Hand-Mouth Coordination in a Pointing Task Requiring Manual Precision

Aleksandra Ćwiek, Susanne Fuchs

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin Germany
Humboldt Universität zu Berlin, Berlin Germany

The synchronization between pointing gesture and speech is an important step in the development (Iverson & Thelen, 1999) and is a frequent phenomenon in our daily life. Various authors have investigated the synchronization between pointing gestures and speech in detail. It has been shown to be linked to various prosodic phenomena, such as focus marking (Rochet-Capellan et al., 2008) and lengthening under prominence (Krivokapic et al., 2018). Past studies reveal that the perturbation of either hand or the mouth affects the other system as well (Pouw & Dixon, 2019; Chu & Hagoort, 2014), what emphasizes the functional link between the two. Yet, hand-mouth coordination involves two different motor systems with specific dynamic properties. While speech articulators consist largely of soft tissue, which is light and fast, hands and arms are built of joints and bones, which are much heavier in mass, thus slower. These different dynamical properties may affect the synchronization behavior e.g., the slower motor system may have to start earlier than the faster motor system while the faster motor system may be more flexible in adjusting to the slower system. Our exploratory study aims at investigating the timing between hand and mouth in a speech-pointing task requiring relatively large hand movements and a high degree of manual precision while reaching the target.

Methods

The experimental task for the participants was to "shoot" cans projected onto a wall in front of them with a laser pointer. Participants stood ca. 1 m in front of the wall with their hands down. They first saw a blank screen. When a can was projected on the wall, participants were asked to point at it and say the word printed on the can (either *piff* or *paff*). After an animation of the can falling, a blank screen was displayed, and subsequently a new can in a different position appeared (for further details, see Ćwiek & Fuchs, 2019).
The arm and lip/jaw movement as well as the acoustic signal were recorded simultaneously using the Optitrack motion capture system (Motive, version 1.9.0) and a Sennheiser ME 64 cardioid microphone. The sampling frequency was 120 Hz for the motion data and 44.1 kHz for the acoustics. A total of 14 markers were placed on the participant's body. Here we will focus on the wrist marker of the pointing arm and the upper lip and jaw markers that were used to calculate the lip aperture during speech production. Thirty-one female German speakers were recorded. Our preliminary analysis is based on a subset of the data. Hand movement is labelled on the velocity signal of the wrist marker (Figure 1) using a 20% threshold criterion. In the acoustic signal, we annotated the burst, the end of the aspiration phase, the vowel, and the final fricative.

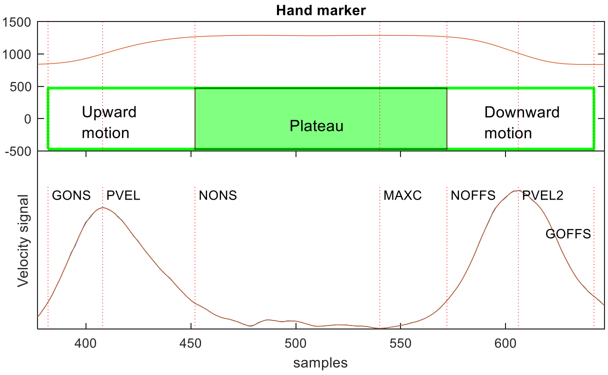

Figure 1: Labelling scheme for hand marker motion (right wrist), upper track: vertical motion and the three gesture intervals (upward motion, plateau, downward motion), lower track: velocity signal and definition of gesture events using Mview (Tiede, 2005).

Results

Figure 2 displays the averaged results for the different hand gesture intervals, aligned at the acoustic burst of the initial voiceless /p/ at time 0. The detectable acoustic speech onset is clearly visible after the hand has been lifted and reached the onset of the plateau phase.

We are currently analyzing the onset of the lip aperture data, which shows a high degree of speaker-specific behavior. While a few speakers close their lips to produce /p/ already when they start lifting their hand, suggesting a large anticipation of the speech gesture, others start to close their lips in preparation for the plosive only around the onset of the hand gesture plateau. The onset of the lip closure in anticipation of the plosive could occur at any time between the onset of hand motion and the onset of speech. Nevertheless, specific alignment with the onset of the upward hand movement or with reaching the plateau – i.e., with beginning or finishing a gesture – are preferred by speakers. These preliminary analyses give further evidence for the synchronization of hand and mouth.

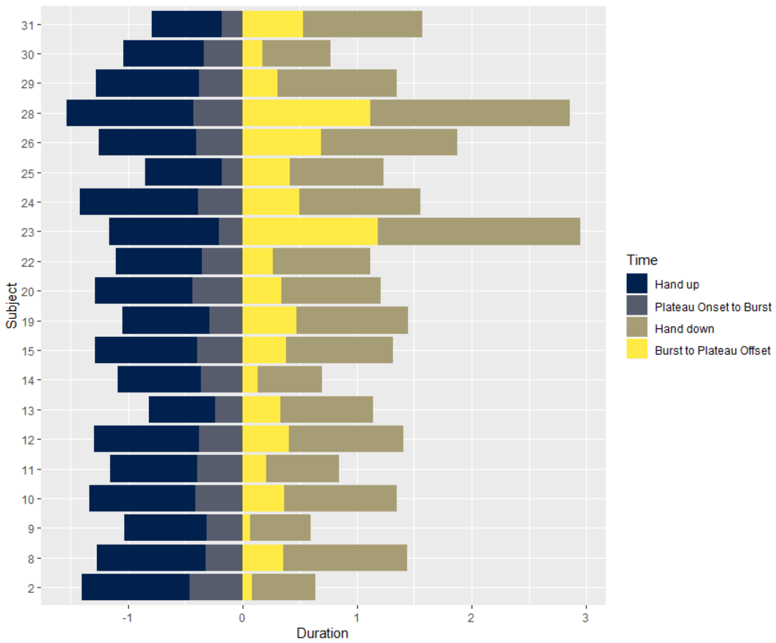

Figure 2: Stack plot for different averaged time intervals (different colors) aligned at the acoustic burst (t = 0); each stack bar corresponds to one subject:

References: ►Chu, M, & Hagoort, P (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*(4), 1726ff. ►Ćwiek, A & Fuchs, S (2019). Iconic Prosody is Rooted in Sensori-Motor Properties: Fundamental Frequency and the Vertical Space. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Montreal, Canada (pp. 1572-1578). ►Iverson, JM, & Thelen, E (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *J of Consc Studies*, *6*(11-12), 19-40. ►Krivokapić J, Tiede MK, Tyrone ME. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Lab Phon*. 8(1): 3ff. ►Pouw, W, & Dixon, JA (2019). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science*, *43*(3), e12721. ►Rochet-Capellan, A, Laboissière, R, Galván, A, & Schwartz, JL (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *JSHLR 51*, 1507–1521 ►Tiede, M (2005). MVIEW: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Labs*.