Co-speech movement behavior at floor exchanges and interruptions

Samantha Danner (USC), Jelena Krivokapić (U. Michigan) & Dani Byrd (USC)
sfgordon@usc.edu

Co-speech bodily movements, e.g. of the hands, head, and face, have been shown to be implicated in speech planning (Butterworth & Beattie 1978), in the prosodic structure of speech (Krivokapić et al. 2017), and in semantic/pragmatic aspects of speech (Ozyurek 2014), making co-speech movement behavior critical to felicitous conversational interaction (Holler et al. 2017).

We examine how interacting speakers deploy co-speech movements at notable moments in conversation: 1) felicitous **floor exchanges**, in which a speaker concludes speaking and another speaker subsequently starts speaking, 2) **non-consummated exchanges** (NC exchanges), in which a speaker concludes speaking, then re-takes the floor, 3) **interruptions**, in which a speaker begins speaking prior to the conclusion of their partner's speech turn, and 4) **non-turn-adjacent speech** (NTA speech). As turn-taking draws on speech planning, prosody and pragmatics, we hypothesize that spatiotemporal characteristics of co-speech movements will quantifiably differ across these contexts (Beňuš et al. 2011; Lee & Narayanan 2010).

From a larger experiment dataset (*N* = 12 native speakers of Am. Engl.), two pairs of previously unacquainted speakers – S5+S6 and S7+S8 – were analyzed for the present report. Positioned with a line of sight to one another at facing desks 2m apart, participants were first instructed to quickly review 22 common English nursery rhymes like "Jack & Jill" or "Humpty Dumpty." Then, dyads were recorded using dual NDI Wave electromagnetic articulometers (EMA) for motion tracking of sensors placed at the lip corners, lower lip, most mobile point of both eyebrows, intraorally on maxilla and jaw (Fig. 1), with video recording of hand movement and directional microphones for speech audio. In this naturalistic yet structured task designed to elicit conversational speech turns, participants collaboratively complete nursery rhymes, prompted initially by the rhyme's first line. Dyad members were instructed to help their partner if the partner was unable to complete a rhyme. 22 rhymes were each presented twice (resulting in 28 min, 46 s of data and 673 turns for S5S6 and 14 min, 44 s of data and 312 turns for S7S8).

Speech audio was transcribed using the Penn forced aligner (Yuan & Liberman 2008). The mean inter-turn interval (ITI) across each dyad's data was used to determine the size of the gestural analysis region (S5S6 ITI = 0.521s; S7S8 ITI = 0.631s). Floor exchanges and NC exchanges were coded as intervals around the endpoint of a participant's acoustic speech turn ± average ITI, interruptions coded as the acoustic starting point at which one speaker interrupts another ± average ITI, and NTA speech as any region of speech falling between turn boundaries + ITI. The analysis assesses 1) mean eyebrow velocity peaks per second (PPS), derived from the Euclidean from the maxilla to right brow sensors), as an index of brow gesture density, and 2) mean angular velocity PPS of rotations of a plane formed by three head reference sensors with respect to the EMA system origin, as an index of head gesture density.

Results indicate that gestural density in head and brow movements pattern similarly to one another and across dyads, despite some individual speaker variation. The greatest co-speech gestural density (mean PPS values) occurs at felicitous **floor exchanges**, while the lowest mean gestural density is observed in **non-turn-adjacent** (**NTA) speech**. **NC exchanges** and **interruptions** exhibit intermediate gestural density, but the PPS in these regions is clearly distinct from baseline **NTA speech** (See Fig. 2 & Fig. 3). Finally, for two individuals there was depression of gesture density by a speaker when they are offered the floor and fail to take it.

In sum, speech acoustic behavior is just one of multiple communicative modalities that reflect speakers' willingness to exchange the conversational floor, and density of co-speech gestures may help distinguish turn-adjacent regions from non-turn-adjacent regions. The research methods and results described here indicate that quantitatively analyzing co-speech movement behavior at critical moments of interaction in conversation may begin to elucidate how face-to-face conversation can proceed so seamlessly. [Supported by NIH R01 DC003172 (Byrd)]

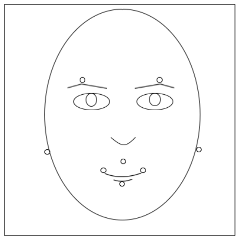


Figure 1: EMA Sensor Placement Schematic

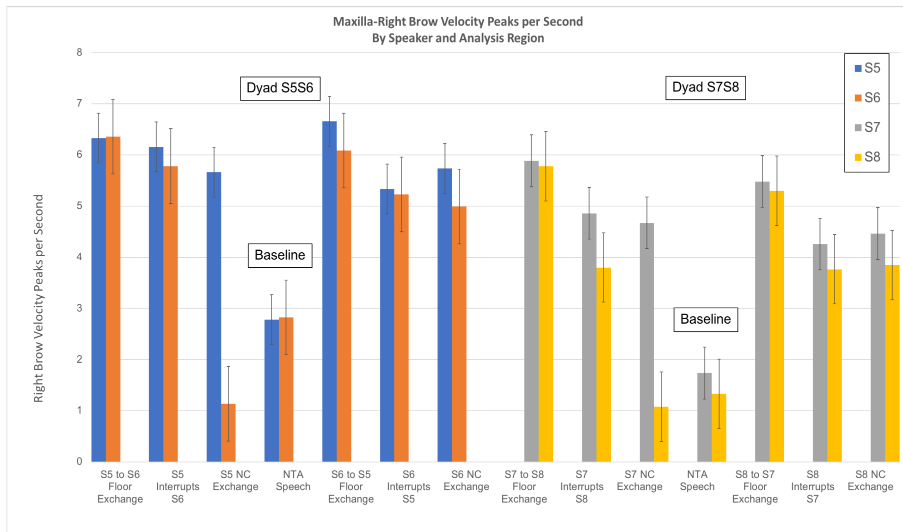
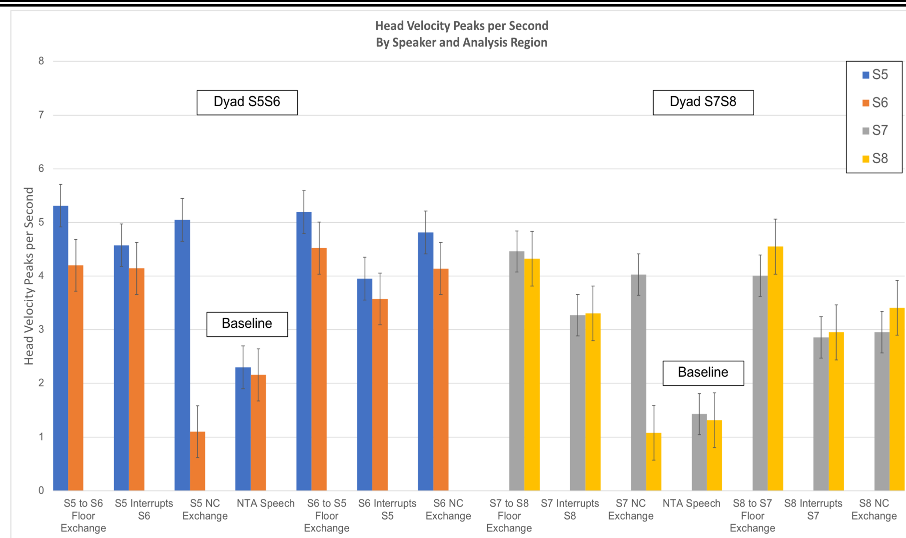


Figure 2: Brow Gesture Density

Figure 3: Head Gesture Density



References

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*, 3001–3027.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell (Ed.), *Recent advances in the psychology of language: formal and experimental approaches* (pp. 347–360). New York: Plenum Press.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2017). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, 1–9.

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology*, *8*, 1–26.

Lee, C. C., & Narayanan, S. (2010). Predicting interruptions in dyadic spoken interactions. In *Proceedings of ICASSP* (pp. 5250–5253). Dallas, Texas.

Ozyurek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Phil. Trans. Roy. Soc. B: Biological Sciences*, *369*, 20130296–20130296.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08* (pp. 6–9).

Acknowledgments: Thanks to Dr. Sungbok Lee, Cheyenne Laroque and Nynaeve Perkins Booker.