PART ONE

CHAPTERS

1
Analysis of Small Groups

1.1 Introduction

Teams are key components of organizations, and although complexity and scale are typical features of large institutions worldwide, much of the work is still implemented by small groups. The small-group meeting, where people discuss around the table, is pervasive and quintessential of collaborative work. For many years now, this setting has been studied in computing, with the goal of developing methods that automatically analyze the interaction using both the spoken words and the nonverbal channels as information sources. The current literature offers the possibility of inferring key aspects of the interaction, ranging from personal traits to hierarchies and other relational constructs, which in turn can be used for a number of applications. Overall, this domain is rapidly evolving and studied in multiple sub-disciplines in computing and engineering as well as the cognitive sciences.

We present a concise review of recent literature on computational analysis of face-to-face small group interaction. Our goal is to provide the reader with a quick pointer to work on analysis of conversational dynamics, verticality in groups, personality of group members, and characterization of groups as a whole, with a focus on nonverbal behavior as information source. The value of the nonverbal channel (including voice, face, and body) to infer high-level information about individuals has been documented at length in psychology and communication (Knapp and Hall (2009)) and is one of the main themes of this volume.

In the chapter, we include pointers to 100 publications appearing in a variety of venues between 2009 and 2013 (discussions about earlier work can found e.g. in Gatica-Perez (2009).) After a description of our methodology (Section 1.2) and a basic quantitative analysis of this body of literature (Section 1.3), we select a few works, due to the limited space, in each of the four aforementioned trends to illustrate the kind of research questions, computational approaches, and current performance available in the literature (Sections 1.4-1.7). Taken together, the existing research on small group analysis is diverse in terms of goals and studied scenarios, relies on state-of-the-art techniques for behavioral feature extraction to characterize group members from audio, visual, and other sensor sources, and is still largely using standard machine learning techniques as tools for computational inference of interaction-related variables of interest. In Section 1.8, we conclude the chapter by providing a few words about what the future can bring in this domain.

1.2 Methodology

For this review, we limited the search for literature on the topic by the following conditions:

1. Publications written in English from 2009 to 2013 (previous surveys cover older literature (Gatica-Perez (2009)).
2. Papers strictly covering small groups, i.e., involving between three and six conversational partners where all of them are human. This condition therefore excludes literature using robots and agents interacting with people, and literature involving only individuals (e.g. lectures or self-presentations), dyads, and large groups.
3. Papers where strictly co-located, face-to-face interactions are studied. This restriction thus leaves aside literature on computer-mediated communication.
4. Papers where some form of sensor processing is done (e.g. audio, video, or motion). This conditions thus excludes papers that focus on analysis using only transcribed speech.
5. Original research work, rather than other review papers or that summarize or revisit existing work.

With the above restrictions, a wide but non-exhaustive search of the literature (using a combination of web searches for terms like "small-group" and "multi-party" and publication venue-specific searches) was conducted in the summer of 2013 and resulted in 100 papers, including 25 journal papers and 75 conference/workshop papers. We then defined seven classification areas that span most of the publication venues where work in computational analysis of small groups with the above restrictions can found. The areas include Audio, Speech, and Language (ASL, including venues like IEEE T-ASLP, ICASSP, Interspeech), Computer Vision (CV, with venues like IVC, CVIU, CVPR, ICCV), Multimodal and Multimedia processing (MM, including venues like IEEE T-MM, ICMI, MM), Human Computer Interaction (HCI, with venues like CHI), Pattern Recognition and Machine Learning (PR, including venues like IEEE T-PAMI, PR, PRL, ICPR), Behavioral, Affective, and Social (BAS, with venues like IEEE T-AC, ACII, SocialCom, SSPW, HBU), and Other (catching publications that could not be clearly associated to any of the previous categories) [1].

1.3 Analysis of main trends

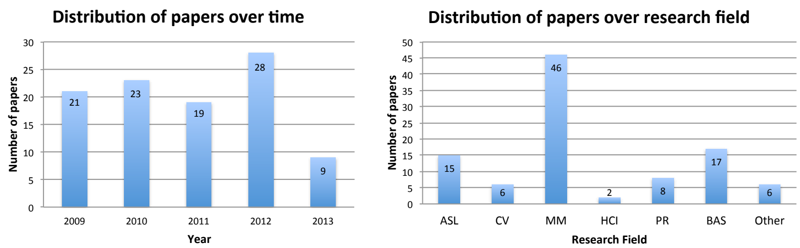
We analyze the trends based on the 100 technical references on small group analysis that we have found based on the methodology described in Section 1.2.

Figure 1.1(b) shows the distribution of the publications over time. The number of publications on small group analysis seem to be stable between 2009 - 2012 with around 20 publications per year. The figures for 2013 is incomplete due to the date that this review was done. In comparison to the period between 2001-2009, reported in Gatica-Perez (2009), we see that there is an increase in the number of publications, around 10 more publications per year, since 2009.

In Figure 1.1(a), we show the distribution of the papers per research field. Almost half of the papers appeared in venues related to multimodal and multimedia processing (column labeled MM in Figure 1.1(a)). This effect might be partly biased by the active participation of the authors' institution in these specific communities, but in general it should be seen as a community effect. Roughly tied in the second place are ASL and BAS. It is interesting that, while ASL is a classic domain, BAS corresponds to publication venues that did not exist before 2009. In comparison to the research disciplines covered for older work (e.g., reviewed in Gatica-Perez (2006, 2009)), we see that more papers are published in multimodal/multimedia venues, and that new venues emerge in parallel to the growing interest on the analysis of social behavior in general and of small groups in particular.

The collected papers investigate the small group interaction based on

[1] For space reasons, we only provide the acronyms for each publication venue, but we anticipate that the reader will be familiar with most of them



(a) (b)

Figure 1.1 Statistics of the 100 technical references on small group analysis reviewed in this paper. (b) Distribution of papers over time and (a) Distribution of papers over research field in journals, conferences and workshops: audio, speech, language (ASL), computer vision (CV), multimodal and multimedia processing (MM), human computer interaction (HCI), pattern recognition and machine learning (PR), behavioral, affective, social (BAS), other.

audio and/or visual recordings through both verbal and nonverbal cues, with a majority of them focusing on nonverbal cues only. As the review did not include venues on NLP, the analysis of text based interaction in small groups can be underrepresented.

We discuss the analysis of small groups in four categories of social constructs, i.e., conversational dynamics, personality, roles-dominance-leadership, and group level analysis. The first three categories look at the social constructs of individuals in a small group setting. In the fourth category, we review papers that focus on the group as a whole, rather than the individuals in the group. In Table 1.1 we list the technical references considered in this paper, grouped in these four categories.

1.4 Conversational dynamics

Conversations in groups involve multiple channels of communication and complex coordination between the interacting parties. The communication process involves taking turns, addressing someone, yielding the floor, and gesturing using head and hand to communicate or acknowledge. Over a decade or so, several works have appeared to extract these basic conversational signals and analyze them further to study the turn-taking, gazing, and gesturing behavior in small groups.

Cristani et al. (2011) present a novel way of analyzing turn-taking patterns by using a GMM model on durations of Steady Conversational

Table 1.1 *List of references for small group analysis in four main categories*

Conversational dynamics	Ba and Odobez (2009); Baldwin et al. (2009); Bohus and Horvitz (2009); Bousmalis et al. (2009); Chen and Harper (2009); Germesin and Wilson (2009); Ishizuka et al. (2009); de Kok and Heylen (2009); Kumano et al. (2009); Lepri et al. (2009b); Otsuka et al. (2009); Vinciarelli (2009); Bachour et al. (2010); Gorga and Otsuka (2010); Subramanian et al. (2010); Sumi et al. (2010); Valente and Vinciarelli (2010); Voit and Stiefelhagen (2010); Ba and Odobez (2011a,b); Bohus and Horvitz (2011); Bousmalis et al. (2011); Campbell et al. (2011); Cristani et al. (2011); Kumano et al. (2011); Wang et al. (2011); Angus et al. (2012); Bruning et al. (2012); Debras and Cienki (2012); Kim et al. (2012a,b); Noulas et al. (2012); Otsuka and Inoue (2012); Pesarin et al. (2012); Prabhakar and Rehg (2012); Rehg et al. (2012); Song et al. (2012); Vinyals et al. (2012); Bousmalis et al. (2013a,b)
Verticality and Roles	Favre et al. (2009); Raducanu and Gatica-Perez (2009); Salamin et al. (2009); Aran and Gatica-Perez (2010); Aran et al. (2010); Charfuelan et al. (2010); Escalera et al. (2010); Glowinski et al. (2010); Hung and Chittaranjan (2010); Poggi and D'Errico (2010); Raducanu and Gatica-Perez (2010); Salamin et al. (2010); Sanchez-Cortes et al. (2010); Valente and Vinciarelli (2010); Varni et al. (2010); Charfuelan and Schroder (2011); Hung et al. (2011); Kalimeri et al. (2011); Raiman et al. (2011); Sanchez-Cortes et al. (2011); Schoenenberg et al. (2011); Vinciarelli et al. (2011a,b); Wilson and Hofer (2011); Feese et al. (2012); Hadsell et al. (2012); Kalimeri et al. (2012); Nakano and Fukuhara (2012); Salamin and Vinciarelli (2012); Sanchez-Cortes et al. (2012a,b); Wöllmer et al. (2012); Wang et al. (2012); Dong et al. (2013); Ramanathan et al. (2013); Sapru and Bourlard (2013); Suzuki et al. (2013)
Personality	Lepri et al. (2009a, 2010a,b); Staiano et al. (2011a,b); Lepri et al. (2012); Aran and Gatica-Perez (2013a,b); Pianesi (2013)
Group level analysis	Camurri et al. (2009); Dai et al. (2009); Jayagopi and Gatica-Perez (2009); Jayagopi et al. (2009a); Kim and Pentland (2009); Dong and Pentland (2010); Hung and Gatica-Perez (2010); Jayagopi and Gatica-Perez (2010); Subramanian et al. (2010); Woolley et al. (2010); Bonin et al. (2012); Dong et al. (2012a,b); Jayagopi et al. (2012); La Fond et al. (2012)

Period (SCP) and then use the discrete cluster classes as observed states of an influence model (Basu et al. (2001)). These low level features are shown to be indeed useful to capture the conversational dynamics by using them to improve the state-of-the-art for classifying roles in group meetings. The authors also argue that SCPs are better than prosodic or phonetic features, used in state-of-the-art algorithms for speech analysis. In this paper, results on role classification on the AMI dataset (Carletta et al. (2005)) are shown to improve w.r.t. an existing baseline with a final accuracy of 90%, with 98 meetings to train, 20 to validate, and 20 to test. The generative approach proposed in the paper has applications in turn-taking decisions for multi-party embodied conversational agents.

Angus et al. (2012), on the other hand, approach the problem of modeling the coupling in human-human communication by quantifying multi-participant recurrence. The words spoken in utterance by a participant are used to estimate the coupling between utterances, both from the same participant as well as other participants. The work proposes a set of multi-participant recurrence metrics to quantify topic usage patterns in human communication data. The technique can be used to monitor the level of topic consistency between participants; the timing of state changes for the participants as a result of changes in topic focus; and, patterns of topic proposal, reflection, and repetition. Finally, as an interesting test case, the work analyzes a dataset consisting of a conversation in a aircraft, involved in an emergency situation. Some of the studied metrics include short-term and long-term topic introduction, repetition, and consistency. The participants involved in this dataset included the captain, first officer, jumpseat captain, ground staff, and others.

Baldwin et al. (2009) study communicative hand gestures for coreference identification, for example, when someone says “you want this” and gestures at a certain speaker, to automatically infer the intention of the speaker, and to undertand whom ‘you’ refers to in this multiparty context. They approach this problem by first formulating a binary classification task to determine if a gesture is communicative or not. Then, every communicative gesture is used to identify if two different linguistic referring expressions actually refer to the same person or object. A diverse set of features that included text, dialogue, and gesture information were used for this task. For this study, a total of six meetings from the Augmented Multi-party Interaction (AMI) data were used with 242 annotated gestures and 1790 referring expressions. The results show that the best accuracy to classify if a gesture is communicative or not, is close to 87% and features such as the duration of the gesture are useful. Also, gestures are shown to improve the performance of coreference identification.

Taken together, the recent literature on conversational analysis show that this area is active, and many open issues remain towards a holistic understanding of the conversational processes. Extracting and analyzing nonverbal and verbal behavior is the basis for all subsequent inferences about individuals and groups.

1.5 Verticality

A second trend in the literature relates to aspects of structure in groups, whether vertical (which in the social psychology literature (Hall et al. (2005)) includes aspects like dominance, status, and leadership) or not (e.g. structure defined by specific roles played by the group members.) In this section, we discuss a few representative works focused on the vertical dimension of interaction, more specifically dominance and leadership (for space reasons, we omit discussions on other aspects of structure like roles, mentioned in Table 1.1.) Dominance can be seen as a manifest act to control others, as a personality trait that elicits such behavior, or as a control behavioral outcome (Dunbar and Burgoon (2005)). Leadership, on the other hand, includes both emerging phenomena and styles related to managing and directing a team (Stein (1975)).

Dominance in small groups was originally studied in computing in the mid 2000s, in works like Rienks and Heylen (2005) and Hung et al. (2007). In the last five years, this line of research has been expanded, among others, by Charfuelan et al. (2010). This particular work used the popular Augmented Multi-party Interaction (AMI) scenario meeting data. The AMI data corresponds to five-minute slices of four-person meetings involving people playing a role-based design scenario. A subcorpus was originally annotated for perceived dominance rankings (from most to least dominant) in Jayagopi et al. (2009b). The goal in Charfuelan et al. (2010) was to investigate whether certain prosody and voice quality signals would be characteristic of most and least dominant individuals. Using a variety of acoustic cues extracted from close-talk microphones and using principal component analysis, the study found that most dominant people tend to speak "louder-than-average voice quality" and, conversely, least dominant people speak with "softer-than-average voice quality". It is important to notice that rather than trying to automatically classify most and least dominant people, this study was interested in identifying acoustic cues useful to synthesize expressive speech corresponding to such social situations. In a subsequent work Charfuelan and Schroder (2011), the same authors applied a similar methodology to two constructs other than dominance, namely speaker roles and dialog acts.

A second construct of interest is leadership. We discuss two variations of this theme found in the recent literature. The first one is emergent leadership, a phenomenon occurring among people who are not acquainted previously to an interaction, in which one of the interactors raises among the others through the interaction itself. One of the first published works is Sanchez-Cortes et al. (2012a), which proposed to identify the emergent leader in a three- to four-person group using a variety of nonverbal cues, including prosody, speaking activity (extracted from a commercial microphone array), and visual activity (estimated from webcam video). The setting is the Winter Survival task - a well known design in psychology to study this phenomenon, in which participants are asked to identify their leader as part of administered questionnaires. Using standard machine learning techniques, this work reported between 72 and 85% of correct identification of the emergent leader on a corpus of 40 group meetings (148 subjects) for various modalities and classification techniques. Through the analysis of the questionnaires, this work also found a correlation between the perception of emergent leadership and dominance.

The other variation is that of leadership styles, which was studied in Feese et al. (2012). Specifically, two contrasting styles in terms of how the leader interacts with the team (authoritarian or considerate) were elicited in a simulated staff selection scenario involving three-person groups, with one of them being the designated leader. A corpus of 44 group discussions was recorded with sensor-shirts. i.e., shirts equipped with Inertial Measurement Units (IMUs) containing accelerometer, gyroscope, and magnetic field sensors. A number of nonverbal body cues were manually annotated and extracted from the IMU sensors, including some measures of behavioral mimicry. This work did not attempt to classify leadership styles, but rather to identify nonverbal features that were significantly different between the two types of leaders. As main results, it was found that authoritarian leaders tend to move their arms more often than considerate ones, and that considerate leaders imitate posture changes and head nods of team members more often than authoritarian ones.

The three examples discussed above show the active interest in understanding and discriminating social constructs related to verticality. It also shows that while current results are promising, additional work is needed to replicate and validate these findings in other settings. A variable closely related to social verticality is the personality of team members. As a closely connected subject, this is discussed in the next section.

1.6 Personality

The automatic analysis of personality has been addressed in a number of works in social computing literature in the last decade. While most works have looked at self-presentations where the individual is the only interacting person, few works also looked at predicting personality of individuals when they interact with others in small groups. The Big-Five model has been the most commonly used model, which factors personality into five different traits (extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience). Among these traits, extraversion has been the one relatively easier to predict, especially in conversational settings. Several audio-visual nonverbal cues have been used and shown to be relatively effective in inferring extraversion. The inference problem can be either formulated as a regression task based on the personality trait scores or as a classification task by quantizing the scores into two or more classes. For the ground truth annotation of personality, current works either use self-reported personality (i.e. the personality of an individual as seen by self) or externally observed (i.e. how the individual is seen by others, also known as impressions.)

Lepri et al. (2012) investigated the automatic classification of the extraversion trait based on meeting behavior, such as speaking time and social attention. They used self-reported personality annotations and investigated the effect of speaking time and social attention as indicators of extraversion based on a thin-slice analysis. Their approach achieved a maximum accuracy of 69% using manually extracted features, and 61% using automatically extracted features with a support vector machine classifier. Their results show that for predicting extraversion, in addition to the target's behavior, the behavior of others in the group should be taken into account. The speaking time or the attention of the target alone did not yield significant accuracies. Besides studying social context in the form of others' behavior, the authors also investigated whether the group composition had any affect on the classification accuracy. They found no significant difference between group variance, and thus concluded that accuracy variability is entirely due to differences among subjects.

Recently, Aran and Gatica-Perez (2013a) studied the inference of personality in cross-domain settings. While collecting data that contains natural face-to-face interaction in everyday life is challenging, social media sites provide a vast amount of human behavioral data. In this study, the authors investigate a cross-domain setting where they used conversational vlogs downloaded from YouTube as the source domain, and video recordings of individuals taken from a small group meeting as the target domain, with personality annotations obtained from external observers. Their results show that, for predicting the extraversion trait, a model learned over body activity cues on conversational vlog data can be useful in a transfer learning setting with face-to-face interaction in small groups as the target domain. The method achieves up to 70% of accuracy in a binary extraversion classification task, by using the source domain data and as few as ten examples from the target data with visual only cues.

While personality is considered to be a stable characteristic of a person, the behavior of people is variable. Although one approach is to consider this variability as noise, another approach would be to use this information to better understand the relationship between personality and behavior. Pianesi (2013) discusses this fact and suggests the characterization of behavior of people in the form of personality states, representing each personality dimension as a distribution of these states. On a similar point, recently, Aran and Gatica-Perez (2013b) also investigated whether thin-slice personality impressions of external observers generalize to the full-meeting behavior of individuals, using a computational setting to predict trait impressions.

In summary, many recent works on the automatic analysis of personality in small groups have focused on the inference of personality of an individual interacting in a group of people, and investigated links between the personality of an individual and the behavior of the other group members. Another research problem is how the characteristics of individuals can affect group formation and interaction. In the next section, we review works that conceptualize groups as units and characterize a group based on the collective behavior of its members.

1.7 Group Characterization

The last thread of work discussed in the chapter is the modeling of collective aspects of groups. A seminal work in this direction is the work on Collective Intelligence by Woolley et al. (2010), which showed how the emergent properties of group intelligence is quite different from the intelligence of the group members. Collective Intelligence is a factor that explains why some groups that do well on a certain task are good at many other tasks (similar to the general intelligence factor of individuals.) The authors show that the Collective Intelligence of a group is uncorrelated with the average or maximum intelligence of the group members. On the contrary, it is shown to be correlated with the communication patterns of the group (particularly egalitarian turn-taking) and the compostion of the group (specifically group with socially-sensitive individuals / more females). This study was conducted with 107 groups, involving 699 people. Wearable badges were used for sensing on a subset of the full dataset (46 groups), particularly to compute the speaking turn distribution. The group tasks were selected from the McGrath Task Circumplex, which included brainstorming, planning, and reasoning tasks. This work establishes the role of group communication on group performance.

With a different goal, Jayagopi et al. (2012) explored the relationship between several automatically extracted group communication patterns and group constructs such as group composition, group interpersonal perception, and group performance. The work proposed a way of characterizing groups by clustering the extracted looking and turn-taking patterns of a group as a whole. The work defined a bag of nonverbal patterns (Bag-of-NVPs) to discretize the group looking and speaking cues. The clusters learnt using the Latent Dirichlet Allocation (LDA) topic model (Blei et al. (2003)), were then interpreted by studying the correlations with the group constructs. Data from eighteen four-people groups was used in this study (a subset of the Emergent Leadership (ELEA) corpus Sanchez-Cortes et al. (2010).) The groups were unacquainted and performed the Winter Survival task. Big-Five Personality traits were used to characterize group composition. Group interpersonal perception questionnaires measured dominance, leadership, liking, and competence. The Survival Task also generates a measure of performance for each group. Several insights about groups were found in this study. The work showed groups with top-two person hierarchy, participated less, while groups without this hierachy participated more. Introverted groups looked at the meeting table more often. Finally, groups which were known to perform better on the task had a competent person as part of their team, and also had more converging gaze on this person during their interaction.

La Fond et al. (2012) approached the problem of group characterization by analyzing who-replies-to-whom patterns, which were manually transcribed. Groups were classified as hub, outlier, and equal types. Similarly, individuals were assigned hub, spoke, outlier, and equal roles. Interestingly, those individuals identified as hub were more assertive, while outliers were not. The groups consisted of three to four individuals solving logic problems. They participated in two phases. The first phase was a distributed session (an online chat session) and the second phase was a face-to-face interaction. In the distributed phase, there were 79 groups of size three and 48 groups of size four, while the face-to-face phase had 27 groups of size three and 35 groups of size four. After the session, the participants evaluated the traits and performance of each member (including themselves), as well as the performance of the group as a whole. The participants evaluated the performance of the group, which included ratings on group cohesion, effectiveness, productivity, trustworthiness, and satisfaction. Models to predict these group evaluation measures using Linear Regression and Decision Trees were learned and tested. The results showed that group effectiveness and trust could be predicted with above 80% accuracy using a Decision Tree classifier.

As a final example, Hung and Gatica-Perez (2010) focused on estimating group cohesion using turn-taking and motion analysis. The work defined several turn-taking features. Compressed-domain motion activity features, which are computationally lighter as compared to pixel-domain features were used to define analogous ”motion turn-taking” features. Group cohesion, unlike the work by La Fond et al. (2012), was defined through external perception or impressions. For the study, 120 two-minute slices of four-people group interaction of the AMI corpus were used. Three annotators answered 27 questions about social and task cohesion. After an analysis of inter-annotator agreement, 61 two-minute slices with 50 high-cohesion score and 11 low-cohesion score with sufficient agreement were used for classification experiments. Accuracies of the order of 90% was achieved on this cohesion classification task.

Overall, the automatic characterization of groups as units is an area for which we anticipate more work in the future, as many open issues need to be addressed. One of them is the need to significantly increase the size of the data sets used for analysis in order to reach more significant conclusions. A second issue is the need to generalize the initial results discussed here across conversational contexts or even cultures. Another direction is in studying nonverbal and verbal behavioral differences between collocated and distributed groups (as in La Fond et al. (2012)), as remote groups interactions have become commonplace. This direction would obviously have links to the literature in Computed-Supported Collaborative Work (CSCW).

1.8 Conclusions and Outlook

In this chapter, we presented a succinct review of the literature on face-to-face small group interaction analysis. From an initial pool of a hundred papers published in the 2009-2013 period, we selected a number of works that illustrate four of the main research trends (conversational dynamics, verticality, personality, and group characterization.) We then briefly discussed the kind of research tasks and approaches that have been proposed using a few illustrative examples for each trend. The review shows that the body of research has grown in numbers in comparison to the previous decade, that it has diversified in terms of goals, and that approaches have gained sophistication in terms of methods to extract behavioral features. In contrast, recent research has made relatively less progress with respect to new computational modeling tools for recognition and discovery tasks: most of the existing work still uses relatively standard machine learning methodologies for automatic inference.

We have argued elsewhere (see Gatica-Perez et al. (2012)) that the future of this area will be shaped by progress along two axes: sensing and modeling. Sensing, literally and metaphorically speaking, is in front of our eyes: smartphones, wearable devices like Google Glass, Android Wear, and Samsung Galaxy Gear, and gaming platforms like Microsoft's XBox One will all give the possibility of sensing interaction quasi-continuously and with higher degree of accuracy than currently possible. While the sensing functionalities will continue to advance, a fundamental point for practical applications is acceptability, both individual and social. There are (and there should be) ethical and legal bounds to recording interaction data. These limits, however, are often not consistent across countries or often not respected; the many stories in the media about privacy intrusion certainly point in the wrong direction. We anticipate privacy to become a much larger research issue in group interaction analysis in the near future.

The second axis is modeling. The possibility of recording interaction in real situations, as enabled by new sensing platforms, will call for methods that integrate both the temporal dimension and the new data scales that will be generated. Regarding time, essentially all of the work discussed in this chapter has examined short-lived interactions, although we know that teams in the real world do not work that way. Methods that are capable of discovering how teams in organizations perform and evolve over weeks, months, or years are needed and likely to appear in the future (existing examples include Olguin et al. (2009); Do and Gatica-Perez (2011)). As a second issue, data scale should also boost new ways of thinking about small group research, moving beyond the small-data-for-small-groups current research trends. It is not hard to anticipate that a big data version of small group research will emerge given the combination of new sensing and modeling methodologies.

1.9 Acknowledgments

We thank the support of the Swiss National Science Foundation (SNSF) through the NCCR IM2, the Sinergia SONVB project, and the Ambizione SOBE (PZ00P2-136811) project.

References

Angus, D., Smith, A.E., and Wiles, J. 2012. Human Communication as Coupled Time Series: Quantifying Multi-Participant Recurrence. *Audio, Speech, and Language Processing, IEEE Transactions on*, **20**(6), 1795–1807.

Aran, O., and Gatica-Perez, D. 2010. Fusing audio-visual nonverbal cues to detect dominant people in small group conversations. In: *20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey*.

Aran, O., Hung, H., and Gatica-Perez, D. 2010. A Multimodal Corpus for Studying Dominance in Small Group Conversations. In: *LREC workshop on Multimodal Corpora, Malta (LREC MMC'10)*.

Aran, Oya, and Gatica-Perez, Daniel. 2013a. Cross-Domain Personality Prediction: From Video Blogs to Small Group Meetings. In: *ICMI*.

Aran, Oya, and Gatica-Perez, Daniel. 2013b. One of a Kind: Inferring Personality Impressions in Meetings. In: *ICMI*.

Ba, Sileye O., and Odobez, Jean-Marc. 2011a. Multiperson Visual Focus of Attention from Head Pose and Meeting Contextual Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 101–116.

Ba, S.O., and Odobez, J.M. 2009. Recognizing visual focus of attention from head pose in natural meetings. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, **39**(1), 16–33.

Ba, S.O., and Odobez, J.M. 2011b. Multi-person visual focus of attention from head pose and meeting contextual cues. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **33**(1), 101–116.

Bachour, K., Kaplan, F., and Dillenbourg, P. 2010. An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *Learning Technologies, IEEE Transactions on*, **3**(3), 203–213.

Baldwin, Tyler, Chai, Joyce Y., and Kirchhoff, Katrin. 2009. Communicative gestures in coreference identification in multiparty meetings. Pages 211–218 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Basu, Sumit, Choudhury, Tanzeem, Clarkson, Brian, and Pentland, Alex. 2001. Learning human interactions with the influence model. NIPS.

Blei, D.M., Ng, A.Y., and Jordan, M.I. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research*, **3**, 993–1022.

Bohus, Dan, and Horvitz, Eric. 2009. Dialog in the open world: platform and applications. Pages 31–38 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Bohus, Dan, and Horvitz, Eric. 2011. Decisions about turns in multiparty conversation: from perception to action. Pages 153–160 of: *Proceedings of the 13th international conference on multimodal interfaces*. ICMI '11. New York, NY, USA: ACM.

Bonin, F., Bock, R., and Campbell, N. 2012. How Do We React to Context? Annotation of Individual and Group Engagement in a Video Corpus. Pages 899–903 of: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*.

Bousmalis, K., Mehu, M., and Pantic, M. 2009. Spotting agreement and disagreement: A survey of nonverbal audiovisual cues and tools. Pages 1–9 of: *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*.

Bousmalis, K., Morency, L., and Pantic, M. 2011. Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition. Pages 746–752 of: *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*.

Bousmalis, Konstantinos, Zafeiriou, Stefanos, Morency, Louis-Philippe, and Pantic, Maja. 2013a. Infinite Hidden Conditional Random Fields for Human Behavior Analysis. *IEEE Trans. Neural Netw. Learning Syst.*, **24**(1), 170–177.

Bousmalis, Konstantinos, Mehu, Marc, and Pantic, Maja. 2013b. Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing*, **31**(2), 203 – 221.

Bruning, Bernhard, Schnier, Christian, Pitsch, Karola, and Wachsmuth, Sven. 2012. Integrating PAMOCAT in the research cycle: linking motion capturing and conversation analysis. Pages 201–208 of: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ICMI '12. New York, NY, USA: ACM.

Campbell, N., Kane, J., and Moniz, H. 2011. Processing YUP! and other short utterances in interactive speech. Pages 5832–5835 of: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*.

Camurri, A., Varni, G., and Volpe, G. 2009. Measuring entrainment in small groups of musicians. Pages 1–4 of: *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*.

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., et al. 2005 (Jul.).

The AMI meeting corpus: A pre-announcement. In: *Proceedings of MLMI Workshop*.

Charfuelan, Marcela, and Schroder, Marc. 2011. Investigating the Prosody and Voice Quality of Social Signals in Scenario Meetings. Pages 46–56 of: Da Mello, Sidney, Graesser, Arthur, Schuller, Bjorn, and Martin, Jean-Claude (eds), *Affective Computing and Intelligent Interaction*. Lecture Notes in Computer Science, vol. 6974. Springer Berlin Heidelberg.

Charfuelan, Marcela, Schroder, Marc, and Steiner, Ingmar. 2010 (Sep). Prosody and voice quality of vocal social signals: the case of dominance in scenario meetings. In: *Interspeech 2010*.

Chen, Lei, and Harper, Mary P. 2009. Multimodal floor control shift detection. Pages 15–22 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Cristani, Marco, Pesarin, Anna, Drioli, Carlo, Tavano, Alessandro, Perina, Alessandro, and Murino, Vittorio. 2011. Generative modeling and classification of dialogs by a low-level turn-taking feature. *Pattern Recognition*, **44**(8), 1785 – 1800.

Dai, Peng, Di, Huijun, Dong, Ligeng, Tao, Linmi, and Xu, Guangyou. 2009. Group Interaction Analysis in Dynamic Context. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, **39**(1), 34–42.

de Kok, Iwan, and Heylen, Dirk. 2009. Multimodal end-of-turn prediction in multi-party meetings. Pages 91–98 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Debras, C., and Cienki, A. 2012. Some Uses of Head Tilts and Shoulder Shrugs during Human Interaction, and Their Relation to Stancetaking. Pages 932–937 of: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*.

Do, T., and Gatica-Perez, D. 2011. GroupUs: Smartphone Proximity Data and Human Interaction Type Mining. In: *IEEE Int.Symp. on Wearable Computers (ISWC), San Francisco*.

Dong, Wen, and Pentland, Alex "Sandy". 2010. Quantifying group problem solving with stochastic analysis. Pages 40:1–40:4 of: *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. ICMI-MLMI '10. New York, NY, USA: ACM.

Dong, Wen, Lepri, Bruno, and Pentland, Alex. 2012a. Automatic Prediction Of Small Group Performance In Information Sharing Tasks. *CoRR*, **abs/1204.3698**.

Dong, Wen, Lepri, B., Kim, Taemie, Pianesi, F., and Pentland, A.S. 2012b. Modeling conversational dynamics and performance in a Social Dilemma task. Pages 1–4 of: *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*.

Dong, Wen, Lepri, B., Pianesi, F., and Pentland, A. 2013. Modeling Functional Roles Dynamics in Small Group Interactions. *Multimedia, IEEE Transactions on*, **15**(1), 83–95.

Dunbar, N. E., and Burgoon, J. K. 2005. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, **22**(2), 207–233.

Escalera, Sergio, Pujol, Oriol, Radeva, Petia, Vitri?, Jordi, and Anguera, M. Teresa. 2010. Automatic Detection of Dominance and Expected Interest. *EURASIP Journal on Advances in Signal Processing*, **2010**.

Favre, Sarah, Dielmann, Alfred, and Vinciarelli, Alessandro. 2009. Automatic role recognition in multiparty recordings using social networks and probabilistic sequential models. Pages 585–588 of: *Proceedings of the 17th ACM international conference on Multimedia*. MM '09. New York, NY, USA: ACM.

Feese, S., Arnrich, B., Troster, G., Meyer, B., and Jonas, K. 2012. Quantifying Behavioral Mimicry by Automatic Detection of Nonverbal Cues from Body Motionc. Pages 520–525 of: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*.

Gatica-Perez, D. 2006. Analyzing Group Interactions in Conversations: a Review. Pages 41–46 of: *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*.

Gatica-Perez, D. 2009. Automatic Nonverbal Analysis of Social Interaction in Small Groups: a Review. *Image and Vision Computing, Special Issue on Human Behavior*, **27**(12), 1775–1787.

Gatica-Perez, D., op den Akken, R., and Heylen, D. 2012. Multimodal Analysis of Small-Group Conversational Dynamics. In: Renals, S., Bourlard, H., Carletta, J., and Popescu-Belis, A. (eds), *Multimodal Signal Processing: Human Interactions in Meetings*. Cambridge University Press.

Germesin, Sebastian, and Wilson, Theresa. 2009. Agreement detection in multiparty conversation. Pages 7–14 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Glowinski, Donald, Coletta, Paolo, Volpe, Gualtiero, Camurri, Antonio, Chiorri, Carlo, and Schenone, Andrea. 2010. Multi-scale entropy analysis of dominance in social creative activities. Pages 1035–1038 of: *Proceedings of the international conference on Multimedia*. MM '10. New York, NY, USA: ACM.

Gorga, Sebastian, and Otsuka, Kazuhiro. 2010. Conversation Scene Analysis based on Dynamic Bayesian Network and Image-based Gaze Detection. In: *ICMI-MLMI 2010*.

Hadsell, R., Kira, Z., Wang, Wen, and Precoda, K. 2012. Unsupervised topic modeling for leader detection in spoken discourse. Pages 5113–5116 of: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*.

Hall, J. A., Coats, E. J., and Smith, L. 2005. Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological bulletin*, **131**(6), 898–924.

Hung, H., and Gatica-Perez, D. 2010. Estimating Cohesion in Small Groups

Using Audio-Visual Nonverbal Behavior. *Multimedia, IEEE Transactions on*, **12**(6), 563–575.

Hung, H., Jayagopi, D., Yeo, C., Friedland, G., Ba, S., Odobez, J-M, Ramchandran, K., Mirghafori, N., and Gatica-Perez, D. 2007 (Sep). Using audio and video features to classify the most dominant person in a group meeting. In: *IACM Int.Conf. on Multimedia (ACM MM)*.

Hung, Hayley, and Chittaranjan, Gokul. 2010. The idiap wolf corpus: exploring group behaviour in a competitive role-playing game. Pages 879–882 of: *Proceedings of the international conference on Multimedia*. MM '10. New York, NY, USA: ACM.

Hung, Hayley, Huang, Yan, Friedland, Gerald, and Gatica-Perez, Daniel. 2011. Estimating Dominance in Multi-Party Meetings Using Speaker Diarization. *IEEE Transactions on Audio, Speech & Language Processing*, **19**(4), 847–860.

Ishizuka, Kentaro, Araki, Shoko, Otsuka, Kazuhiro, Nakatani, Tomohiro, and Fujimoto, Masakiyo. 2009. A speaker diarization method based on the probabilistic fusion of audio-visual location information. Pages 55–62 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Jayagopi, D., Raducanu, B., and Gatica-Perez, D. 2009a. Characterizing conversational group dynamics using nonverbal behavior. In: *Icme*.

Jayagopi, D. B., and Gatica-Perez, D. 2010. Mining group nonverbal conversational patterns using probabilistic topic models. *IEEE Transactions on Multimedia*.

Jayagopi, D. B., Hung, H., Yeo, C., and Gatica-Perez, D. 2009b. Modeling Dominance in Group Conversations from Nonverbal Activity Cues. *IEEE Trans. on Audio, Speech, and Language Processing, Special Issue on Multimodal Processing for Speech-based Interactions,*, **17**(3), 501–513.

Jayagopi, D.B., and Gatica-Perez, D. 2009. Discovering group nonverbal conversational patterns with topics. In: *International Conference on Multimodal Interfaces*.

Jayagopi, Dineshbabu, Sanchez-Cortes, Dairazalia, Otsuka, Kazuhiro, Yamato, Junji, and Gatica-Perez, Daniel. 2012. Linking speaking and looking behavior patterns with group composition, perception, and performance. Pages 433–440 of: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ICMI '12. New York, NY, USA: ACM.

Kalimeri, Kyriaki, Lepri, Bruno, Kim, Taemie, Pianesi, Fabio, and Pentland, AlexSandy. 2011. Automatic Modeling of Dominance Effects Using Granger Causality. Pages 124–133 of: Salah, AlbertAli, and Lepri, Bruno (eds), *Human Behavior Understanding*. Lecture Notes in Computer Science, vol. 7065. Springer Berlin Heidelberg.

Kalimeri, Kyriaki, Lepri, Bruno, Aran, Oya, Jayagopi, Dinesh Babu, Gatica-Perez, Daniel, and Pianesi, Fabio. 2012. Modeling dominance effects on nonverbal behaviors using granger causality. Pages 23–26 of: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ICMI '12. New York, NY, USA: ACM.

Kim, S., Valente, F., and Vinciarelli, A. 2012a. Automatic detection of conflicts in spoken conversations: Ratings and analysis of broadcast political debates. Pages 5089–5092 of: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on.*

Kim, Samuel, Filippone, Maurizio, Valente, Fabio, and Vinciarelli, Alessandro. 2012b. Predicting the conflict level in television political debates: an approach based on crowdsourcing, nonverbal communication and gaussian processes. Pages 793–796 of: *Proceedings of the 20th ACM international conference on Multimedia.* MM '12. New York, NY, USA: ACM.

Kim, T., and Pentland, A. 2009. Understanding effects of feedback on group collaboration. *Association for the Advancement of Artificial Intelligence*, 25–30.

Knapp, M. L., and Hall, J. A. 2009. *Nonverbal Communication in Human Interaction.* 7 edn. Wadsworth Publishing.

Kumano, S., Otsuka, K., Mikami, D., and Yamato, J. 2011. Analyzing empathetic interactions based on the probabilistic modeling of the co-occurrence patterns of facial expressions in group meetings. Pages 43–50 of: *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.*

Kumano, Shiro, Otsuka, Kazuhiro, Mikami, Dan, and Yamato, Junji. 2009. Recognizing communicative facial expressions for discovering interpersonal emotions in group meetings. Pages 99–106 of: *Proceedings of the 2009 international conference on Multimodal interfaces.* ICMI-MLMI '09. New York, NY, USA: ACM.

La Fond, T., Roberts, D., Neville, J., Tyler, J., and Connaughton, S. 2012. The Impact of Communication Structure and Interpersonal Dependencies on Distributed Teams. Pages 558–565 of: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom).*

Lepri, B., Mana, N., Cappelletti, A., Pianesi, F., and Zancanaro, M. 2009a. Modeling the Personality of Participants During Group Interactions. Pages 114–125 of: *UMAP 2009.*

Lepri, Bruno, Mana, Nadia, Cappelletti, Alessandro, and Pianesi, Fabio. 2009b. Automatic prediction of individual performance from "thin slices" of social behavior. Pages 733–736 of: *Proceedings of the 17th ACM international conference on Multimedia.* MM '09. New York, NY, USA: ACM.

Lepri, Bruno, Subramanian, Ramanathan, Kalimeri, Kyriaki, Staiano, Jacopo, Pianesi, Fabio, and Sebe, Nicu. 2010a. Employing social gaze and speaking activity for automatic determination of the Extraversion trait. Pages 7:1–7:8 of: *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction.* ICMI-MLMI '10. New York, NY, USA: ACM.

Lepri, Bruno, Kalimeri, Kyriaki, and Pianesi, Fabio. 2010b. Honest Signals and Their Contribution to the Automatic Analysis of Personality Traits - A Comparative Study. Pages 140–150 of: Salah, AlbertAli, Gevers, Theo, Sebe, Nicu, and Vinciarelli, Alessandro (eds), *Human Behavior*

Understanding. Lecture Notes in Computer Science, vol. 6219. Springer Berlin Heidelberg.

Lepri, Bruno, Ramanathan, Subramanian, Kalimeri, Kyriaki, Staiano, Jacopo, Pianesi, Fabio, and Sebe, Nicu. 2012. Connecting meeting behavior with extraversion - a systematic study. *Affective Computing, IEEE Transactions on*, **3**(4), 443–455.

Nakano, Yukiko, and Fukuhara, Yuki. 2012. Estimating conversational dominance in multiparty interaction. Pages 77–84 of: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ICMI '12. New York, NY, USA: ACM.

Noulas, A., Englebienne, G., and Krose, B.J.A. 2012. Multimodal Speaker Diarization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **34**(1), 79–93.

Olguin, D. Olguin, Waber, B. N., Kim, T., Mohan, A., Ara, K., and Pentland, A. 2009. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, **39**(1), 43–55.

Otsuka, Kazuhiro, Araki, Shoko, Mikami, Dan, Ishizuka, Kentaro, Fujimoto, Masakiyo, and Yamato, Junji. 2009. Realtime meeting analysis and 3D meeting viewer based on omnidirectional multimodal sensors. Pages 219–220 of: *Proceedings of the 2009 international conference on Multimodal interfaces*. ICMI-MLMI '09. New York, NY, USA: ACM.

Otsuka, Y., and Inoue, T. 2012. Designing a conversation support system in dining together based on the investigation of actual party. Pages 1467–1472 of: *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*.

Pesarin, A, Cristani, M, Murino, V, and Vinciarelli, A. 2012. Conversation Analysis at Work: Detection of Conflict in Competitive Discussions through Automatic Turn-Organization Analysis. *Cognitive Processing*, **13**(2), 533–540.

Pianesi, F. 2013. Searching for Personality [Social Sciences]. *Signal Processing Magazine, IEEE*, **30**(1), 146–158.

Poggi, Isabella, and D'Errico, Francesca. 2010. Dominance Signals in Debates. Pages 163–174 of: Salah, AlbertAli, Gevers, Theo, Sebe, Nicu, and Vinciarelli, Alessandro (eds), *Human Behavior Understanding*. Lecture Notes in Computer Science, vol. 6219. Springer Berlin Heidelberg.

Prabhakar, Karthir, and Rehg, JamesM. 2012. Categorizing Turn-Taking Interactions. Pages 383–396 of: Fitzgibbon, Andrew, Lazebnik, Svetlana, Perona, Pietro, Sato, Yoichi, and Schmid, Cordelia (eds), *ECCV 2012*. Lecture Notes in Computer Science, vol. 7576. Springer Berlin Heidelberg.

Raducanu, B., and Gatica-Perez, D. 2009 (Apr). You are fired! nonverbal role analysis in competitive meetings. In: *Icassp*.

Raducanu, Bogdan, and Gatica-Perez, Daniel. 2010. Inferring competitive role patterns in reality TV show through nonverbal analysis. *Multimedia Tools and Applications*, 1–20.

Raiman, Nimrod, Hung, Hayley, and Englebienne, Gwenn. 2011. Move, and i will tell you who you are: detecting deceptive roles in low-quality data. Pages 201–204 of: *Proceedings of the 13th international conference on multimodal interfaces*. ICMI '11. New York, NY, USA: ACM.

Ramanathan, Vignesh, Yao, Bangpeng, and Fei-Fei, Li. 2013. Social Role Discovery in Human Events. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rehg, J. M., Fathi, A., and Hodgins, J. K. 2012. Social interactions: A first-person perspective. Pages 1226–1233 of: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society.

Rienks, R. J., and Heylen, D. 2005. Automatic dominance detection in meetings using easily detectable features. In: *Workshop Mach. Learn. for Multimodal Interaction (MLMI'05), Edinburgh, U.K.*

Salamin, H, and Vinciarelli, A. 2012. Automatic Role Recognition in Multiparty Conversations: an Approach Based on Turn Organization, Prosody and Conditional Random Fields. *IEEE Transactions on Multimedia*, **13**(2), 338–345.

Salamin, H, Favre, S, and Vinciarelli, A. 2009. Automatic Role Recognition in Multiparty Recordings: Using Social Affiliation Networks for Feature Extraction. *IEEE Transactions on Multimedia*, **11**(7), 1373–1380.

Salamin, Hugues, Vinciarelli, Alessandro, Truong, Khiet, and Mohammadi, Gelareh. 2010. Automatic role recognition based on conversational and prosodic behaviour. Pages 847–850 of: *Proceedings of the international conference on Multimedia*. MM '10. New York, NY, USA: ACM.

Sanchez-Cortes, D., Aran, O., Schmid-Mast, M., and Gatica-Perez, D. 2010. Identifying Emergent Leadership in Small Groups using Nonverbal Communicative Cues. In: *12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI'10), Bejing, China*.

Sanchez-Cortes, D., Aran, O., and Gatica-Perez, D. 2011 (Nov). An audio visual corpus for emergent leader analysis. In: *Icm-mlmi'11: workshop on multimodal corpora for machine learning: taking stock and road mapping the future*.

Sanchez-Cortes, D., Aran, O., Mast, M.S., and Gatica-Perez, D. 2012a. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *Multimedia, IEEE Transactions on*, **14**(3), 816–832.

Sanchez-Cortes, Dairazalia, Aran, Oya, Jayagopi, Dinesh Babu, Schmid Mast, Marianne, and Gatica-Perez, Daniel. 2012b. Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*.

Sapru, A., and Bourlard, Hervé. 2013. Automatic Social Role Recognition In Professional Meetings Using Conditional Random Fields. In: *Proceedings of Interspeech*.

Schoenenberg, K., Raake, A., and Skowronek, J. 2011. A conversation analytic approach to the prediction of leadership in two to six-party audio conferences. Pages 119–124 of: *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*.

Song, Yale, Morency, Louis-Philippe, and Davis, Randall. 2012. Multimodal human behavior analysis: learning correlation and interaction across modalities. Pages 27–30 of: *Proceedings of the 14th ACM international conference on Multimodal interaction*. ICMI '12. New York, NY, USA: ACM.

Staiano, J., Lepri, B., Kalimeri, K., Sebe, N., and Pianesi, F. 2011a. Contextual Modeling of Personality States' Dynamics in Face-to-Face Interactions. Pages 896–899 of: *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*.

Staiano, Jacopo, Lepri, Bruno, Ramanathan, Subramanian, Sebe, Nicu, and Pianesi, Fabio. 2011b. Automatic modeling of personality states in small group interactions. Pages 989–992 of: *Acm multimedia*.

Stein, R. T. 1975. Identifying Emergent Leaders from Verbal and Nonverbal Communications. *Personality and Social Psychology*, **32**(1), 125–135.

Subramanian, Ramanathan, Staiano, Jacopo, Kalimeri, Kyriaki, Sebe, Nicu, and Pianesi, Fabio. 2010. Putting the pieces together: multimodal analysis of social attention in meetings. Pages 659–662 of: *Proceedings of the international conference on Multimedia*. MM '10. New York, NY, USA: ACM.

Sumi, Yasuyuki, Yano, Masaharu, and Nishida, Toyoaki. 2010. Analysis environment of conversational structure with nonverbal multimodal data. Pages 44:1–44:4 of: *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. ICMI-MLMI '10. New York, NY, USA: ACM.

Suzuki, Noriko, Kamiya, Tosirou, Umata, Ichiro, Ito, Sadanori, Iwasawa, Shoichiro, Sakata, Mamiko, and Shimohara, Katsunori. 2013. Detection of Division of Labor in Multiparty Collaboration. Pages 362–371 of: *HCI (15)*.

Valente, F, and Vinciarelli, A. 2010. Improving Speech Processing trough Social Signals: Automatic Speaker Segmentation of Political Debates using Role based Turn-Taking Patterns. Pages 29–34 of: *Proceedings of the International Workshop on Social Signal Processing*.

Varni, G., Volpe, G., and Camurri, A. 2010. A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. *IEEE Transactions on MultiMedia*, **12**(6), 576–590.

Vinciarelli, A. 2009. Capturing Order in Social Interactions. *IEEE Signal Processing Magazine*, **26**, 133–152.

Vinciarelli, A, Salamin, H, Mohammadi, G, and Truong, K. 2011a. More than Words: Inference of Socially Relevant Information from Nonverbal Vocal Cues in Speech. Pages 24–34 of: *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*. LNCS, vol. 6456. Springer Verlag.

Vinciarelli, A., Valente, F., Yella, S.H., and Sapru, A. 2011b. Understanding social signals in multi-party conversations: Automatic recognition of socio-emotional roles in the AMI meeting corpus. Pages 374–379 of: *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on.*

Vinyals, Oriol, Bohus, Dan, and Caruana, Rich. 2012. Learning speaker, addressee and overlap detection models from multimodal streams. Pages 417–424 of: *Proceedings of the 14th ACM international conference on Multimodal interaction.* ICMI '12. New York, NY, USA: ACM.

Voit, Michael, and Stiefelhagen, Rainer. 2010. 3D user-perspective, voxel-based estimation of visual focus of attention in dynamic meeting scenarios. Pages 51:1–51:8 of: *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction.* ICMI-MLMI '10. New York, NY, USA: ACM.

Wang, Wen, Precoda, Kristin, Richey, Colleen, and Raymond, Geoffrey. 2011. Identifying Agreement/Disagreement in Conversational Speech: A Cross-Lingual Study. Pages 3093–3096 of: *INTERSPEECH.*

Wang, Wen, Precoda, K., Hadsell, R., Kira, Z., Richey, C., and Jiva, G. 2012. Detecting leadership and cohesion in spoken interactions. Pages 5105–5108 of: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on.*

Wilson, Theresa, and Hofer, Gregor. 2011. Using linguistic and vocal expressiveness in social role recognition. In: *Proceedings of the International Conference on Intelligent User Interfaces (IUI).*

Wöllmer, Martin, Eyben, Florian, Schuller, Björn W., and Rigoll, Gerhard. 2012. Temporal and Situational Context Modeling for Improved Dominance Recognition in Meetings. In: *INTERSPEECH.*

Woolley, Anita W., Chabris, Christopher F., Pentland, Alex, Hashmi, Nada, and Malone, Thomas W. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, **330**(6004), 686–688.