Speech airflow outside the mouth

Donald Derrick[1], Mark Jermy[2], Bryan Gick[3]

[1]University of Canterbury, New Zealand Institute of Language, Brain & Behaviour, [2]University of Canterbury, Mechanical Engineering Department, [3]University of British Columbia, Department of Linguistics

During speech, people produce sound, facial motion, and airflow via their mouths and noses. We can normally understand speech from audio information alone, but particularly in noise, visual[1,2] and aero-tactile[3,4] information can also affect speech perception. Airflow's perceptual influence is small compared to visual speech[5] and may only affect classification of word-initial consonants[6]. This is likely because speech airflow travels slowly, and dissipates quickly as it moves from the speaker, unlike audible or visible information[7]. However, we do not know how speech airflow changes as it moves from the mouth. Here we use schlieren imaging to show speech airflow as it moves up to 30 cm away from a speaker's mouth.

We collected speech airflow from 13 native English-speaking participants. We used a single mirror schlieren system with a 400 mm diameter parabolic mirror, a knife edge on a linear stage with a micrometer adjustment, and Photron SA5 camera (1024x1024 px, 250 frames per second). The field of view of the schlieren images was 400 mm and included minimally the participant's lips. We simultaneously recorded audio and high-speed video. The setup can be seen in Fig. 1. Participants were seated, given shade 5 welding goggles for eye protection, and a head-mounted microphone for audio recording. Participants were given 3 blocks of 4 phrases to read, and due to computer capture times, each participant took 1.5 hours to record.

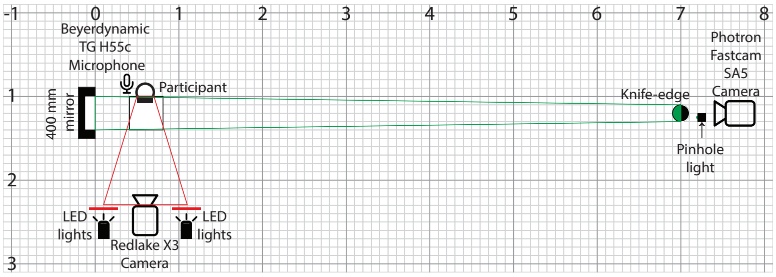

Fig. 1: Schematic of experimental setup. 1 meter scale [17]

Audio recordings were labelled and transcribed in PRAAT[8]. Transcriptions were based on Wells' Lexical sets[9]. Audio files and schlieren images were visually aligned within 1 frame using alignment of auditory release bursts and airflow at the mouth.

We grayscaled and balanced video data using zsh, ffmpeg[10], and imagemagick[11]. We high pass filtered data to remove body heat from airflow using R[12] and the signal[13] package. We used OpenOpticalFlow[14] in MATLAB to convert schlieren air density into airflow velocity. The velocity images were aligned using a procrustean fit based on duration of phrase audio. We used MGCV[15] in R[12] for generalized additive mixed effects modelling to show regions of significant differences in speech airflow velocity at set distances from the lips at 5, 10, 15, 20, 25 and 30 cm from the lips; here we show air flow velocity at 20 cm distance in Fig 2.

The Y axis shows height along an arc of points equidistant from the lips. The X axis shows average times for the phrase "The beige hue on the waters of the loch impressed all". Topographical colors show high velocity (blue) and low velocity (orange) airflow regions. Black bands represent standard errors in airflow velocity.

Results show that speech airflow from some nasals, stops and fricatives reach 20 cm past the lips in running speech, with flow from intense fricatives and aspirated stops reaching as far as 30 and 35 cm.

This data provides detailed information on the range of airflow outside the mouth. We are using this data to test models of multimodal sound categories in phonetics-phonology. It can also be used to reproduce artificial airflow from speech for use in improved behavioural and brain response research. It also provides information that can be mapped to skin sense response to speech airflow for the three major skin mechanoreceptors that detect touch based on estimation of airflow's effects on skin intendation[16].

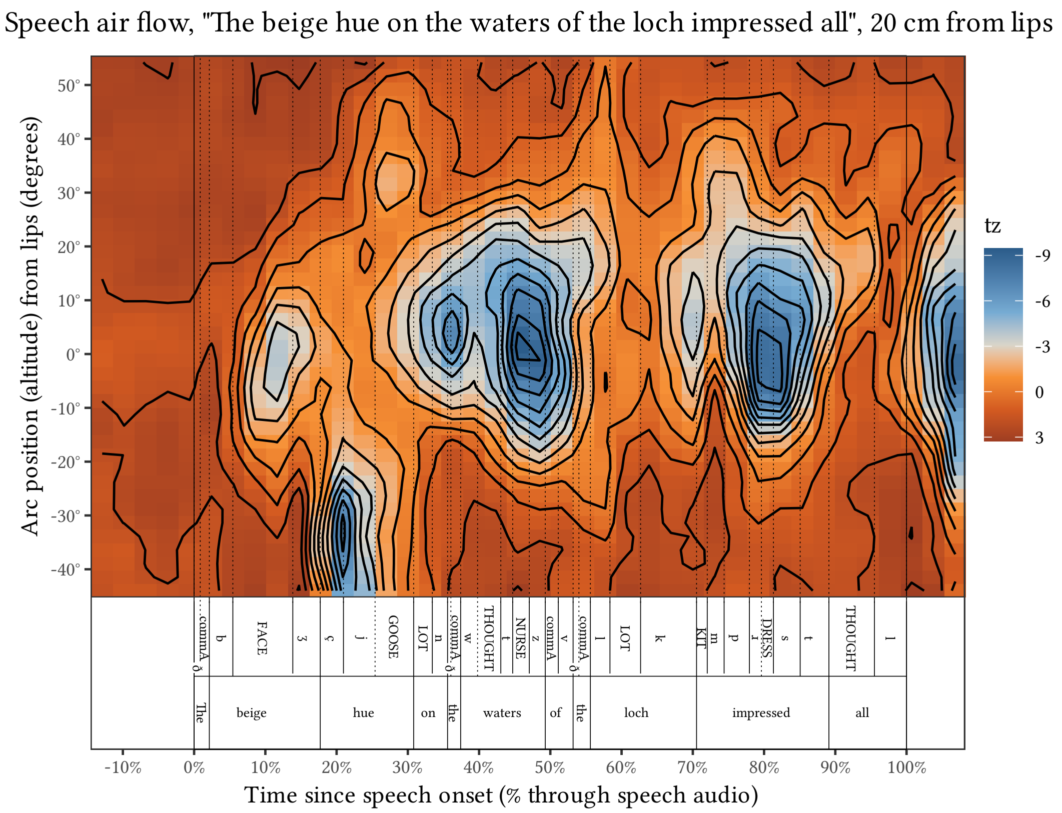

Fig. 2: Generalized additive mixed-effects model topographical map of speech airflow, each line represents approximately 2 standard deviations of airflow velocity difference. Lower number = faster non-dimensional speech air flow.

References

[1] Sumby, W. H. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America,* 26 (1), 212–215. https://doi.org/https://doi.org/10.1121/1.1907309

[2] McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748 https://doi.org/https://doi.org/10.1038/264746a0

[3] Gick, B. & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504. https://doi.org/10.1038/nature08572

[4] Derrick, D. & Gick, B. (2013). Aerotactile integration from distal skin stimuli. *Multisensory Research,* 26, 405–416. https://doi.org/https://doi. org/10.1163/22134808-00002427

[5] Derrick, D., Hansmann, D. & Theys, C. (2019). Tri-modal speech: Audio-visual- tactile integration in speech perception. *Journal of the Acoustical Society of America,* 146 (5), 3495–3504. https://doi.org/https://doi.org/ 10.1121/1.5134064

[6] Derrick, D., Madappallimattam, J. & Theys, C. (2019). Aero-tactile integration during speech perception: Effect of response and stimulus characteristics on syllable identification. *Journal of the Acoustical Society of America,* 146 (3), 1605–1614. https://doi.org/https://doi.org/10.1121/1. 5125131

[7] Derrick, D., Anderson, P., Gick, B. & Green, S. (2009). Characteristics of air puffs produced in English 'pa': Experiments and simulations. *Journal of the Acoustical Society of America,* 125 (4), 2272–2281. https: //doi.org/https://doi.org/10.1121/1.3081496

[8] Boersma, P. & Weenink, D. (2019). *Praat: doing phonetics by computer* [computer program]. Version 6.0.52.

[9] Wells, J. C. (1982). *Accents of English.* Vol. 1: An Introduction (pp. i–xx, 1–278), Vol. 2: The British Isles (pp. i–xx, 279–466), Vol. 3: Beyond the British Isles (pp. i–xx, 467–674). Cambridge University Press.

[10] FFmpeg Developers (2023). *ffmpeg tool* [sofware]. [software] http://ffmpeg.org/

[11] The ImageMagick Development Team (2023). *Imagemagick* [software]. https://imagemagick.org

[12] R Development Core Team. (2023). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria (2018). ISBN 3-900051-07-0.

[13] Signal developers. (2014). *signal: Signal processing* [software] URL http://r-forge. r-project.org/projects/signal/.

[14] Liu, T. (2017). OpenOpticalFlow: An open source program for extraction of velocity fields from flow visualization images. *Journal of Open Research Software*, 1(5):1-29.

[15] S. N. Wood. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association,* 99(467):673-686.

[16] Ouyang, Q. and Wu, J. and Shao, Z., and Chen, D., and Bisley, J. W. (2021). A Simplified Model for Simulating Population Responses of Tactile Afferents and Receptors in the Skin. *IEEE Transactions on Biomedical Engineering*, 68(2):556-567.

[17] Derrick, D., Kabaliuk, N., Longworth, L., Pishyar-Dehkordi, P., and Jermy, M. (2022). Speech air flow with and without face masks. *Scientific Reports*, 12(837):1–10. https://doi.org/10.1038/s41598-021-04745-z