Coordination of head movements and speech in first encounter dialogues

Patrizia Paggio
University of Copenhagen and University of Malta
paggio@hum.ku.dk
patrizia.paggio@um.edu.mt

This paper presents an analysis of the temporal alignment between head movements and associated speech segments in the NOMCO corpus of first encounter dialogues [14]. Our results show that head movements tend to start slightly before the onset of the corresponding speech sequence and to end slightly after, but also that there are delays in both directions in the range of -/+ 1s. Various factors that may influence delay duration are investigated. Strong correlations are found between delay length and the duration of the speech sequences associated with the head movements. Effects due to the different head movement types are also discussed.

Many studies have claimed that speech and gesture, in particular hand gestures, are two manifestations of the same underlying cognitive mechanism [12], [13], [7], [8], [2]. One aspect of this tight relation is the temporal coordination between the two modalities. It is generally agreed that hand gestures are coordinated with prosodic events, such as pitch accents and prosodic phrase boundaries [1], [6], [10], [11]. It has also been shown experimentally that subjects are sensible to asynchrony, especially when gesture strokes are made to lag behind the accompanying speech [9], and also that coordination with prosody contributes to the well-formedness of multimodal signals [3].

These studies deal with hand gestures, especially beats. Head movements often have the same quality of manual beats, by being rapid, simple and often repeated movements. Therefore, we would expect them also to show tight temporal synchronisation with the words they co-occur with. Coordination between head movements and speech is discussed in [4], where it is claimed that speakers' head movements are attuned to prosody in establishing peaks and prosodic boundaries especially in cases of high intensity. Furthermore, in [5], it is argued that coordination with speech, together with physical properties of head movements (cyclicity, amplitude, duration) are indicative of the diverse communicative functions of the movements themselves. However, the temporal synchronisation between the two modalities is not described in detail, and the datasets explored in these papers only consists of a couple of hundreds of head movements.

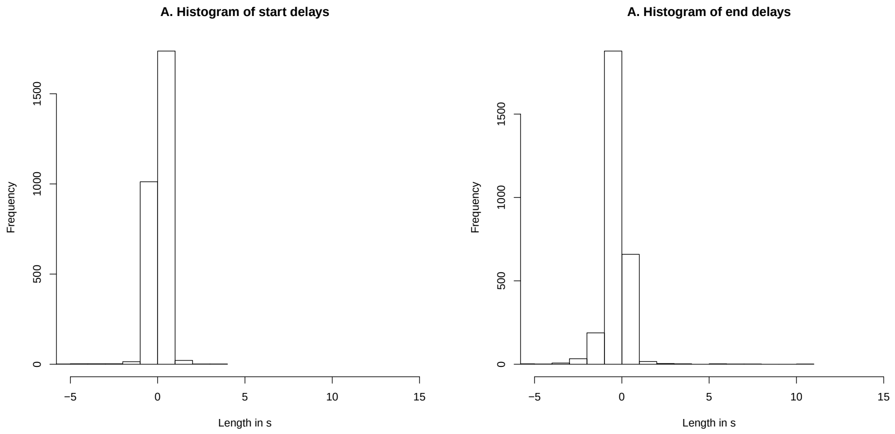

Figure 1: Start and end delays in the NOMCO corpus. In histogram A, bars to the left of zero (negative) correspond to movements starting after the onset of the corresponding speech, and those to the right to movements starting before speech. In histogram B, bars to the left of zero count movements ending after speech offset, and those to the right movements ending before.

In this paper, we look at temporal synchronisation at the level of onsets and offsets of movements and associated speech in the NOMCO corpus [14].

The total number of head movements in the NOMCO corpus is 3117. We are only interested in head movements that are linked to word sequences in the gesturer's own speech stream, and ignore unimodal head movements performed while the interlocutor is speaking. That leaves a subset of 2795 movements, which will be used to analyse movement-speech synchronisation in this study. The duration of most head movements in this dataset is around 1s, although there are occurrences of up to 7s (mean = 0.93s, SD = 0.58s).

On average, in our data head movements tend to start 0.05s before the onset of the associated speech sequence (SD = 0.40s), and to end 0.28s after its offset (SD = 0.64s). The histograms in Figure 1 show that in more than 2500 cases, delays range between -1 and 1, and that about 1750 delays are actually positive delays in the range 0 to 1. Most of the end delays, on the other hand, are distributed in the range -1 to 0, meaning that in most cases, the head movement ends up to 1s after speech offset. About 700 cases, however, fall in the range 0 to 1. To have an intuition of what the size of the delays means, we can compare it with the mean word duration in the whole NOMCO corpus, which is 0.21s, or the mean length of a linked speech sequence in the dataset, which is as we saw 0.59s. It can also be mentioned that in the already cited study in [9] it is found that subjects are sensible to asynchrony of as little as 0.2 seconds if a gesture lags behind speech, whereas in [3] it is claimed that subjects react to gesture-speech misalignments of at least 0.5 seconds.

Thus, a delay of 1s is not negligible, in that it corresponds to four words, or two speech sequences.

We have analysed a number of factors which may have an influence on the polarity and duration of the delays, including effects due to individual speaker variation, different conversations, movement type, as well as movement and linked speech segment durations.

The strongest effect we find is the correlation between delay length and the duration of the linked speech sequence. A strong negative correlation, in fact, can be observed between start delay length and the duration of the speech segments (Pearson's r = -0.58), while there is a strong positive correlation between end delay length and speech segment duration (Pearson's r = 0.51). In general, this means that the longer the speech chunk associated with the head movement is, the later the head movement starts and the earlier it ends. This, in turn, can be interpreted as a general tendency for the overlap between head movement and speech sequence to be maximised. This general pattern, however, varies depending on the movement type, with some types showing a more systematic adherence to the general tendency than others. While these differences seem to be related to the internal duration of the head movement in some cases (jerks, shakes, waggles), duration alone cannot explain the different behaviours of other movement types (e.g. nods, tilts and side turns). A more precise characterisation of the synchronisations patterns for these movement types probably needs to take into account the alignment between movement stroke and prosodic peak, or kinetic features such as amplitude and intensity.

References

[1] Bolinger, D. *Intonation and its parts: Melody in spoken English.* Stanford, CA: Stanford, 1986.

[2] De Ruiter, J.-P. The production of gesture and speech. In *Language and Gesture.* Cambridge University Press, 2000.

[3] Giorgolo, G., and Verstraten, F. A. Perception of 'speech-and-gesture' integration. In *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008* (2008), pp. 31–36.

[4] Hadar, U., Steiner, T., Grant, E. C., and Rose, F. C. Head movement correlates of juncture and stress at sentence level. *Language and Speech 26*, 2 (1983), 117–129.

[5] Hadar, U., Steiner, T., and Rose, F. C. The timing of shifts of head postures during conversation. *Human Movement Science 3*, 3 (1984), 237–245.

[6] Kendon, A. Gesture and speech: two aspects of the process of utterance. In *Nonverbal Communication and Language*, M. R. Key, Ed. Mouton, 1980, pp. 207–227.

[7] Kendon, A. *Gesture.* Cambridge University Press, 2004.

[8] Kita, S., and Özyürek, A. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language 48*, 1 (2003), 16–32.

[9] Leonard, T., and Cummins, F. The temporal relation between beat gestures and speech. *Language and Cognitive Processes 26*, 10 (2010), 1457–1471.

[10] Loehr, D. P. *Gesture and Intonation.* PhD thesis, Georgetown University, 2004.

[11] Loehr, D. P. Aspects of rhythm in gesture and speech. *Gesture 7*, 2 (2007).

[12] McNeill, D. *Hand and Mind: What Gestures Reveal About Thought.* University of Chicago Press, Chicago, 1992.

[13] McNeill, D. *Gesture and thought.* University of Chicago Press, 2005.

[14] Paggio, P., Allwood, J., Ahlsén, E., Jokinen, K., and Navarretta, C. The NOMCO multimodal nordic resource - goals and characteristics. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (Valletta, Malta, 2010), European Language Resources Association (ELRA).