

# Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus

Maria Koutsombogera, Carl Vogel

Trinity College Dublin  
College Green, Dublin 2, Ireland  
{koutsomm, vogel}@cs.tcd.ie

## Abstract

We present a multimodal corpus that has been recently developed within the MULTISIMO project and targets the investigation and modeling of collaborative aspects of multimodal behavior in groups that perform simple tasks. The corpus consists of a set of human-human interactions recorded in multiple modalities. In each interactive session two participants collaborate with each other to solve a quiz while assisted by a facilitator. The corpus has been transcribed and annotated with information related to verbal and non-verbal signals. A set of additional annotation and processing tasks are currently in progress. The corpus includes survey materials, i.e. personality tests and experience assessment questionnaires filled in by all participants. This dataset addresses multiparty collaborative interactions and aims at providing tools for measuring collaboration and task success based on the integration of the related multimodal information and the personality traits of the participants, but also at modeling the multimodal strategies that members of a group employ to discuss and collaborate with each other. The corpus is designed for public release.

**Keywords:** multimodal corpus, multiparty interaction, collaboration, task success, social behavior, personality

## 1. Introduction

The development of multimodal corpora is an essential step in the investigation of the human behavior, as the knowledge reflected in them can lead to models of social behavior to be integrated in intelligent interfaces, making the human-computer interaction as effective, natural and intuitive as possible. Multimodal corpora provide information about the way different expressive modalities shape the structure of the interaction (i.e. turn management) and convey the speakers' cognitive and affective state in any given moment (including feedback responses and emotions), thus demonstrating the speakers' interactional and social behavior (Vinciarelli et al., 2015; Esposito et al., 2015). While two-party interaction is already a rich and informative setup, multiparty interaction is even more challenging because of the dynamics developed among group members (Gatica-Perez et al., 2017).

In recent years, various multimodal and multiparty corpora have been created to study and analyse human behavior and collaboration aspects of it, in two-party (HCRC Map Task Corpus (Thompson et al., 1993)) and in group interaction settings, such as the AMI corpus (Carletta et al., 2005), the Mission Survival Corpus-2 (Mana et al., 2007), the Canal 9 political debates satabase (Vinciarelli et al., 2009) and the Idiap Wolf Corpus (Hung and Chittaranjan, 2010), to name a few. The importance of analyzing such corpora lies in decoding communicative patterns involving verbal and non-verbal modalities. By definition, group dialogue is a canvas where different communicative intentions, personalities, lexical choices that may affect the outcome and the effectiveness of the interaction are manifested by the participants. In terms of behavior modelling, efforts focus on automatically analyzing various facets of group interactions and collecting this knowledge to improve the quality of the interaction either in human-human or in human-machine settings. Related work that exploits group dia-

logue and multiparty corpora studies dominance and leadership (Jayagopi et al., 2009; Nakano and Fukuhara, 2012) personality traits (Mohammadi and Vinciarelli, 2012), deception and competition (Hung and Chittaranjan, 2010).

In this work we discuss the design and implementation of the MULTISIMO corpus, a multimodal corpus consisting of collaborative group interactions where two players need to provide answers to a quiz and are guided by a facilitator. Participants work together while the facilitator monitors their progress and provides feedback and hints when needed. In this setup, collaboration refers to the process where the two players coordinate their actions to achieve their shared goal, i.e. find the appropriate answers and rank them. Collaboration will be measured based on communicative features that may show equal participation, for example the number of turns that each player produces, the number of times persons address their co-players, the mutual turn exchanges, the mutual gazes, etc.; or collaborative turn organization, e.g. in the cases where the facilitator allocates turns to each of the players, or when players address their co-participant to discuss different options instead of providing themselves a direct answer.

Though the development of multimodal and multiparty corpora is not a new domain, this corpus serves to fill in the gap in the investigation of the factors that influence collaboration and task success in a three-party setting and in providing tools for measuring group success. Specifically, the corpus will be exploited to investigate the factors that contribute to collaborative behaviors, e.g. the impact of personality on the participants' behaviors; the impact of dominant participants on the quality of collaboration; participants' attitudes; description and justification of the turn managing practices of the facilitator; multimodal strategies that the group participants employ to collaborate with each other; gender, nationality, language nativeness and familiarity correlations with collaborative behavior. Also, the automatic measurements and manual annotations on the corpus form

materials that can be viewed as the individual speakers profiling and can thus be investigated and revisited at various levels, i.e. linguistic, acoustic, visual, psychological and sociolinguistic.

Our goal is to investigate the way people discuss and collaborate with each other in order to complete a task. We seek to improve the understanding of the structure of collaborative interactions as well as to interpret the behavior of the participants. The corpus will serve as the knowledge base for identifying measurable behavioral and psychological variables of group members with the goal of creating behavioral models. These models may be exploited in human-computer interfaces, and specifically in the design of embodied conversational agents, i.e. agents that need to be able to extract information about their interlocutors to increase the intuitiveness and naturalness of the interaction. The expectations of a user interacting with such interfaces would be that the interaction is ideally as natural as possible. Specifically, users expect that the interface can understand who's talking, recognize signals of human behavior, understand them, coordinate the discussion and address humans in a natural form and at the correct time, as a human would do, facilitating overall collaboration on a common problem (Bohus and Horvitz, 2010; Al Moubayed et al., 2012).

The aforementioned challenges can be pursued if appropriate corpora are available to develop and train models. In this context, the design and implementation of a multimodal, as naturalistic as possible, data collection involving humans engaged in interaction tasks become meaningful in that they help us clarify and answer our research questions. In the next sections we present the experimental design and the data acquisition setup, the data collection, and ongoing work on the corpus analysis.

## 2. Scenario and Experimental Setup

The scenario was designed in a way that would elicit the desired behavior from the participants, that is, encourage their collaboration towards a goal. We thus designed sessions, in which 3 members of a group, 2 players and 1 facilitator, collaborate with each other to solve a quiz. The sessions were carried out in English and the task of the players was to discuss with each other, provide the 3 most popular answers to each of 3 questions (based on survey questions posed to a sample of 100 people), and rank their answers from the most to the least popular. Participants expressed and exchanged their personal opinions when discussing the answers, and they announced the facilitator the ranking once they reached a mutual decision.

The questions were taken from a database related to the Family Feud game.<sup>1</sup> No specific knowledge nor terminology was required to answer the specific questions; instead, the questions were selected so that they would be easy to address for both native and non-native English speakers, trigger the discussion among the players, and also elicit multimodal behavior, such as performing gestures when de-

scribing an object or an idea.<sup>2</sup> After the end of each session the participants filled in a brief questionnaire to assess their impression of the experiment.

Participants were assisted by the facilitator who coordinated this discussion, i.e. provided the instructions of the game and confirmed participants' answers, but also helped participants throughout the session and encouraged them to collaborate. The facilitator role is of key importance in the setup design, considering that it is a role that would be modeled for an embodied conversational agent that would coordinate group interaction and would help participants achieve their goals. In this respect, the facilitator role was designed in a way that will enable the extraction of behavioral cues for the development of an agent responsible for managing the interaction and choosing actions that maximize the collaboration effort and the performance of the group participants.

Participants' recruitment was announced on the online and wall notice boards at Trinity College Dublin (TCD) and via mailing lists. In total, 49 participants were recruited, the majority of them being students or researchers at TCD. 46 were assigned the role of players and were paired in 23 groups. The remaining 3 participants shared the role of the facilitator throughout the 23 sessions. Facilitators were selected in advance and were briefly trained before the actual recordings, i.e. they were given the quiz questions and answers and they were instructed to monitor the flow of the discussion and, if necessary, intervene to help players or to balance their participation. For this role we looked for participants who were teaching or tutoring professionals, who had a pedagogical training and were familiar with instructing and guiding groups of people in completing a task, properties that are desirable for a facilitator to have. To ensure consistency, the facilitators were of the same gender, nationality, professional background and level of English language competence (i.e. female, Greek, English teachers).

### 2.1. Technical Setup

The recording of the sessions took place in the premises of the School of Computer Science and Statistics (SCSS) in a quiet, though not sound-proof, room. The participants were seated around a table. The equipment used includes three HD cameras, one 360 camera, three head-mounted microphones, one omnidirectional microphone and one Kinect 2 sensor. Two of the HD cameras (1920x1080 px, 29.97 fps) were placed opposite each of the two players capturing their front view (Figure 1). The third HD camera (1920x1080 px, 25 fps) was placed opposite the facilitator and captures the whole scene (Figure 2); its zoomed angle is used to isolate the facilitator's front view. The 360 camera (3840x2160 px, 29.97 fps) was placed in the middle of the table to capture the whole scene from a low angle (Figure 3). The head-mounted microphones were recording the individual audio signals (SR 44.1 kHz), while the omnidirectional microphone was used as a backup audio

<sup>2</sup>The 3 questions were: (a) *name a public place where it's likely to catch a cold or a flu bug*; (b) *name 3 instruments you can find in a symphony orchestra*; and (c) *name something that people cut*.

<sup>1</sup><http://familyfeudfriends.arjdesigns.com/>, last accessed 15.02.2018

source (SR 44.1 kHz). Finally, the Kinect 2 was placed in a way that it would perform the skeletal tracking of all participants (Figure 4). The experiment, together with two surveys that participants filled in, i.e. a personality test (cf. section 2.3) and an experience assessment questionnaire, lasted about 40 minutes per participant and each participant was rewarded with a 10 euro token.



Figure 1: Front views of participants.



Figure 2: The setup of a corpus session.



Figure 3: 360 camera angle of a group session.

## 2.2. Ethics Management

Since the data collection involves humans, the experimental process leading to it was supervised by the SCSS ethics committee and followed its ethical standards. Moreover, our aim was that the corpus, i.e. the recorded data together with the survey materials (personality tests and experience assessment questionnaires) will be made open for reuse and repurposing. However, since the core information is located in the audio and video signals, complete data anonymization was not an option. To assure participants' right to

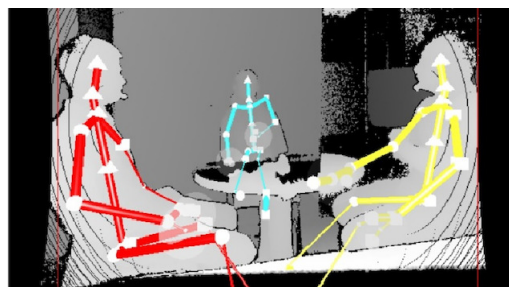


Figure 4: 3D view of Kinect 2 performing skeletal tracking.

control their personal information, participants were fully informed of the overall process and signed the respective consent forms (Koutsombogera and Vogel, 2017). Participants needed to select one of the three options in the consent forms, i.e. to allow the use of their data (a) for the purposes of the specific project, (b) for teaching and academic research purposes, and (c) for sharing in an open access repository.

The majority of the participants gave their consent to their data being publicly available in the future, therefore allowing the sharing of a large part of the corpus after the end of the project. The corpus to be released will be accompanied by an appropriate licensing scheme and will be linked to an open research repository that will provide long-term access.<sup>3</sup>

## 2.3. Personality Traits

One of our aims has been to investigate the effect of personality traits on the task success but also on participants' collaborative behavior, including their engagement, attention and the tendency to create or the ability to manage conversational conflicts. Personality variables are an important tool for the interpretation of social behavior. At the same time it has been widely acknowledged by personality researchers that there is a necessity to have an accepted classification scheme to categorise empirical findings and that the 5-factor model is a robust and meaningful framework enabling the formulation and testing of hypotheses related to individual differences in personality (Goldberg, 1992).

Therefore, out of the wide spectrum of personality measures, we opted for the big five personality traits<sup>4</sup> and especially those that have a communicative value with regard to the interaction behaviour of the speaker, such as conscientiousness and extraversion. Before the recordings participants completed the Big Five Inventory (BFI), a self-report inventory designed to measure the Big Five dimensions (John et al., 1991; John et al., 2008). The test consists of 44 items (statements) and the participants were asked to rate each statement to indicate the extent to which they agree or disagree with it. As a result, a list of scores per

<sup>3</sup>More information and access details will be published at <https://www.scss.tcd.ie/clg/MULTISIMO/>

<sup>4</sup>The five personality traits assessed are: Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience.

personality trait and per participant was created. The percentile rank of each participant across the five personality traits was then calculated using local norms, i.e. the norms were constructed upon the groups population. Details about the calculation of scores and percentiles are included in the corpus documentation.

### 3. Corpus Description

The collected dataset consists of 23 sessions of average duration of 10 minutes (min=6, max=16), resulting in a total duration of approximately 4 hours. The pairing of players was randomly scheduled, and was based on their availability to attend the recordings. While in other corpora the familiarity among group participants is a controlled variable (Thompson et al., 1993), in our case there was no attempt to pair players based on whether they know each other or not. This was decided to ensure that the least number of constraints is imposed on the experimental design, allowing for more flexibility in the formulation of research questions. In most of the sessions the participants don't know each other, although there are a few cases (i.e. in four groups) where the players are either friends or colleagues. The average age of the participants is 30 years old (min=19, max=44). Furthermore, gender is balanced, i.e. with 25 female and 24 male participants. Nevertheless, the gender distribution varies, depending on the pairing of the players. For example, there are groups where both of the players are female, or groups with male players, and groups with both genders. The participants come from different countries and span eighteen nationalities, one third of them being native English speakers. Tables 1 and 2 present the details about the gender of the participants, the number of groups whose players are familiar with each other or not, the number of native and non-native English speakers, as well as their different nationalities.

The corpus includes all files from the different cameras and microphones employed, all fully synchronized. It is also complemented by personality test results, that give evidence of various personality types, as well as by the experience assessment survey ratings. Sample data from the corpus are available for viewing at <https://www.scss.tcd.ie/clg/MULTISIMO/>.

Gender	
Male	24
Female	25
Language	
EN Native	16
EN Non-native	33
Familiarity (groups)	
Familiar	4
Non-familiar	19

Table 1: Distribution of corpus participants per gender, language and familiarity among group players

#### 3.1. Corpus Annotations

The data analysis is based on manual annotations, automatically extracted features and the scores of the surveys used,

#### Participants' Nationalities

Greek	13	Croatian	1
Irish	13	Egyptian	1
French	4	Italian	1
Brazilian	2	German	1
British	2	Kazach	1
Indian	2	Mexican	1
Pakistani	2	Romanian	1
American	1	Slovenian	1
Chinese	1	Thai	1

Table 2: Distribution of participants' nationalities

i.e. personality self-assessments and experience assessments. Manual and automatic annotations form the ground truth on which some components will be trained and tested upon (e.g. automatic personality perception from speech and video).

##### 3.1.1. Manual Annotations

Manual encoding includes annotations on the audio and video streams, as well as perceived assessments, as is the case with experiments involving the other-assessment of personality traits. All corpus sessions are fully transcribed using Transcriber,<sup>5</sup> including segmentation of speaker turns. Video annotation tasks in progress include the labelling of gaze focus, head movements and the communicative functions that these entail, i.e. the semantics of speakers' actions in terms of feedback and turn management. For the last two elements we use the relevant values of the MUMIN coding scheme (Allwood et al., 2007). Video annotation is performed on ELAN<sup>6</sup> editor. Figure 5 presents a sample of an annotated file in ELAN, with information about speech transcription, speaking turns and gaze direction. Finally, manual annotations also include a set of assessments from annotators who are ranking the perceived dominance and perceived personality traits of the sessions participants.

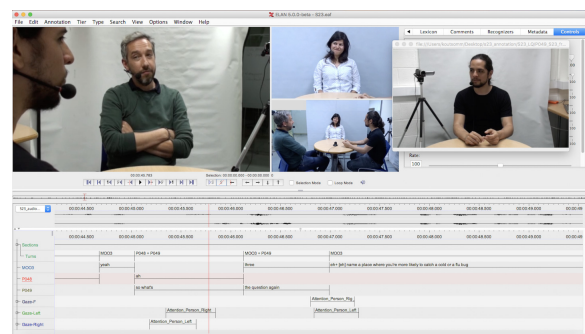


Figure 5: Sample of an annotated file in ELAN (multicamera view).

<sup>5</sup><http://trans.sourceforge.net/> last accessed 15.02.2018

<sup>6</sup><https://tla.mpi.nl/tools/tla-tools/elan/> last accessed 15.02.2018

### 3.1.2. Automatic Measurements

In parallel to the manual transcriptions, that provide information about speaking activity, voice activity and pauses have been automatically extracted. Audio files have been processed with Praat<sup>7</sup> to extract pitch and loudness features. The front view videos of the participants were processed with the Emotient module of iMotions (iMotions A/S, 2016), resulting in a set of measurements for each of seven emotions,<sup>8</sup> the facial action units involved, as well as the (visual) emotion valence. For tasks in progress related to gaze detection from video and personality perception and emotion recognition from audio, the OpenFace (Amos et al., 2016) and openSMILE (Eyben et al., 2013) toolkits are used to extract visual and acoustic features respectively.

## 4. Research Topics in Corpus Analysis

The corpus is considered a source that will provide tools for measuring collaboration and task success in group interactions. In this respect, we are currently investigating the following research questions: (a) the multimodal strategies that the group participants employ to collaborate with each other; (b) the factors that affect success in a collaborative discussion; and (c) behavioral cues as a source of speaker profiling.

### 4.1. Multimodal Turn-taking

The verbal and non-verbal strategies of the speakers (including facial expressions, gestures, body pose and head shifts, speaking times etc.) provide evidence about the structure of turn-taking, but also inform about the degree of participation of the speakers in the discussion, revealing possibly unbalanced participation or signs of dominance from certain speakers. In this particular setting, studying the role of the facilitators becomes significant, in the sense that it's crucial to understand the mechanisms they use to coordinate the interaction and to encourage equal participation of both players. Thus, the focus is on the turn-taking mechanism based on both verbal and non-verbal cues, examining who takes a turn and when, how turn allocation is performed, the number and type of overlaps, as well as on the role of pauses in turn management.

### 4.2. Collaboration and Task Success

We will follow a principled approach to measuring collaboration and task success. Their estimation will be based on measurable features, such as (a) the number of correct answers; (b) the amount of time groups need to complete a task; and most importantly (c) the level of participation of both players in a session, as measured by the number of turns and the duration of their speech activity, but also the amount of mutual gazes and the number of times the players address each other. In this respect, successful completion of the task does not always entail successful collaboration between players. For example, in cases where one of the players is a dominant speaker and manages to answer all questions correctly and quickly, the task may be successful

(in the sense that all answers are provided), however collaboration has not been successful, as the two players have actually not shared their ideas nor participated equally in the discussion.

Interestingly, due to the nature of the game, correctness of answers is mainly a matter of popularity and not suitability. For example, players may express options that are reasonable possible answers to the specific question; however, if they're not the exact, the 3 most popular answers according to the game database, then it is considered that the correct answer is still missing. This criterion (correctness) is thus a subjective one; however, it has been noticed that it encourages the expression of emotions (e.g. surprise, curiosity, fun or frustration) and stimulates the creativity of the players in that they keep guessing and they approach their co-player to collaborate with him/her.

Furthermore, the success of the interaction and the task at hand will be correlated with the participants' personality self-assessment scores at an individual level, but also at group level, that is, combinations of personalities within the groups. The personality test scores will serve as an objective tool to interpret aspects of participants' behavior that relates to their attitudes towards the task but also towards their co-player. In parallel, personality other-assessments will be performed by observers, who will listen to and watch slices of audio and video respectively, and decide upon the participants' big 5 traits. These assessments are an additional source of information, and they will be linked to the extraction of acoustic and visual features that are related to personality, to train relevant models of personality perception.

Conversational dominance is a phenomenon closely linked to collaboration, in the sense that a participant may wish to control or dominate the other player's behavior, resulting in asymmetry in participation. Dominance in our corpus is being computed based on audio (number of turns per speaker, turn length, number of words, number of successful interruptions) and visual (mutual gaze, attention to either speaker) cues. Perception experiments are also being carried out to assess the dominance level of the players, where observers watch the videos and rank participants' dominance levels on a scale from 1 to 5.

### 4.3. Speaker Profiling

The corpus data, the annotations performed either manually or automatically, as well as the assessment scores, give access to a valuable source of information related to the corpus speakers. In this respect, this set of multi-faceted information constitutes the profiling of the speakers that participate in the corpus. Table 3 lists a set of cues related to speakers, that can be accessible from the corpus, its annotation and analysis.

In terms of group composition and demographic data of the speakers, we believe that gender, language nativeness or non-nativeness and familiarity level are important features to be correlated with collaboration and task success, as well as with the interpretation of the participants' social behavior. Speech cues provide important measurements about the overall activity of speakers within the sessions, but also the type of activity, e.g. whether they ask a lot of questions

<sup>7</sup><http://www.fon.hum.uva.nl/praat/> last accessed 15.02.2018

<sup>8</sup>Joy, anger, surprise, fear, sadness, disgust and contempt.

or they show a tendency to dominate the interaction by interrupting their co-participants. Acoustic and visual cues provide quantitative low-level features (e.g. action units) that characterize the speakers and may correlate with annotations or assessments to determine high-level qualitative features (e.g. type of emotion). At a lexical level, the transcripts are a rich source of information that may reveal vocabulary variation (type/token ratio) of speakers, lexical or syntactic patterns they may use, as well as a set of markers that are related to politeness, engagement and co-operation. Therefore, the dataset is profiled according to the aforementioned cues and can be revisited accordingly at various levels and signal- or natural language processing tasks.

<b>Speech</b>	Number of turns Turn length Number of overlaps Number of interruptions Number of questions asked
<b>Acoustic</b>	Pitch and intensity Features related to emotions Features related to personality Features related to dominance
<b>Visual</b>	Emotions Emotion polarity Action units Gaze patterns Head poses
<b>Lexical</b>	Transcripts Number of words per speaker Type/token ratio Discourse markers
<b>Perceptual</b>	Personality other-assessment scores Dominance other-assessment scores
<b>Survey</b>	Personality self-assessment scores Experience self-assessment scores
<b>Demographic</b>	Gender Native language Familiarity Age

Table 3: Available cues for each speaker in the corpus

## 5. Conclusion

In this paper we introduced a new multimodal corpus of group collaborative interactions and we presented work in progress regarding the exploitation of the corpus in the analysis and modeling of human-human group interactions. Considering the fact that the quality of the models depends on the dataset quality, this work was driven by the need to design and implement multimodal, as naturalistic as possible, data collections involving humans engaged in interaction tasks. A suitable amount of data was collected and will enable us to draw conclusions on aspects of collaborative interactions as well as ways of measuring collaboration and task success, based on the exploitation of visual, audio and text features.

The corpus itself is a major output of this research in the sense that a large part of it (i.e. the sessions for which the participants have given their consent) will be publicly available for research purposes, so that it can be reused by other researchers to explore further questions. Also, the documentation regarding the materials used is rich enough so that the recordings experiment can be duplicated, including details about the experimental protocol and the guidelines and training that the facilitators had before the recordings. The corpus release will include audio, video (in high and low resolution formats) and Kinect files, the annotation scheme, manual annotations and automatic measurements, together with the perception experiments and surveys scores and the related documentation.

## 6. Acknowledgements

The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 701621 (MULTISIMO). We would also like to thank the TCD Speech Communication Lab for contributing to the experimental setup.

## 7. Bibliographical References

- Al Moubayed, S., Beskow, J., Skantze, G., and Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In Anna Esposito, et al., editors, *Cognitive Behavioural Systems. Lecture Notes in Computer Science*, pages 114–130. Springer, Berlin, Heidelberg.
- Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., and Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3/4):273–287.
- Amos, B., Ludwiczuk, B., and Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science.
- Bohus, D. and Horvitz, E. (2010). Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI ’10*, pages 5:1–5:8, New York, NY, USA. ACM.
- Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, I., Post, W., Reidsma, D., and Wellner, P. (2005). The AMI meeting corpus: A pre-announcement. In Steve Renals et al., editors, *MLMI’05: Proceedings of the Workshop on Machine Learning for Multimodal Interaction*, number 3869 in LNCS, pages 28–39. Springer-Verlag.
- Esposito, A., Esposito, A. M., and Vogel, C. (2015). Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognition Letters*, 66:41–51.
- Eyben, F., Wenginger, F., Gross, F., and Schuller, B. (2013). Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In *Proceedings of*

- the 21st ACM International Conference on Multimedia, MM '13, pages 835–838, New York, NY, USA. ACM.
- Gatica-Perez, D., Aran, O., and Jayagopi, D. (2017). Analysis of small groups. In Judee K. Burgoon, et al., editors, *Social Signal Processing*, pages 349–367. Cambridge University Press.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26–42.
- Hung, H. and Chittaranjan, G. (2010). The Idiap Wolf Corpus: Exploring group behaviour in a competitive role-playing game. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 879–882, New York, NY, USA. ACM.
- iMotions A/S. (2016). iMotions Biometric Research Platform 6.0. Copenhagen, Denmark.
- Jayagopi, D., Hung, H., Yeo, C., and Gatica-Perez, D. (2009). Modeling dominance in group conversations from non-verbal activity cues. *IEEE Transactions on Audio, Speech and Language Processing*, 17(3):501–513.
- John, O. P., Donahue, E. M., and Kentle, R. L. (1991). The big five inventory versions 4a and 54. Technical report, University of California, Berkeley, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3:114–158.
- Koutsombogera, M. and Vogel, C. (2017). Ethical responsibilities of researchers and participants in the development of multimodal interaction corpora. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 277–282.
- Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., and Zancanaro, M. (2007). Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In *Proceedings of the 2007 Workshop on Tagging, Mining and Retrieval of Human Related Activity Information*, TMR '07, pages 9–14, New York, NY, USA. ACM.
- Mohammadi, G. and Vinciarelli, A. (2012). Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing*, 3(3):273–284.
- Nakano, Y. and Fukuhara, Y. (2012). Estimating conversational dominance in multiparty interaction. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, pages 77–84, New York, NY, USA. ACM.
- Thompson, H. S., Anderson, A., Bard, E. G., Doherty-Sneddon, G., Newlands, A., and Sotillo, C. (1993). The HCRC map task corpus: Natural dialogue for speech recognition. In *Proceedings of the Workshop on Human Language Technology*, HLT '93, pages 25–30, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Vinciarelli, A., Dielmann, A., Favre, S., and Salamin, H. (2009). Canal9: A database of political debates for analysis of social interactions. In *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–4.
- Vinciarelli, A., Esposito, A., André, E., Bonin, F., Chetouani, M., Cohn, J. F., Cristani, M., Fuhrmann, F., Gilmartin, E., Hammal, Z., et al. (2015). Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions. *Cognitive Computation*, 7(4):397–413.