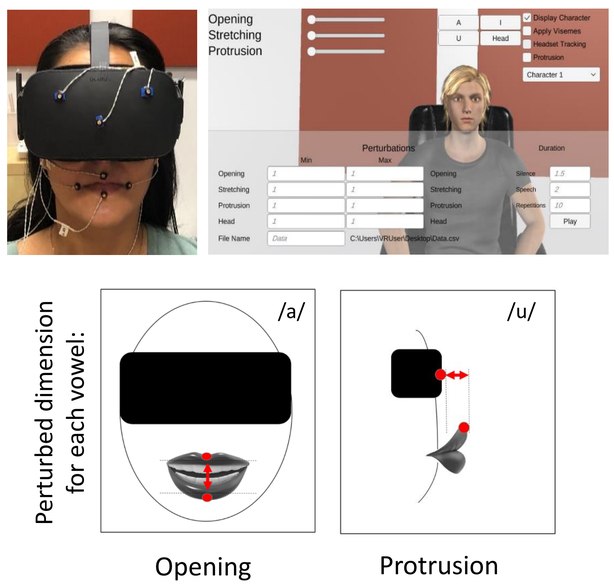
Speech perception involves auditory and visual cues (Sumby and Pollack, 1954; McGurk and MacDonald, 1976; Sams *et al.*, 2005);

At the production level, produced lip shapes tend to match those seen on others (Miller *et al.*, 2010; Gentilucci and Bernardis, 2007);

Seeing one's own lip movements affects the lip movements produced (Sams *et al.*, 2005);

OBJECTIVE

Our objective was to examine the effect of visually perceived self-produced lip movements on the control of the vowels /i/, /a/ and /u/, using a real-time self-avatar.



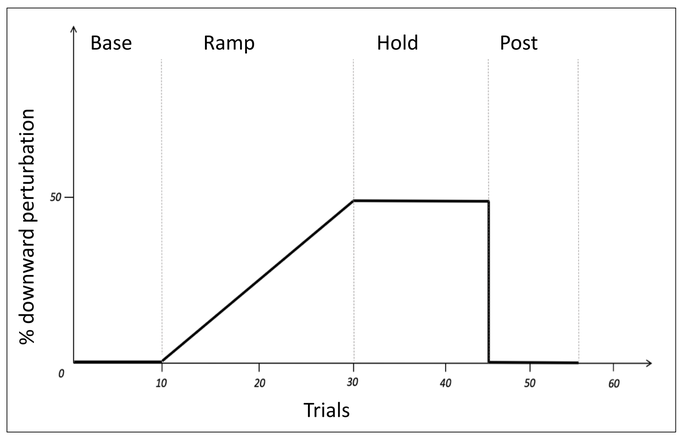
METHOD

Participants
-24 adults (9 females, 15 males)
-All participants had normal auditory thresholds, unimpaired vision, and were native speakers of Canadian French.

Experimental setup
-A virtual animated avatar was developed in Unity 3D and presented with an Oculus Rift© virtual reality head-mounted display (HMD).
-An Optotrak Certus 3020 was used to track head and lips movements of the participants and animate their avatar in real-time.
-Lip opening (for /a/), lip protrusion (for /u/) and lip stretching (for /i/) were gradually scaled down.

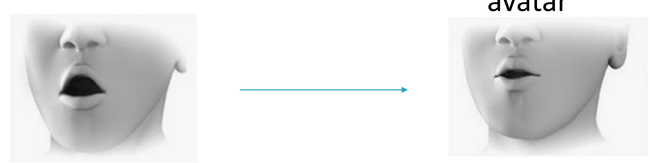
Data analysis (results of /a/ and /u/ are reported here)
-For each participant, vowel and trial, ratio of lip opening and lip protrusion relative to the baseline;
-3 groups of participants depending on the values of the ratios in the Hold phase: compensators (ratios > 1.05), resistants (0.95 < ratios < 1.05) and followers (ratios < 0.95);
-Linear mixed effects models built ratios, with group (compensators, resistants, followers), phase (base, ramp, hold, psot), and vowel (/i/, /a/, /u/) as fixed effects and speaker as a random effect.

Task
Real-time feedback perturbation paradigm:



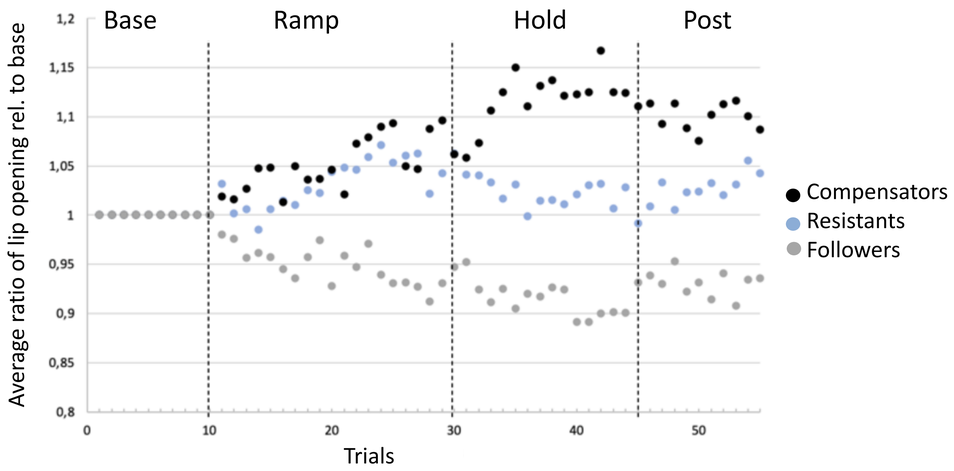
-Example of visual feedback perturbation: the tracked lip positions are modified such that lip aperture is reduced by 50%. Those values are used to animate the avatar in real-time.

Recorded → Animated avatar



RESULTS

Lip opening /a/



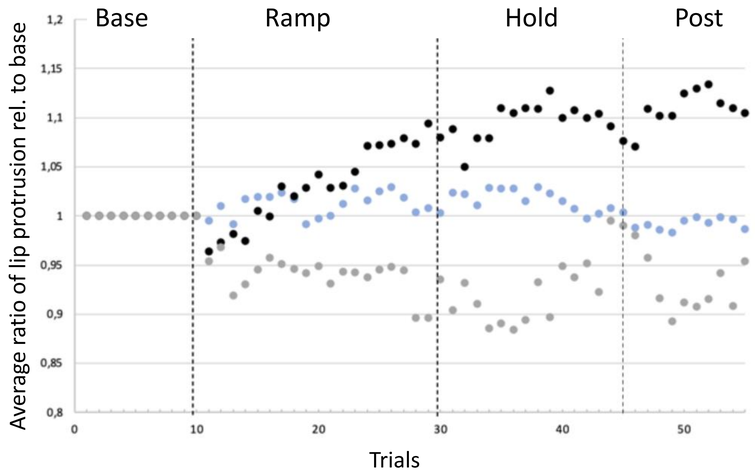
Results summary for lip opening in /a/:

Compensators (n=9): **Base < Ramp; **Ramp < Hold; Hold = Post

Resistants (n=8): **Base < Ramp; Ramp = Hold; Hold = Post = Base

Followers (n=6): **Base > Ramp; *Ramp > Hold; Hold = Post

Lip protrusion /u/



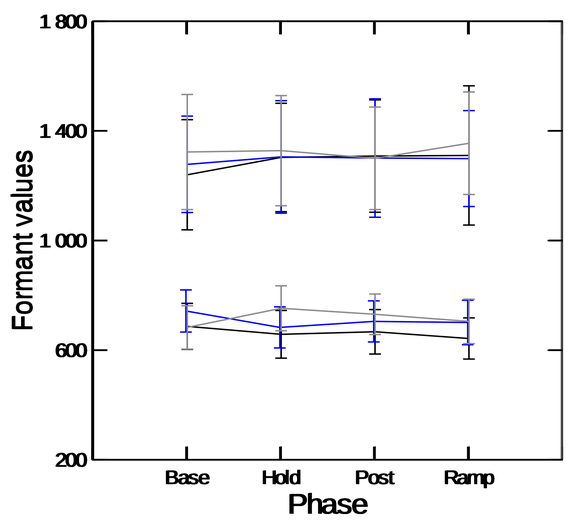
Results summary for lip protrusion in /u/:

Compensators (n=6): Base = Ramp; **Ramp < Hold; Hold = Post

Resistants (n=12): Base = Ramp = Hold = Post

Followers (n=5): **Base > Ramp; Ramp = Hold; Hold = Post

No difference in formant values across groups throughout the trials (ex. of /a/):



DISCUSSION

-When perturbing in real-time the visual feedback of speaker's lips, compensatory responses are produced in some participants. Some other participants produce lip movements that mirror those that they observed (followers), which is likely an effect of convergence. Finally, some participants do not react to the visual perturbation (resistants). Speaker-specific and vowel-specific weight of the visual channel?
-Since acoustical differences are not found in the produced vowels throughout the 55 trials, compensators and followers react to the discrepancy created by the perturbation between the expected visual consequences of their articulatory positions and the seen visual consequences.
-Despite the fact that, for the orofacial articulators, visual feedback is not a closed loop, speakers have mapped the links between articulatory positions and sensory (here, visual) consequences.
-Further analyses are underway and follow-up experiments will be done after COVID!

REFERENCES

Sumby and Pollack (1954): Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, *26*(2), 212-215. https://doi.org/10.1121/1.1907309
McGurk and MacDonald (1976): Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. https://doi.org/10.1038/264746a0
Sams, M., Möttönen, R., & Sihvonen, T. (2005): Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*(2-3), 429-435. https://doi.org/10.1016/j.cogbrainres.2004.11.006
Miller, R. M., Sanchez, K. and Rosenblum, L. D. (2010): Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72 (6), 1614-1625.
Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologica*, *45*(3): 608–15.

*This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC).