Coordination of facial expressions and head movements in first encounter dialogues

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

Costanza Navarretta
University of Copenhagen
costanza@hum.ku.dk

The coordination of different signals in human communication has been studied especially as regards gesture and speech, and there is considerable agreement that hand gestures are coordinated with prosodic events, such as pitch accents and prosodic phrase boundaries (Bolinger, 1986; Kendon, 1980; Loehr, 2004; Loehr, 2007). Experimental work has also clearly shown that people are sensitive to disruptions of the natural temporal alignment between the two modalities (Leonard and Cummins, 2010; Giorgolo and Verstraten, 2008). Coordination between head movements and speech, and how this is mediated by prosody, is discussed in Hadar et al. (1983) and (1984). More recently, Paggio (2016) and Paggio and Navarretta (2016) investigated the temporal alignment between head movements and co-occurring speech segments in multimodal data, and discussed a number of factors that affect the alignment.

Studies dealing with the relation between facial expressions and other expressive modalities have looked at smiles co-occurring with gaze and head movements towards or away from the interlocutor (Keltner, 1995); correlations between lip-corner displacement in smiles and head or eye movement (Cohn et al., 2004b); and occurrence of eyebrow raising with forward head movements (Cohn et al., 2004a). Work where multimodal coordination of different expressions is used to model the behaviour of Embodied Conversational Agents include Cassell et al. (1999), and Lee and Marsella (2006). Finally, a study of how smiles and laughters can be generated based on the interlocutor's smiling and laughing behaviour, is in El Hadded et al. (2016).

In this paper, we focus on the coordination between facial expressions and head movements in cases in which there is indeed an overlap between the two modalities. In particular, we look at how the onset of facial expressions is coordinated with the first overlapping head movement, in other words which of the two modalities precedes the other and possibly why. The motivation for the analysis is to shed light on a less studied aspect of multimodal communication – an aspect that is relevant to the generation of natural multimodal expressions in embodied conversational agents.

The data for this study consist of 1448 facial expressions and 3117 head movements extracted from an annotated corpus of twelve first encounter dialogues. The average duration of the facial expressions is 1.98s (sd=1.6). The spread of the duration is remarkable, with the shortest expression lasting 0.16s, and the longest 12.12. Smiles are the expressions showing the most variation in duration, with scowls showing the least. Head movements are shorter. Their mean duration is 0.93s (sd=0.58), with up-turns providing the shortest and least varying movements, and head shakes the longest outlier (7.08s).

Head movements can be single or repeated. In our dataset there are 2315 single head movements, and 794 are repeated ones. The mean duration for single movements is 0.82s (sd=0.48s), while it is 1.28 for repeated ones (sd=0.70s).

Table 1: Facial expression onset overlaps

Facial expression type	Start before head	Start after head	Total
Smile	303	308	611
Laughter	84	116	200
FrownScowl	66	55	121
EyebrowRaise	190	161	351
FaceOther	46	45	91
Total	689	685	1374

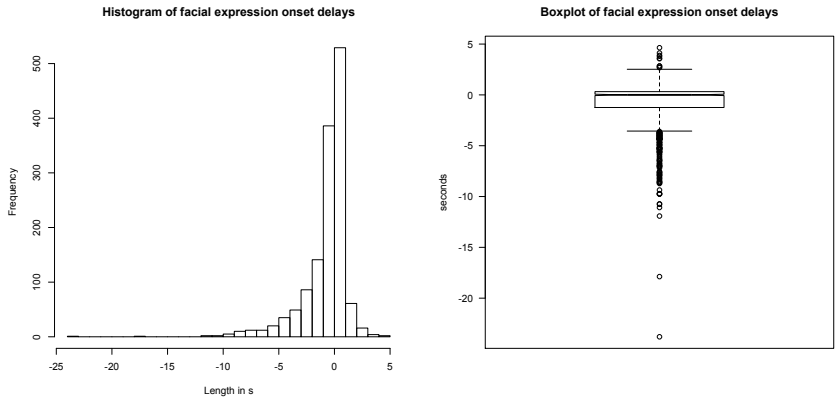

Figure 1: Distribution of onset delays between facial expressions and the first overlapping head movement shown as a histogram (on the left), and a boxplot (on the right). Positive delays indicate facial expressions whose onset follows the onset of the corresponding head movement.

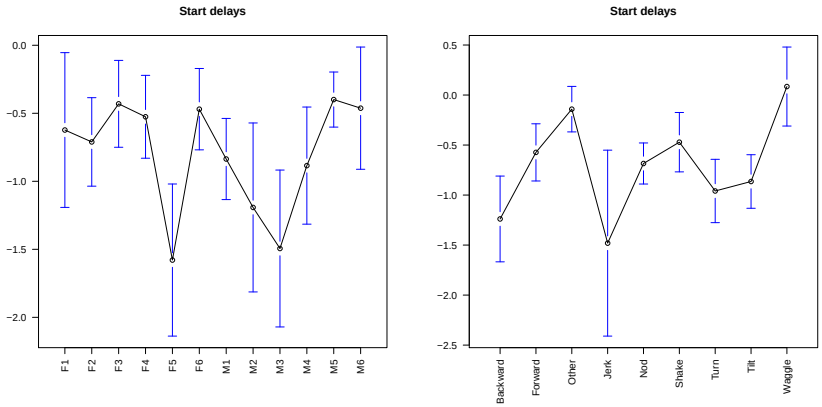

Figure 2: Mean values and confidence intervals for the duration of start delays according to individual speakers (plot to the left), and associated head movement (plot to the right).

When we look at the number of facial expressions that overlap with the other modality at the onset, we see that 689 of them, i.e. 48% of the total facial expressions in the corpus, start before or at the same time as an overlapping head movement. Conversely, 685 facial expressions, i.e. 47% of the total, start after the first overlapping head movement. Frequency counts of the various facial expression types against their onset relation with the first co-occurring head movement are shown in Table 1. In general, it can be concluded that there is a very high likelihood for facial expressions to be accompanied by head movements (1374 cases out of 1448, i.e. 95% in total). However, whether the onset of the facial expression precedes or follows the onset of the head movement is equally likely. Nevertheless, a χ-squared test of independence showed that the type of onset delay depends on the facial expression type (χ^2=8.5563, df=4, p-value=0.0732). This dependency is due to the significant difference between the two types of delay in the case of *Laughter* and *EyebrowRaise*, where we see that laughters tend to start after an overlapping head movement, whilst the opposite is true of eyebrow raises. An earlier study found that children used eyebrow raises preceding head movements in connection with visual search (Jones and Konner, 1970). In our data, this difference may well be due to different physical characteristics of the signals. Thus, eyebrow movements are quite small and their onset may become more quickly visible compared to that of the accompanying head movement. Conversely, laughters, which also imply a vocalisation, may be slower at the onset although planned together with the head movement.

The two plots in Figure 1 show the distribution of the duration of the onset delays between facial expressions and the first overlapping head movement. As can be seen, most of the delays are in the area between -1s (facial expression starting before the head movement), and +1s (facial expression starting

after the head movement). There are, however, quite a number of outliers in the negative range, as clearly shown by the boxplot, so that the distribution does not conform to normality.

Statistical analysis shows a main effect of individual speaker variation (Kruskal-Wallis: χ^2= 33.384, df=11, p-value$<$0.001), and an effect of head movement type (Kruskal-Wallis: χ^2=37.001, df=8, p-value$<$0.001) on the distribution of the start delay size. The effect of facial expression type, on the contrary, does not reach significance (Kruskal-Wallis: χ^2=8.3289, df=4, p-value=0.08025).

As can be seen from the plots showing mean values and confidence intervals in Figure 2, two of the speakers, F5 and M3, stand out in that they display a mean negative delay onset of around 1.5s. As for the head movement type, negative delays are seen especially together with HeadBackwards and Jerks (UpNods). The two movement types are physically similar in that they both imply a backward movement of the neck which may physically be slightly more demanding than a forward movement, and have the effect of the movement becoming visible after the onset of the facial expression. Conversely, waggles tend to precede the associated facial expressions. Waggles are rather complex movements and relatively long on average (mean duration=1.2s), characteristics which may explain why they are initiated before the associated facial expressions.

To conclude, our data clearly show that facial expressions have a strong tendency to co-occur with head movements. We have also found interesting patterns concerning the delays between the two types of signal. Thus, laughters and eyebrow raises behave in opposite ways, with laughters slightly following and raises slightly preceding the associated head movement. Similarly, the type of head movement also has an effect on the direction of the delay, with head movement that imply an upward movement of the neck following the associated facial expression and long complex head movements preceding it. It seems reasonable to explain these effects at least partially in terms of the physical characteristics of the movements. Considering their function in the conversation, however, may shed additional light on the fine-grained coordination between different signals. This is left for future research.

References

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

J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 520–527.

J. F. Cohn, L. I. Reed, Z. Ambadar, Jing Xiao, and T. Moriyama. 2004a. Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 1, pages 610–616, Oct.

J. F. Cohn, L. I. Reed, T. Moriyama, Jing Xiao, K. Schmidt, and Z. Ambadar. 2004b. Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pages 129–135, May.

Kevin El Haddad, Hüseyin Çakmak, Emer Gilmartin, Stéphane Dupont, and Thierry Dutoit. 2016. Towards a listening agent: A system generating audiovisual laughs and smiles to show interest. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 2016, pages 248–255, New York, NY, USA. ACM.

Gianluca Giorgolo and Frans A.J. Verstraten. 2008. Perception of 'speech-and-gesture' integration. In *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008*, pages 31–36.

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

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

N. G. Blurton Jones and M.J. Konner. 1970. An experiment on eyebrow-raising and visual searching in children. *Journal of Child Psychology and Psychiatry*, 11(4):233–240.

D. Keltner. 1995. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality & Social Psychology*, 68:441–454.

Adam Kendon. 1980. Gesture and speech: two aspects of the process of utterance. In M: R. Key, editor, *Nonverbal Communication and Language*, pages 207–227. Mouton.

Jina Lee and Stacy Marsella. 2006. Nonverbal behavior generator for embodied conversational agents. In *International Workshop on Intelligent Virtual Agents*, pages 243–255. Springer.

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

Daniel P. Loehr. 2004. *Gesture and Intonation*. Ph.D. thesis, Georgetown University.

Daniel P. Loehr. 2007. Aspects of rhythm in gesture and speech. *Gesture*, 7(2).

Patrizia Paggio and Costanza Navarretta. 2016. The Danish NOMCO corpus multimodal interaction in first acquaintance conversations. *Journal of Language Resources and Evaluation*, pages 1–32.

Patrizia Paggio. 2016. Coordination of head movements and speech in first encounter dialogues. In *Proceedings of the 3rd European Symposium on Multimodal Communication*, Linkping Electronic Conference Proceedings, pages 69–74.