COORDINATION OF EYEBROW MOVEMENT WITH SPEECH ACOUSTICS AND HEAD MOVEMENT

Kevin D. Roon[1,2], Mark K. Tiede[2], Katherine M. Dawson[1], D. H. Whalen[1,2,3]

[1]CUNY Graduate Center, [2]Haskins Laboratories, [3]Yale University
kroon@gc.cuny.edu, tiede@haskins.yale.edu, kdawson2@gradcenter.cuny.edu, dwhalen@gc.cuny.edu

ABSTRACT

Studies on the relationship between eyebrow movement and other aspects of speech production have focused on large, discrete movements of the eyebrows. Using integrated optical and electromagnetic point tracking, we measured eyebrow movements relative to the skull with a high level of precision. These data in combination with a correlational analysis method that accommodates varying phasing between the signals enabled the investigation of the relationship between continuous eyebrow movements, speech acoustics, and head movements. Our results show that there was a correlation between eyebrow movements and speech acoustics, though there was notable variation within and across participants. The relationship between eyebrow and head movements was much closer, with a strong correlation between eyebrow movement and subsequent head movement, which held across participants. We discuss the implications for theories of gestural control in speech production.

Keywords: speech production, eyebrow movement, head movement, gestural control.

1. INTRODUCTION

Vocal tract gestures are not the only movements that occur during speech production. Orofacial movements have been shown to be linguistically relevant, and are highly correlated with acoustics and vocal tract movements [17, 19]. These movements also play a role in speech perception, evidenced by the fact that seeing videos of only the face inferior to the eyes can aid in identifying segments [11, 15]. Facial movements superior to the eyes have also been shown to play a role in speech production and perception. Eyebrow movements have been correlated with fundamental frequency (F0) [2, 5], phrasal stress [9, 14, 17], and turn-taking [10]. Seeing the entirety of the forehead aids perception [15], providing cues to prosodic prominence [6, 10, 11, 14], "expressiveness" [3], and emotional content [11].

In addition to facial skin movements, head movement has been shown to correlate with F0 [19] and specifically with lexical stress [9]. Head movements also contribute to speech perception [4]. Head and forehead skin movements make independent contributions to perception, with syllables most accurately identified when noise-masked natural speech was accompanied by computer-generated natural head and face movements [13]. Identification is worse when face movements are included but head movements are not, but acoustics accompanied by face movements without head movements are more easily perceived than acoustics only.

Eyebrow movements have been characterized as being either discrete or continuous [10, 18], but the majority of eyebrow studies have focused on discrete movements [2, 6, 9, 10, 14, 18]. Continuous movements of the face inferior to the eyes have been studied to some extent [17, 19], but the relationship between continuous eyebrows movements and other movements during speech production is unexplored. There is debate on whether the underlying control mechanism(s) that generate vocal-tract, head, and facial movements are the same, and if not, how they might be coordinated [10, 17, 18]. The relationship between continuous eyebrow movements and other movements during speech production is an important and missing piece in this puzzle.

The experimental setup of the present study allowed for precise measurement of continuous and discrete eyebrow movements relative to the skull, making it possible to investigate the relationship between eyebrow movements and both speech acoustics and head (i.e., skull) movements.

2. EXPERIMENT

2.1. Participants

Data were recorded from 8 participants (4 female), ranging in age from 22–50 years. Six were native speakers of American English, one of Southern British English (P1F), five of American English (P2M, P3F, P5F, P6M, P8M). P4F was a native speaker of Dutch and P7M was a native speaker of Hindi, but both have professional-level English. No participant reported linguistic or neuromuscular disorders.

2.2. Data collection

3D movement data were collected simultaneously using an integrated optical camera (Optotrak Certus) and a WAVE electromagnetic articulography (EMA, [7]) system (both by NDI, Northern Digital Inc.), both at a sampling rate of 100 Hz. Movement data recorded from both systems were registered within a common coordinate system by NDI-provided data-capture software. Ten infrared emitting diodes (IREDs) were affixed to each participant's forehead skin using double-sided tape, with no special preparation of the skin. One IRED was placed at the nasion. The other nine were affixed as three rows of three, with the most inferior row having one IRED just above each eyebrow, most superior row as close as possible to the hairline, and one row in between. IREDs were tracked by the Optotrak camera, placed ~2.13 m away. Skull position was tracked with an EMA sensor either affixed to the gingiva just above the upper incisors (UI) using dental glue (P1F–P4F) or embedded in a dental mold over the UI (P5F–P8M). Concurrent audio was recorded at 44.1 kHz.

2.3. Procedure

Data for this study came from recordings of short fragments of spontaneous conversation between the participant and the first author. Each had a fixed duration of 7 or 10 s, depending on participant. The number of trials varied by participant, from one (P2M) to six (P8M), for a total of 28 trials.

2.4. Data processing

Rare missing frames of EMA data (skull movement) due to machine error during data collection were filled in using local spline interpolation. EMA data were low-pass Butterworth filtered at a cut-off of 5 Hz to compensate for noise during signal-capture. There was very little noise in the Optotrak signal capture, so Optotrak data were not filtered. The 3D Euclidian distance from each IRED to the UI EMA sensor was calculated at each time frame.

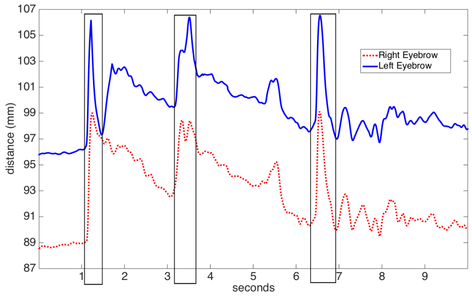
2.5. Analyses

Data from 9 trials were discarded due to lack of sufficient movement to be analyzed meaningfully, determined by an exclusion criterion of the standard deviation (SD) of the distance from the UI to each eyebrow IRED being ≤ 0.5 mm across the trial. We determined whether and to what degree the eyebrow movement data from our participants could be accurately characterized as discrete vs. continuous, and whether any discrete movements corresponded to turn-taking and/or prominence as found in previous studies. We then used correlation map analysis [1] to investigate the relationship between eyebrow movements and speech acoustics, and then between eyebrow movements and head movement.

2.5.1. Discrete movements

Discrete movements were identified by visual inspection of the distance data, but in order for a movement to be considered discrete, its maximum distance had to be greater than two SDs away from the median distance (nearer or further) of that trial. Twenty-five discrete movements were identified. The onset of discrete movements was defined as the distance minimum immediately preceding the peak in cases of raising, or the preceding maximum in cases of lowering. Fig. 1 shows distance data for one trial that contains 3 discrete movements, all raises, which are indicated in boxes. The left edge of each box is aligned with the onset of discrete movement.

Figure 1:Movements for one trial for P7M. Dotted red line: Euclidean distance from the UI EMA sensor to the IRED above the Right Eyebrow; solid blue line: IRED above the Left Eyebrow



There were insufficient trials containing discrete movements to perform statistical analyses, but a few qualitative observations are possible. Of the 25 discrete movements (across all participants), 18 were eyebrow raises and 7 were eyebrow lowerings. These directional differences were not participant-specific. Except for participant P4F, whose one discrete movement was downward, the other 6 downward movements came from two participants (P6M and P8M) who had roughly balanced up and down movements (2 vs. 3 and 5 vs. 3, respectively). Discrete movements usually occurred for both eyebrows, whose onsets and magnitudes were generally similar (see Fig. 1). The onset of a discrete movement most often accompanied a pause that came before either a resumption of a sentence, an emphasized word, or a backchannel. This was the case for 16 of the 25 discrete movements and across all 6 participants. The onset of the other 9 discrete movements aligned either with a prominent word in the discourse (7/9), with the onset of a quotation (1/9), or the end of an utterance (1/9). The onset of discrete movements most often preceded the element of speech with which they seemed coordinated (17/25), though sometimes they were roughly time-aligned (5/25) or even followed (3/25). The lag after the onset of a discrete movement and an element of the speech acoustics that it preceded ranged from 20–470 ms, and varied within participant. While discrete movements were identifiable and appeared where expected, Fig. 1 illustrates that they account for a small amount of the time that eyebrows move.

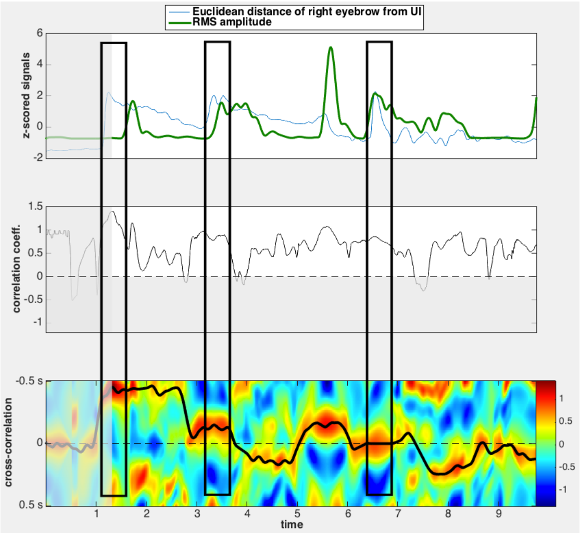
2.5.2. Eyebrow movement and acoustics

The relationship between eyebrow movement and acoustics was investigated through a series of correlation analyses. Of the 19 trials not excluded for insufficient movement, an additional 5 trials were discarded because the SD of distance from the UI was ≤ 0.5 mm when discrete movements were ignored. Two participants (P3F, P5F) had no remaining data, leaving 14 trials across 4 participants.

Movement was correlated with root mean squared (RMS) acoustic amplitude. RMS was calculated with a 200 ms sampling window in order to capture cyclic aspects of the acoustics that could plausibly be correlated with physical movement of the skin and head, and then down-sampled to 100 Hz to match the movement data sampling rate. Acoustic and movement signals were z-scored within trial before correlation since the units of the two signals resulted in very different magnitudes. Distance data for whichever eyebrow moved more (determined by SD) was used for the correlation.

Correlation map analysis (CMA) was used (see [1] for a more detailed explanation), though we used a simpler calculation without a decay parameter. Unlike simple correlation between two signals, which calculates correlation with a constant temporal offset (usually none), CMA can capture correlation across varying temporal offsets between signals, i.e., when the phasing of the oscillations of two signals is not constant. The top panel of Fig. 2 shows the z-scored RMS amplitude (thick green) and 3D distance of the right eyebrow from the UI (thin blue) for the same trial shown in Fig. 1. The bottom panel of Fig. 2 is the correlation map that shows the correlation coefficient between RMS amplitude and eyebrow movement, calculated within sliding windows of a length of 50 ms. Each signal was detrended within each window to minimize spuriously high correlations due to the inherent correlation across time samples of physical movement data. The *x* axis indicates time. Dark red indicates strong positive correlation between the signals. Dark blue indicates strong negative correlation. The *y* axis indicates the temporal offset between the two signals (from –0.5 to +0.5 s). Correlation in the upper half of the map (negative *y*-axis values) indicates correlation of right eyebrow movement with later samples of RMS amplitude, i.e., eyebrow movement precedes RMS amplitude changes. Correlation values in the lower half thus indicate that RMS amplitude changes precede eyebrow movement. The black line indicates the optimal path through the largest positive correlation coefficients of the map across all samples, similar to the maximum correlation path used by [16]. The middle panel shows the correlation coefficient values along the black line in the map. Since the path tracks positive correlation, negative correlation ranges are greyed out, as is the period of silence at the beginning of the trial. CMA is a good tool for exploring the relationship between two signals, especially when the phasing between them is unknown or variable, but space limitations prevent showing all maps from all trials here. We therefore show representative maps and describe the rest.

Figure 2: Correlation map of right eyebrow movement with RMS amplitude the same trial shown in Fig. 1. Grey shading indicates no speech. Boxes surround discrete movements.

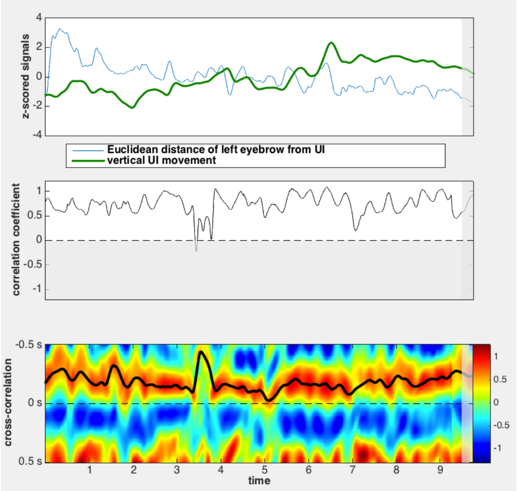


The CMA for the trial in Fig. 2 shows that the first two discrete movements in this trial preceded the RMS amplitude associated with the beginning of an utterance, by about 0.45 and 0.10 s respectively, and the third discrete movement was synchronous with the RMS amplitude rise. These movements are consistent with those based on the movement onsets reported in section 2.5.1. While there are other periods of reasonably high positive correlation between the signals during continuous movement of the right eyebrow, the offset varies between negative and positive. This pattern was characteristic of other trials and participants to varying degrees. Coefficients between the two signals tended to have stretches of strong positive correlation (coeffs. > 0.5), but the offsets varied between acoustics preceding, eyebrow movement preceding, and the two signals being in-phase. There were also periods in several trials with no strong correlation.

2.5.3. *Eyebrow movement and head movement*

We analyzed the correlation between eyebrow movement and head movement, as indexed by the vertical position of the UI EMA sensor in coordinate space, for the same 14 trials analyzed in section 2.5.2. UI position was z-scored and then detrended within correlation windows as above.

Figure 3: Correlation of left eyebrow movement with UI vertical movement (one trial from P6M).



For 3 of 4 participants, there was strong correlation between the distance of the eyebrow from the UI and the vertical movement of the UI, with movement of the eyebrow preceding the UI movement by approximately 0.2 seconds. A typical trial is shown in Fig. 3. The exception was that the correlations for P8M were much more varied. However, for P8M the correlation between eyebrow movement and UI movement in the anterior-posterior plane was very strong, similar to the vertical movement observed for the other participants. The difference for this participant may be due to idiosyncratic movement, which has been noted as a property of head and face movements [12], or to the fact that the participants' body posture was not forced to be exactly aligned with the 3 dimensions of the machine coordinate space.

3. DISCUSSION AND CONCLUSION

We have quantified movements of the eyebrows relative to the skull by using integrated optical and electromagnetic tracking to analyze the relationship of those movements with speech acoustics and with head movements using correlational map analysis.

We replicated the previously noted [10, 18] distinction between discrete and continuous eyebrow movements relative to the head. Discrete movements were aligned with expected properties of speech, most notably prosodic prominence. There were few instances of turn-taking in our corpus, but many discrete eyebrow movements were aligned with (and usually preceded) onsets of utterances, consistent with using eyebrow movement to signal turn-taking.

Our methodology also allowed for the exploration of the relationship between continuous eyebrow movements, which has not been addressed in the literature. There was sufficient correlation between eyebrow movement and RMS amplitude to suggest further study of these movements would be fruitful. We also found high levels of correlation and very consistent phasing between eyebrow movements and head movements, with the former slightly preceding the latter across participants.

It has been argued that orofacial movements arise from the same control source that generates vocal-tract movements [8, 17, 19]. Evidence in favor of this view are: findings linking eyebrow height to F0 [8], that observed facial movements could be predicted by EMG of lingual muscles [17], that orofacial movements sometimes predict RMS amplitude better than vocal tract flesh points [19], and that orofacial movements could be estimated from tongue movements better than vice versa [19].

Other researchers propose an independent source for non-vocal-tract movements [6, 10, 18], claiming they are coordinated by some other mechanism. This position is motivated by the fact that relationship between discrete eyebrow movements and, e.g., F0 are not systematic or mandatory [2, 6].

Our results are more consistent with the latter view, since even those of our participants who did produce eyebrow movements during speech did not produce them in all expected locations. The relationship between those eyebrow movements and acoustics showed some regular correlation, but varied within and across participants. However, our data suggest that speech-accompanying movements of the eyebrows and the head are likely to share a control mechanism, given their high degree of correlation within and across participants.

In conclusion, our findings, while preliminary, suggest that continuous eyebrow movements, not just discrete movements, are correlated with the acoustic signal to some extent, and that they are strongly coordinated with head-movement gestures produced during speech. These findings highlight that further development of theoretical accounts of the control mechanisms of all movements that are involved in speech production include continuous as well as discrete eyebrow movements.

4. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from NIH Grant DC-002717 to Haskins Laboratories and the City University of New York.

5. REFERENCES

[1] Barbosa, A. V., Déchaine, R.-M., Vatikiotis-Bateson, E., Yehia, H. C. 2012. Quantifying time-varying coordination of multimodal speech signals using correlation map analysis. *J. Acoust. Soc. Am.* 131, 2162–2172.

[2] Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R. 1996. About the relationship bettween eyebrow movement and F0 variations. *Proc. 4th ICSLP*, Philadelphia, PA, 2175–2178.

[3] Davis, C., Kim, J. 2006. Audio-visual speech perception off the top of the head. *Cognition* 100, B21–B31.

[4] Foxton, J. M., Riviere, L.-D., Barone, P. 2010. Cross-modal facilitation in speech prosody. *Cognition* 115, 71–78.

[5] Granström, B., House, D., Lundeberg, M. 1999. Prosodic cues in multimodal speech perception. *Proc. 14th ICPhS*, San Francisco, CA, 655–658.

[6] Guaïtella, I., Santi, S., Lagrue, B., Cavé, C. 2009. Are eyebrow movements linked to voice variations and turn-taking in dialogue? An experimental investigation. *Language and Speech*, 52, 207–222.

[7] Hoole, P., Zierdt, A. 2010. Five-dimensional articulography. In: Maasen, B., van Lieshout, P. (eds), *Speech Motor Control: New developments in basic and applied research*. Oxford: Oxford University Press, 331–349.

[8] Huron, D., Shanahan, D. 2013. Eyebrow movements and vocal pitch height: Evidence consistent with an ethological signal. *J. Acoust. Soc. Am.* 133, 2947–2952.

[9] Keating, P. A., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E. T., Bernstein, L. E. 2003. Optical phonetics and visual perception of lexical and phrasal stress in English. *Proc. 15th ICPhS*, Barcelona, 2017–2074.

[10] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57, 396–414.

[11] Lansing, C. R., McConkie, G. W. 1999. Attention to facial regions in segmental and prosodic visual speech perception tasks. *JSLHR* 42, 526–539.

[12] Munhall, K. G., Buchanan, J. N. 2004. Something in the way she moves. *TiCS* 8, 51–53.

[13] 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. *Psych. Sci.* 15, 133–137.

[14] Swerts, M., Krahmer, E. 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *J. Phon.* 36, 219–238.

[15] Thomas, S. M., Jordan, T. R. 2004. Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *J Exp. Psych: Human Perception and Performance* 30, 873–888.

[16] Vatikiotis-Bateson, E., Barbosa, A. V., Best, C. T. 2014. Articulatory coordination of two vocal tracts. *J. Phon.* 44, 167–181.

[17] Vatikiotis-Bateson, E., Munhall, K. G., Hirayama, M., Kasahara, Y., Yehia, H. C. 1996. Physiology-based synthesis of audiovisual speech. *Proc. 4th Speech Production Seminar: Models and Data*. Autrans, France, 241–244.

[18] Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S. 2004. The timing of speech-accompanying gestures with respect to prosody. Proc. From sound to sense: 50+ years of discoveries in speech communication, MIT, Cambridge, MA, C97–C102.

[19] Yehia, H. C., Rubin, P. E., Vatikiotis-Bateson, E. 1998. Quantitative association of vocal-tract and facial behavior. *Speech Communication* 26, 23–43.