

# Detection of Communicative Co-speech Gesturing in Conversations

**Martin Vels**

Institute of Computer Science  
University of Tartu  
martin.vels@ut.ee

**Kristiina Jokinen**

Institute of Computer Science  
University of Tartu  
Kristiina.jokinen@ut.ee

## Abstract

This paper concerns the analysis of conversational videos and especially automatic recognition of body movement of the conversing partners. We discuss a modification of our earlier movement recognition algorithm to decompose the bulk of body movement into three parts so as to distinguish body, head and feet gesturing as separate activities. The recognized gestures are visualised together with the participants' speech, to show how the conversational participants' gesture activity correlates with their speaking and listening.

## 1 Gesticulation and gesturing

Conversations form a social system whereby the interlocutors aim at exchanging information as well as conveying intentions, interests, and feelings. The listener is involved verbally (*as you said; what do you think?*) and by multimodal means (tone of voice, laugh, gesturing, head and body posture, etc.) to give feedback and construct common understanding with the speaker.

Interactive gestures form a class with the common function of including the listener in the conversation. They occur at specific moments in time and particular points in space, and can efficiently exert coordination of the conversation and provide meanings as the dialogue goes on.

Gestural signs are formed by the cognitive system that is also used in the movement of the body in the physical environment. Gesturing requires spatio-motoric thinking and ability to orient body parts toward a target in the physical environment, as well as the ability to track the target when it moves (Kita 2000). Kendon (2004) uses the term gesticulation to refer to the gesture as a whole (with the preparatory, peak, and recovery phases), while the term “gesture” refers

to visible action that participants distinguish and treat as governed by openly acknowledged communicative intent. Human body movements can be said to form a continuum from movement without overt communicative meaning to communicatively significant gesturing.

Body movements and the flow of speech are closely linked in one's communication system and between the interlocutors. For instance, it is noted that the peak of the gesture coincides with the conceptual focal point of the speech unit, and each new representational gesture appears with a new unit of meaning. Both utterance and gesture derive from a deeper idea unit source that they represent co-expressively.

## 2 Data and method

The 23 dialogues used for the experimentation are from the MINT (Multimodal INTERaction) project collected at University of Tartu (Jokinen and Tenjes 2012). They are first-encounter dialogues where the speakers are unfamiliar with each other and make acquaintance with their partner (cf. the Paggio et al. 2010). Original Full HD (1920x1080px) videos were resized to 640x360px and 25 frames per second.

The original algorithm (Vels and Jokinen 2015) allows us to find the positions of the moving persons in a video frame using a contour detection algorithm. It introduces a novel idea of initializing a background model from a single frame using 8-neighbourhood of each of the frame pixel and randomly choosing 20 neighbour-pixels instances to build the model.

This approach is extended by decomposing the participant's contour box into head, torso, and leg parts and retrieving their precise coordinates in the frame. First a very precise location and size of the head position is provided, and using this, the whole body contour is located within which the coordinates for the torso and legs can be retrieved.

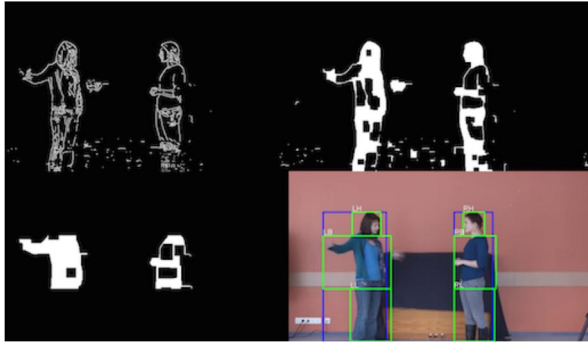


Figure 1 Some segmentation steps: background subtraction, closing, body contour, and final result with detected head, body and leg coordinates.

### 3 Results and discussion

We will describe the algorithm and the various steps in more detail in the full paper. We will also report interesting results on detecting interaction patterns between conversation partners, e.g. movements related to a person with beat gesturing (Fig. 2) and bending whole body forward (Fig. 3).

We will also discuss gestures when a person is laughing and pointing or waving her hands around. The results show gesturing as an activity related to speaking, cf. Battersby (2011). Gestures help the speaker to speak, yet depend on the linguistic choices made when constructing an utterance. We regard gesture as a separate and distinct mode of expression with its own properties, but which can

be brought into a cooperative relationship with spoken utterance, so that the two modes of expression being used in a complementary way.

### References

- Battersby, S. 2011. Moving Together: the organization of Non-verbal cues during multiparty conversation. PhD Thesis, Queen Mary, University of London.
- Campbell, N., Scherer, S. 2010. Comparing Measures of Synchrony and Alignment in Dialogue Speech Timing with respect to Turn-taking Activity. *Proceedings of Interspeech*. Makuhari, Japan
- Canny, J. 1986. A Computational Approach to Edge Detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698.
- Jokinen, K. 2009. *Constructive Dialogue Modelling: Rational Agents and Speech Interfaces*. Chichester: John Wiley.
- Jokinen, K. and Tenjes, S. 2012. Investigating Engagement – Intercultural and Technological Aspects of the Collection, Analysis, and Use of Estonian Multiparty Conversational Video Data. In *Proceedings of LREC'12*, pp. 2764 – 2769. Istanbul, Turkey: ELRA.
- Kendon, A. 2004. *Gesture: Visible action as utterance*. New York: Cambridge University Press.
- Kita, S. 2000. How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture*. pp. 162-185. Cambridge: Cambridge University Press. Open Access accepted version: <http://wrap.warwick.ac.uk/66257/>
- Patrizia Paggio, Jens Allwood, Elisabeth Ahlsén, Kristiina Jokinen & Costanza Navaretta. 2010. The NOMCO Multimodal Nordic Resource – Goals and Characteristics. In *Proceedings of LREC'10*, Valetta, Malta: ELRA.

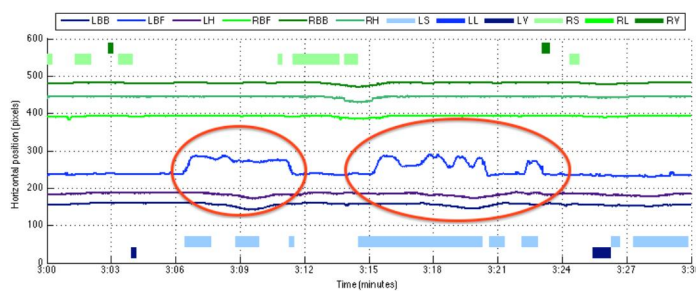


Figure 2 Beat gesturing simultaneous with speech (indicated by a light blue bar underneath the movement).

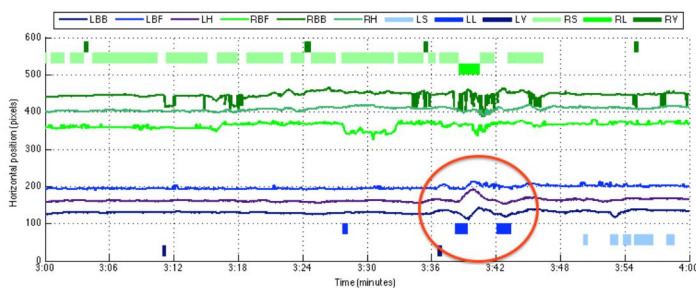


Figure 3 Leaning body movements in between speech and simultaneous with the partner's speech and gesturing.