Laughter and Body Movements as Communicative Actions in Interactions

Kristiina Jokinen
AIRC AIST Tokyo Waterfront, Japan
kristiina.jokinen@aist.go.jp

Trung Ngo Trong
University of Eastern Finland, Finland
trung.ngotrong@uef.fi

Abstract

This paper focuses on multimodal human-human interactions and especially on the participants' engagement through laughter and body movements. We use Estonian data from the Nordic First Encounters video corpus, collected in situations where the participants make acquaintance with each other for the first time. This corpus has manual annotations of the participants' head, hand and body movements as well as laughter occurrences. We examine the multimodal actions and employ machine learning methods to analyse the corpus automatically. We report some of the analyses and discuss the use of multimodal actions in communication.

1. Introduction

Human multimodal communication is related to the flow of information in dialogues, and the participants effectively use non-verbal and paralinguistic means to coordinate conversational situations, to focus the partner's mind on important aspects of the message, and to prepare the partner to interpret the message in the intended way.

In this paper we investigate the relation between body movements and laughter during first encounter dialogues. We use the video corpus of human-human dialogues which was collected as the Estonian part of the Nordic First Encounters Corpus, and study how human gesturing and body posture are related to laughter events, with the ultimate aim to get a better understanding of the relation between the speaker's affective state and spoken activity. We estimate human movements by image processing methods that extract the contours of legs, body, and head regions, and we use speech signal analysis for laughter recognition. Whereas our earlier work (Jokinen et al. 2016) focussed on the video frame analysis and clustering experiments on the Estonian data, we now discuss laughter, affective states and topical structure with respect to visual head and body movements.

We focus on human gesticulation and body movement in general and pay attention to the frequency and amplitude of the motion as calculated automatically from the video recordings. Video analysis is based on bounding boxes around the head and body area, and two features, speed of change and speed of acceleration, are derived based on the boxes. The features are used in calculating correlations between movements and the participants' laughing.

Our work can be compared with Griffin et al. (2013) who studied how to recognize laughter from body movements using signal processing techniques, and Niewiadomski et al. (2014, 2015) who studied rhythmic body movement and laughter in virtual avatar animation. Our work differs from these in three important points. First, our corpus consists of first encounter dialogues which are a specific type of social situation and may have an impact on the interaction strategies due to participants conforming to social politeness norms. We also use a laughter classification developed in our earlier studies (Hiovain and Jokinen, 2016) and standard techniques from OpenCV. Moreover, our goal is to look at the co-occurrence of body movement and laughter behaviours from a novel angle in order to gain insight into how gesturing and laughing are correlated in

human interaction. Finally, and most importantly, we wanted to investigate the relation using relatively simple and standard automatic techniques which could be easily implemented in human-robot applications, rather than develop a novel laughter detection algorithm.

The paper is structured as follows. Section 2 briefly surveys research on body movements and laughter in dialogues. Section 3 discusses the analysis of data, video processing and acoustic features, and presents results. Section 4 draws the conclusion that there is a correlation between laughter and body movements, but also points to challenging issues in automatic analysis and discusses future work.

2. Multimodal data

Gesturing and laughing are important actions that enable smooth communication. In this section we give a short overview of gesturing and laughing as communicative means in the control and coordination of interaction.

2.1 Body Movements

Human body movements comprise a wide range of motions including hand, feet, head and body movements, and their functions form a continuum from movements related to moving and object manipulation in the environment without overt communicative meaning to highly structured and communicatively significant gesturing. Human body movements can be estimated from video recordings via manual annotation or automatic image processing (see below) or measured directly through motion trackers and biomechanical devices (Yoshida et al. 2018). As for hand movements, 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 a visible action that participants distinguish as a movement and is treated as governed by a communicative intent.

Human body movement and gesturing are multifunctional and multidimensional activities, simultaneously affected by the interlocutor's perception and understanding of the various types of contextual information. In conversational situations gestural signals create and maintain social contact, express an intention to take a turn, indicate the exchanged information as parenthetical or foregrounded, and effectively structure the common ground by indicating the information status of the exchanged utterances (Jokinen 2010). For example, nodding up or nodding down seems to depend on the presented information being expected or unexpected to the hearer (Toivio and Jokinen 2012), while the form and frequency of hand gestures indicate if the

referent is known to the interlocutors and is part of their shared understanding (Gerwing and Bavelas 2014, Holler and Wilkin 2009, McNeill 2005). Moreover, co-speech gesturing gives rhythm to speech (beat gestures) and can synchronously occur together with the partner's gesturing, indicating alignment of the speakers on the topic. Although gesturing is culture-specific and precise classification of hand gestures is difficult (cf. Kendon 2004; McNeill 2005), some gesture forms seem to carry meaning that is typical to the particular hand shape. For instance, Kendon (2004) identified different gesture families based on the general meaning expressed by gestures: "palm up" gestures have a semantic theme related to offering and giving, so they usually accompany speech when presenting, explaining, and summarizing, while "palm down" gestures carry a semantic theme of stopping and halting, and co-occur in denials, negations, interruptions and when considering the situation not worthwhile for continuation.

Also body posture can carry communicative meaning. Turning one's body away from the partner is a strong signal of rejection, whereas turning sideways to the partner when speaking is a subtle way to keep the turn as it metaphorically and concretely blocks mutual gaze and thus prevents the partner from interrupting the speaker.

In general, body movements largely depend on the context and the task, for instance a change in the body posture can be related to adjusting one's position to avoid getting numb, or to signalling to the partner that the situation is uncomfortable and one wants to leave. Leaning forward or backward is usually interpreted as a sign of interest to the partner or withdrawal from the situation, respectively, but backward leaning can also indicate a relaxed moment when the participant has taken a comfortable listener position.

Interlocutors also move to adjust their relative position during the interaction. Proxemics (Hall 1966) studies the distance between interlocutors, and different cultures are generally associated with different-sized proximity zones. Interlocutors intuitively position themselves so that they feel comfortable about the distance, and move to adjust their position accordingly to maintain the distance.

2.2 Laughter

Laughter is usually related to joking and humour (Chafe 2003), but it has also been found to occur in various socially critical situations where its function is connected to creating social bonds as well as signalling relief of embarrassment (e.g. Jefferson 1984; Truong and van Leeuwen 2007; Bonin 2016; Hiovain and Jokinen 2016). Consequently, lack of laughter is associated with serious and formal situations where the participants wish to keep a distance in their social interaction. In fact, while laughing is an effective feedback signal that shows the participants' benevolent attitude, it can also function as a subtle means to distance oneself from the partner and from the discussed topics and can be used in a socially acceptable way to disassociate oneself from the conversation.

Vöge (2010) discusses two different positionings of laughter: same-turn laughter, where the speaker starts to laugh first, or next-turn laughter, where the partner laughs first. Same-turn laughter shows to the other participants how the speaker wishes their contribution to be taken and thus allows shared ground to be created. Laughter in the second position is potentially risky as it shows that the partner has found something in the previous turn that is laughable; this may increase the participants' disaffiliation, since the speaker may not have intended that their contribution had such a laughable connotation, and the speakers must restore their shared understanding.

Bonin (2016) did extensive qualitative and quantitative studies of laughter and observed that the timing of laughing follows the underlying discourse structure: higher amounts of laughter occur in topic transition points than when the interlocutors continue with the same topic. This can be seen as a signal of the interlocutors' engagement in interaction. In fact, laughter becomes more likely to occur within the window of 15 seconds around the topic changes, i.e. the participants quickly react to topic changes and thus show their participation and presence in the situation.

Laughter has been widely studied from the acoustic point of view. Although laughter occurrences vary between speakers and even in one speaker, it has been generally observed that laughter has a much higher pitch than the person's normal speech, and also the unvoiced to voiced ratio is greater for laughter than for speech.

Laughter occurrences are commonly divided into free laughter and co-speech laughter, and the latter further into speech-laughs (sequential laughter often expressing real amusement) and speech-smiles (expressing friendliness and a happy state of mind without sound, co-occurring with a smile). Tanaka and Campbell (2011) draw the main distinction between mirthful and polite laughs, and report that the latter accounts for 80% of the laughter occurrences in their corpus of spontaneous conversations. A literature survey of further classifications and quantitative laughter detection can be found in Cosentino et al. (2016).

There are not many studies on the multimodal aspects of laughter, except for Griffin et al. (2013) and Niewiadomski et al. (2015). In the next section we will describe our approach which integrates bounding-box based analysis of body movement with a classification of laughs and emotional states in conversational first encounter videos.

3. Analysis

3.1 First Encounter Data

We use the Estonian part of the Nordic First Encounters video corpus (Navarretta et al. 2010). This is a collection of dialogues where the participants make acquaintance with each other for the first time. The interlocutors do not have any external task to solve, and they were not given any particular topic to discuss. The corpus is unique in its ecological validity and interesting for laughter studies, because of the specific social nature of the activity.

The Estonian corpus was collected within the MINT project (Jokinen and Tenjes, 2012), and it consists of 23 dialogues with 12 male and 11 female participants, aged between 21-61 years. The corpus has manual annotations of the participants' head, hand and body movements as well as laughter occurrences. The annotation for each analysis level was done by a single annotator in collaboration with another one, whose task was to check the annotation and discuss problematic cases until consensus was achieved.

3.2 Laughter annotation

We classify laughter occurrences into free laughs and speech-laughs, and further into subtypes which loosely

relate to the speaker's affective state (see Hiovain and Jokinen 2016). The subtypes and their abbreviations are:

- · b: (breath) heavy breathing, smirk, sniff;
- · e: (embarrassed) speaker is embarrassed, confused,
- · m: (mirth) fun, humorous, real laughter,
- · p: (polite) polite laughter showing positive attitude towards the other speaker
- · o: (other) laughter that doesn't fit in the previous categories; acoustically unusual laughter

The total number of laughs is 530, average 4 per second. The division between free and speech laughs is rather even: 57% of the laugh occurrences are free laughs. However, the different subtypes have unbalanced distribution which may reflect the friendly and benevolent interaction among young adults: 35% are mirthful, 56% are breathy, and only 4% are embarrassed and 4% polite. This can be compared with the statistics reported by Hiovain and Jokinen (2016) on a corpus of free conversations among school friends who know each other well: 29% of their laughs were mirthful, 48% breathy, and a total of 21% embarrassed.

Most people laughed for approximately 0.8 seconds, and the laughing is rarely longer than 2 seconds. Speech-laughs tend to be significantly longer than free laughs (1.24s vs. 1.07s), and mirthful laughs the longest while breathy and polite types were the shortest. The longest type of laugh was embarrassed speech laugh produced by both female and male participants. Figure 1 gives a box plot of the laugher events and their durations, and also provides a visualisation of the total duration of the various laughs.
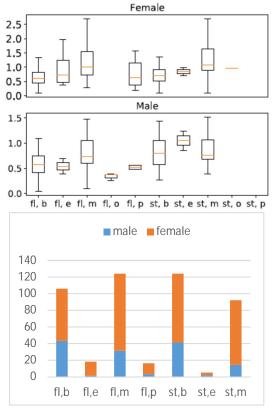

Figure 1. Box plots of the duration of laughter events (upper part) and the total duration of the laughter events (lower part) in seconds, with respect to affective states for male and female speakers. fl = free laugh, st = speech laugh, b= breathy, e = embarrassed, m = mirthful, p = polite, o = other. There were no occurrences of polite or the other speech laughs for males, and polite speech laugh or other free laugh for women.

3.3   Video analysis

To recognize gestures and body movement, we use a variant of the well-known bounding-box algorithm. As described in Vels and Jokinen (2014), we use the edge detector (Canny 1986) to obtain each frame's edges and then subtract the background edges to leave only the person edges. Noise is reduced by morphological dilation and erosion (Gonzales and Woods 2010), and to identify human head and body position coordinates, the contours in the frame are found (Suzuki and Abe 1985), with the two largest ones being the two persons in the scene.

The contours are further divided into three regions for head, body and legs, exploiting the heuristics that the persons are always standing in the videos. The top region of the contour contains the head, and the middle region the torso, arms and hands. The lower region contains the legs, but the contour is unfortunately not very reliable so it is omitted from the analysis. Labelled bounding boxes are drawn around the head, body and leg contours, with a time stamp, as shown in Figure 2. The boxes are labelled LH (left person head), LB (left person body), LL (left person legs) and similarly RH, RB, RL for the right person head, body and legs.
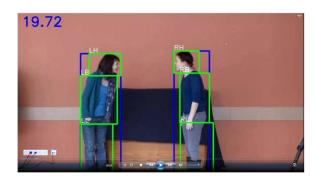

Figure 2 Video frame with bounding boxes for heads, bodies and legs of laughing persons.

In Jokinen et al. (2016) we studied the relation between gesturing and laughter, assuming a significant correlation between laughing and body movements. We experimented with several algorithms (e.g. Linear Discriminant Analysis, Principal Component Analysis, and t-distributed Stochastic Neighbor Embedding), and found that the best results were obtained by Linear Discriminant Analysis (LDA).

By forming a pipeline where data is first transformed using LDA and then used to train a classifier to discriminate between laughs and non-laughs it was possible to get an algorithm which performed decently on the training set. Unfortunately LDA fails to capture the complexity of all the laughing samples, and it seems that certain laughing and non-laughing frames are inherently ambiguous, since all the algorithms mixed them up. It was concluded that laughing bears a relation to head and body movement, but the details of co-occurrence need more studies.

3.4   Laughter and discourse structure

The video annotations show that the interlocutors usually laugh in the beginning of the interaction when greeting each other, and as the conversation goes on, laughing can be an expression of joy or occur quietly without any overt action. Considering the temporal relation between laughter and the evolving conversation, we studied the distribution of laughter events in the overall discourse structure. In

order to provide a comparable laughter timeline among the dialogues, we quantized the time of each laughter event (in seconds), and the position of the laughter was calculated based on its relative position within the utterance. To compensate for the different lengths of the conversations we divided the conversations into five equal stages: Opening, Feedforward, Discuss, Feedback, and Closing, which are the bins that each laughter events is quantized to.

The results of the temporal distribution are depicted in Figure 3. As can be seen, in our corpus openings mostly contain embarrassed speech-laughs, while closings contain breathy free laughs, and discussion mirthful speech-laughs. The feedback part is likely to contain free laughs of embarrassed or mirthful affect, or breathy speech laughs.
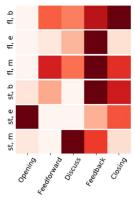

Figure 3 Temporal distribution of the affective laughter types in a dialogue structure consisting of five stages. The laughter abbreviations are as in Figure 1.

3.5 Laughter and acoustic features

Acoustic analysis of laughter is large (see Section 2), and it is only natural to include speech features in the analysis. We tried pitch (acoustic frequency) and MFCC features (mel-frequency cepstral coefficients, short-term power spectrum representation), and noticed that Linear Discriminant Analysis (LDA) can separate non-laugh and laugh signals for both pitch and MFCC, while Principal Component Analysis (PCA) seems to work only for MFCC and pitch features introduce confusion. We processed MFCCs with a 25ms window size, and experimented with different context sizes to capture all necessary information that characterises laughing. We group multiple 25ms windows into larger features called "context windows". For instance, a context length of 10 windows means that we add 5 windows in the past and 5 windows in the future to create a "super vector" feature. The longer the context, the further the non-laugh and laugh events are pushed from each other. In our Estonian experiments, we used MFCC features and a context length of 24 windows.

Figure 4 (left side) visualises how speech laugh is separated from free laugh using LDA on MFCC features with a context length of 10 windows, and the right side shows the same for the more detailed laughter classes with affective states. Concerning the laugh types on the left, speech-laugh can be clearly separated from free-laugh using LDA, and we can see that the laugh types can be recognized given the mixed information of the MFCC and affective states.

The right side of Figure 4 illustrates the difficulty in extracting detailed affective state information from all the laughter annotations. We have highlighted the dense area

of the three most popular affective states: breathy, embarrassed, and mirthful, and their overlapping circles show confusion between the different affective states.

On the other hand, when comparing the left and right sides, we notice that the green zone of speech-laugh on the left matches the turquoise mirth zone on the right. This indicates a strong relationship between speech-laughs and mirthful laughter events. Unfortunately the blue zone of free laugh overlaps with the breathy and embarrassed laugh types, thus indicating a more mixed situation.
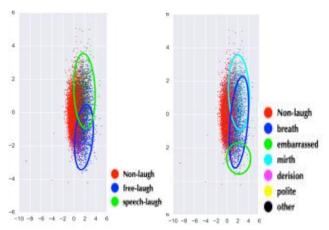

Figure 4 Applying LDA on MFCC features, with rings showing laugh types on the left (blue = free laugh, green = speech laugh) and affective states on the right (blue = breathy, green = embarrassed, turquoise = mirth).

3.6 Laughter and communicative actions

Laughter is a complex behaviour related to the speaker's affective state and interaction with the environment. Body movements and laughter are usually unconscious rather than deliberate actions in this context, although their use in communicative situations can be modelled with the help of temporal segmentation and categorisation. For instance, movements can be described via physical action ontologies related to categorisation of different forms and functions as with hand gestures, and also include internal structure such as preparation, climax, and ending, proposed for gestures as well as laughter events. Unfortunately, the bounding box technique used in this study does not allow detailed gesture analysis so it is not possible to draw inferences concerning e.g. Kendon's gesture families, or co-occurrence of certain types of movement and speech or laughter. For instance, it has been noted that the peak of the gesture coincides with the conceptual focal point of the speech unit (Kendon 2004), and the commonly used audio correlate for gestures, the word, may be too big a unit. Gesture strokes seem to co-occur with vocal stress corresponding to an intonation phrase of a syllable or a mora, rather than a whole word.

In the Gesture-for-Conceptualization Hypothesis of Kita et al. (2017), gestures are generated from the same system that generates practical actions such as object manipulation, but distinguished from them in that gestures represent information. We extend this hypothesis to take the speaker's affective state into consideration, and consider it as the starting point for communication. It leads us to study body movements and laughter, together with spoken utterances and dialogue topics, as actions initiated by the agents based on their affective state, and co-expressively represented by body movements, gesturing, laughter and

spoken utterances. For instance, the cascade model, where the analysis starts from sensory data, integrates the results of decision processes, and finally ends up with a response to the stimuli, has been replaced by a uniform view that regards action control and action representation as two sides of the same coin (Gallese 2000).

When designing natural communication for robot agents, cross-modal timing phenomena become relevant as the delay in the expected synchrony may lead to confusion or total misunderstanding of the intended message (Jokinen and Pelachaud 2013). Manual modelling of the general semantics encoded in the different gesture forms in the robot application as in Jokinen and Wilcock (2014) or in animated agents (André and Pelachaud 2009) is an important aspect in these studies, and can be further deepened by automatic analysis and detection algorithms as in the current study. Body movements and speech flow are closely linked in one's communicative system and between interlocutors in their synchronous behaviour, although the hypothesis of the motor origin of language still remains speculative. An interesting issue in this respect concerns cross-modality annotation categories and the minimal units suitable for anchoring the correlations.

Communicative action generation and annotation are related to the broader issue of the relationship between action and perception in general, and it would be possible to investigate how humans embody the knowledge of communicative gestures via action and interaction, and how communicative gestures are related to acting and observing someone else acting. We can assume that a higher-level control module takes care of the planning and action control, whereas the perception system provides continuous feedback about the selected actions and their effect on the environment. Connections are based on the particular context in which the actions occur, so representations require interpretation of the agent's goals and action purposes for which the act and representation are used in the given context. For instance, extending one's index finger may be executed to point to an object, grasp a mug, rub one's nose, or play with fingers, so the same movement becomes a different action depending on the purpose. Communicative gestures are perceived, learnt, and executed for certain communicative purposes, so perception of a certain type of hand gesture is connected to the assumed communicative action, with the purpose for example to provide information, direct the partner's focus of attention, or stop the partner from speaking.

4. Conclusions and Future Work

We studied laughter and body movements in natural first encounter interactions. The most common laughter type in our Estonian corpus is mirthful, humorous laugh, which includes both free laugh and speech laugh. The longest laughter events are of mirthful types, whereas the polite and breathy laughs were the shortest.

The study gives support for the conclusion that laughing bears a relation to head and body movement, but also highlights the need for accurate and sophisticated movement detection algorithms to capture the complexity of the movements involved in laughing. On the basis of the experiments, it seems that the bounding box approach and the associated speed and acceleration of the movements are too coarse features to infer correlation of the body movements with laughter. For instance, it is not easy to model temporal aspects and intensity of laughter occurrences as they seem to include a complex set of behaviours where body, hand, and head movements play different roles. The bounding box approach collapses all these movements into the two features of velocity and acceleration and is prone to information loss concerning the finer aspects related to body movements and laughter.

On the other hand, the bounding box approach potentially adapts to different settings of the camera angle (front, or sideways recordings of the participants), and it serves well for the particular dataset with the manual annotation of laughter. For instance, we experimented with affective states related to the commonly used emotional descriptions of laughter events (mirthful, embarrassment, politeness), and noticed that these classes can be detected with the bounding box techniques, although there is much confusion between the types.

Due to the roughness of bounding boxes to detect human head and body position we also started to investigate Dense Optical Flow (Brox et al. 2004), which is used for action modelling and has been successfully deployed to action recognition. Compared with Canny edge detector, it does not suffer from dynamic changes in the video frames such as varying lightning conditions, and can thus provide stability and more coverage for different video types. Moreover, it may be possible to use Optical Flow to study specific types of body motion and if they occur during laughter which cannot be captured by frame difference models like bounding boxes.

The work contributes to our understanding of how the interlocutors' body movements are related to their affective state and experience of the communicative situation. From the point of view of interactive system design, a model that correlates the user's affective state and multimodal activity can be an important component in interaction management. It can be used to support human-robot interaction, as the better understanding of human behaviour can improve how the robot interprets the user's laughter or anticipates certain reactions based on the observed body movements. It can also be used as an independent module to determine the robot's own behaviour and to plan more natural responses in terms of gesturing and laughter. Such precise models are valuable when developing the robot agent's capabilities towards natural interaction for practical applications like various care-taking situations in social robotics (Jokinen and Wilcock, 2017).

Future work includes experimenting with larger interaction data and the more recent computer vision methods, and exploring more specific features to associate body movements and laughter. We also plan to upload a more precise model in the robot to experiment with human-robot interactions.

5. Acknowledgements

6. Bibliographical References

E. André and C. Pelachaud: Interacting with Embodied Conversational Agents. In Jokinen, K. and Cheng, F. (Eds.) Speech-based Interactive Systems: Theory and Applications. Springer, 2009.

F. Bonin: Content and Context in Conversations: The Role of Social and Situational Signals in Conversation Structure. PhD thesis, Trinity College Dublin, 2016.

T. Brox, A. Bruhn, N. Papenberg, and J. Weickert: High accuracy optical flow estimation based on a theory for warping. In ECCV 2004.

J. Canny: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679-698, 1986.

W. Chafe: The Importance of Being Earnest. The Feeling Behind Laughter and Humor. Amsterdam: John Benjamins Publishing Company, 2007.

S. Cosentino, S. Sessa, and A. Takanishi: Quantitative Laughter Detection, Measurement, and Classification–A Critical Survey. IEEE Reviews in Biomedical Engineering, 9, 148-162, 2016.

V. Gallese: The inner sense of action: Agency and motor representations. Journal of Consciousness Studies 7:23-40, 2000.

J. Gerwing and J.B. Bavelas: Linguistic influences on gesture's form. Gesture, 4, 157–195, 2004.

R. C. Gonzales and R. E. Woods. Digital Image Processing (3rd edition). Pearson Education, Inc., 2010.

H. J. Griffin, M. S. H. Aung, B. Romera-Parades, G. McKeown, W. Curran, C. McLoughlin and N. Bianchi-Berthouze: Laughter Type Recognition from Whole Body Motion. ACII 2013.

K. Hiovain and K. Jokinen: Acoustic Features of Different Types of Laughter in North Sami Conversational Speech. Proceedings of the LREC-2016 Workshop Just Talking. 2016.

J. Holler and K. Wilkin: Communicating common ground: how mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task. Language and Cognitive Processes, 24, 267-289, 2009.

G. Jefferson: On the organization of laughter in talk about troubles. In: Atkinson, J. Maxwell, J., Heritage, J. eds. Structures of Social Action: Studies in Conversation, 1984.

K. Jokinen: Pointing Gestures and Synchronous Communication Management. In Esposito, A., Campbell, N., Vogel, C., Hussain, A., and Nijholt, A. (Eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. Springer, LNCS Volume 5967, pp. 33-49, 2010.

K. Jokinen and S. Tenjes: Investigating Engagement: Intercultural and technological aspects of the collection, analysis, and use of Estonian Multiparty Conversational Video Data. Proceedings of the Language Resources and Evaluation Conference (LREC-2012). Istanbul, Turkey, 2012.

K. Jokinen, T. Ngo Trung, and G. Wilcock: Body movements and laughter recognition: experiments in first encounter dialogues. Proceedings of the ACM-ICMI Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction (MA3HMI '16), 2016.

K. Jokinen and C. Pelachaud: From Annotation to Multimodal Behaviour. In Rojc, M. and Campbell, N. (Eds.) Co-verbal Synchrony in Human-Machine Interaction. Chapter 8. CRC Press, Taylor & Francis Group, New York. 2013.

K. Jokinen and G. Wilcock: Dialogues with Social Robots – Enablements, Analyses, and Evaluation. Springer, 2017.

K. Jokinen and G. Wilcock: Multimodal Open-domain Conversations with the Nao Robot. In: Mariani et al. (Eds.) Natural Interaction with Robots, Knowbots and Smartphones - Putting Spoken Dialog Systems into Practice. Springer Science+Business Media. pp. 213-224, 2014.

A. Kendon: Gesture: Visible action as utterance. Cambridge University Press, 2004.

S. Kita, M.W. Alibali, and M. Chu: How Do Gestures Influence Thinking and Speaking? The Gesture-for-Conceptualization Hypothesis. Psychological Review, 124(3): 245–266, 2017.

D. McNeill: Gesture and thought. Chicago: University of Chicago Press, 2005.

C. Navarretta, E. Ahlsen, J. Allwood, K. Jokinen, and P. Paggio: Feedback in Nordic first-encounters: a comparative study. Proceedings of 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, 2012.

R. Niewiadomski, M. Mancini, G. Varni, G. Volpe, and A. Camurri: Automated Laughter Detection from Full-Body Movements. IEEE Transactions on Human-Machine Systems, 2015.

R. Niewiadomski, M. Mancini, Y. Ding, C. Pelachaud, and G. Volpe: Rhythmic body movements of laughter. ICMI 2014.

S. Suzuki and K. Abe: Topological structural analysis of digitized binary images by border following. *CVGIP*, 30:32-46, 1985.

H. Tanaka and N. Campbell: Acoustic features of four types of laughter in natural conversational speech. Proceedings of XVIIth ICPhS, Hong Kong, 2011.

E. Toivio and K. Jokinen: Multimodal Feedback Signalling in Finnish. Proceedings of the Human Language Technologies – The Baltic Perspective. Published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License. 2012.

K.P. Truong and D.A van Leeuwen: Automatic discrimination between laughter and speech. Speech Communication, 49(2): 144-158, 2007.

M. Vels and K. Jokinen: Recognition of human body movements for studying engagement in conversational video files. Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication, 2014.

M. Vöge: Local identity processes in business meetings displayed through laughter in complaint sequences. In Wagner, J. and Vöge, M. (Eds.) Laughter in Interaction. Special Issue in the Honor of Gail Jefferson. Journal of Pragmatics, 42/6: 1556-1576, 2010.

Y. Yoshida, T. Nishimura and K. Jokinen: Biomechanics for understanding movements in daily activities. Proceedings of the LREC 2018 Workshop Language and Body in Real Life, Miyazaki, Japan, 2018.