**Head Movement and F0 Variation in Blind and Sighted Speakers' Speech**

Yadong Liu, Arian Shamei, Una Chow, Gina Pineda, Alexander Angsongna, Gillian De Boer, and Bryan Gick

Previous work has shown that head movement correlates with speaker fundamental frequency (F0)[1]. The present study tests whether this F0-related head movement (F0HM) serves a strictly physiological function in the control of F0[2], or whether speakers make use of it as a deliberate visual aspect of the linguistic signal. To investigate this, we compare F0HM in congenitally blind and sighted speakers. Previous studies indicate that sighted speakers may intentionally amplify physiological movements to convey linguistic information visually; for example, blind speakers produce less lip protrusion for rounding than sighted speakers [3]. If congenitally blind speakers employ F0HM identically to sighted speakers, this would suggest that head movement serves some physiological function in speech production. If sighted speakers exhibit larger F0-related head movements, this would suggest that sighted speakers use these visible movements as a deliberate linguistic cue.

**Methods.** Naturalistic speech from audio-visual interviews of two adult English speaker groups were downloaded from YouTube: blind from birth (BB n=8, 14334 frames) and sighted (SS; n=10, 8892 frames). These videos were downloaded in 720p with a standardized frame rate (30 frames/second). Audio from YouTube videos were saved to a sound file at 44.1 kHz in a 16-bit mono channel. Vertical head movement measurements (in millimetres: mm) were extracted from each video using facial point tracking in OpenFace 2.0 [4], focused on the nose of each speaker. The speakers' F0 values were extracted from each sound file using prosody analysis functions in Prosogram [5] and Praat [6]. F0 values were converted to semitones (ST) with a reference of 100 Hz. (For voiceless frames, the F0 was interpolated across voiceless frames of continuous speech with a duration below 250 milliseconds). Head movement values and their corresponding F0 values were normalized by calculating the difference between the value at each point and the value from the initial frame. A Pearson product-moment correlation test was used to evaluate the relationship between F0 and head movement. To quantify the degree of head movement corresponding to F0 change by speaker group, a ratio was calculated at each frame for head movement (in mm) and the corresponding F0 change (in ST), in this case, head movement values and their corresponding F0 values were normalized by subtracting the value at each frame from the values of the preceding frame.

**Results.** The mean (M) and standard deviation (SD) of the frame-to-frame ST change for the two groups were equivalent: BB ($M = 1.02$, $SD=1.64$) and SS ($M = 1.07$, $SD=1.72$ $p=0.016$). However, significantly more head movement is observed from SS ($M = 1.17$, $SD=2.1$) compared to BB ($M = 0.61$, $SD = 1.14$ , $p<0.001$). Figure 1 displays scatterplots with the correlation between normalized vertical head movement and F0 values for the data collected to date. Pearson's product-moment correlation tests reveal a weak but significant ($p < 0.001$) positive correlation between these two variables for both speaker groups, with BB showing a stronger correlation ($r=0.12$, $df=7$) than SS ($r=0.05$, $df=9$). The ratio of head movement per ST change (BB: 9.0mm/ST; SS: 18.6mm/ST) suggests that SS produce twice as much head movement than BB for each corresponding change in ST.

**Discussion.** These results support the view that at least some head movement serves a physiological function in F0 production for all speaker groups. Even without access to visual cues, significant positive correlations between head movements and pitch variation are observed in blind speakers. In addition, for sighted speakers, increased head movement for each

corresponding change in F0 suggests that head movement also serves as a visual speech cue. This latter observation may also be reflected in the weaker correlation value.

**Key words.** blind speaker speech, fundamental frequency, head movement, prosody, visual cues

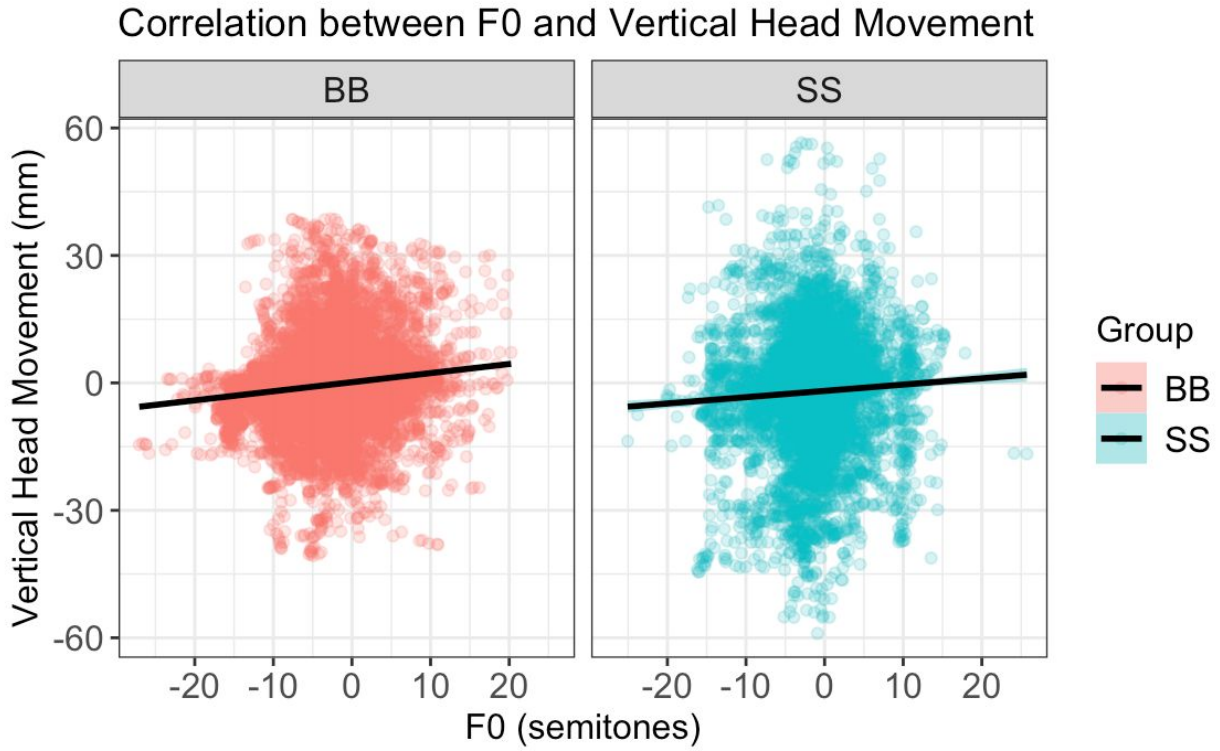


**Figure 1.** Scatterplots with correlation between normalized vertical head movement (mm) and F0 (semitones) for the blind and sighted speaker groups.

## References

[1] Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological science*, *15*(2), 133-137.

[2] Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, *30*(3), 555-568.

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

[4] Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 59-66). IEEE.

[5] Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004, Nara (Japan)* (pp. 549-552).

[6] Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer* (Version 5.3.24).