Identifying Dominant People in Meetings from Audio-Visual Sensors

Hayley Hung[1]
[1] IDIAP Research Institute
Switzerland
hhung@idiap.ch

Daniel Gatica-Perez[1,2]
[2] Ecole Polytechnique Federale de Lausanne (EPFL)
Switzerland
gatica@idiap.ch

Abstract

This paper provides an overview of the area of automated dominance estimation in group meetings. We describe research in social psychology and use this to explain the motivations behind suggested automated systems. With the growth in availability of conversational data captured in meeting rooms, it is possible to investigate how multi-sensor data allows us to characterize non-verbal behaviors that contribute towards dominance. We use an overview of our own work to address the challenges and opportunities in this area of research.

1. Introduction

Human group behavior is a complex and highly dynamic, time-varying process which defines our role and identity in a group through social interactions. An initial face-to-face encounter between unacquainted individuals commences immediately with an establishment of hierarchy between the interactants [15]. These encounters can be as subtle as non-verbal communication through eye-gaze with other participants, showing that establishing hierarchy is innate part of human behavior in social interactions.

Through the recent growth in meeting rooms equipped to capture multi-sensor data, it has become more plausible to study group dynamics in conversations and task-driven interactions. One key element of group dynamics is dominance. Being able to identify dominant behaviour in conversational settings could potentially allow us to analyze the effectiveness of teams or to search or browse meeting data. We believe there is a real need for automated systems for estimating dominance and this has led to an emergence of research which tries to cross the divide between social psychology, and machine learning and perceptual computing. As shall be explained later, dominance has been defined in many ways by social psychologists, making the automated estimation and evaluation of it more challenging.

This paper summarizes contributions to this challenging topic in both social psychology and automated dominance estimation. A study of major problems and solutions are discussed in more detail through an overview of our own work in this area. Specifically, we investigate the properties of different non-verbal audio-visual cues used individually or in combination. We also discuss a systematic study of how variations in human judgments can affect performance and discuss the importance of using natural meeting data.

For the remainder of this paper, Section 2 provides a summary of investigations in social psychology on defining and understanding dominant behavior; Section 3 describes work in automated dominance analysis; Section 4 presents an overview of our recent contributions to this field using audio, visual and audio-visual measures of activity; Section 5 discusses several remaining challenges and potentially promising solutions. We conclude in Section 6.

2. What Is Dominance?

Dominance has been studied in social psychology for several decades where psychologists have tried to define dominance or find indications of it. Dominance can be viewed as a personality characteristic, a person's status within a group or the power they have within it [10]. However, Dunbar and Burgoon [6] suggested that power, influence and dominance were not the same. They suggest that power is the "capacity to produce intended effects, and in particular, the ability to influence the behavior of another person...Because power is an ability...it is not always exercised...its magnitude may not be fully evident unless it is pitted against a counterforce of appropriate strength" (p. 208). On the other hand, "dominance is necessarily manifest. It refers to context and relationship-dependent interactional patterns in which one actors assertion of control is met by acquiescence from another" (p.208). This definition of dominance was defined by Rogers-Millar and Millar [14] who defined dominance as two separate control variables: 'one-up' to 'one-down' maneuvers. In addition, Dunbar and Burgoon suggest that dominance is a set of "expressive, relationally based communicative acts by which power is exerted and influence achieved" (p. 208).

This idea of assertion and acquiescence was also suggested by Dovidio and Ellyson who defined a visual dominance ratio [5] to infer the level of dominance of two individuals. This was based on the ratio of the proportion of time someone spent addressing the other person divided by the time they spent looking and listening to the other.

Studies have quantified the effect of different facets of non-verbal activity cues on a person's perceived dominance levels. Schmid Mast found through a meta-analysis of 40 articles spanning 5 decades, that dominance could be inferred and expressed through speaking time [10] much more for scenarios where leader roles were assigned. Later, Dunbar and Burgoon [6] conducted a study into decoding dominance through non-verbal cues which they categorized as vocalic and kinesic features, referring to speech (e.g. speaking time, loudness or energy, speaking rate, pitch vocal control or interruptions [17]) and gesture based cues (e.g. body movement, posture and elevation, facial expressions, gestures or eye gaze [5]) respectively.

In terms of a human's perception of dominance, social psychologists have shown that it is possible to do this either as a participant or an observer of the interaction [5], though there may be differences in perception [6]. This is particularly relevant to the evaluation of automated systems where manual (first or third party) annotations are required. Dunbar and Burgoon commented that "Perhaps coders' perception of dominance correspond more closely with objective measures of verbal and non-verbal dominance than those of participants themselves...However, the coders' observations are limited to the behaviors in a particular interaction, whereas participants are privy to the ongoing interaction that is part of a continuing relationship." [6] (pp. 228). More details on understanding dominance from a social psychology perspective can be found in [6, 3].

3. Automated Dominance Estimation

To our knowledge, Basu et al. [2] were the first to investigate influence in group discussions. Their approach treated verbal exchanges on a dyadic basis and modeled all group interactions in terms of Markov chains where the transitions depended on the influence that one participant could exert on another. In each discussion, two out of five participants were required to debate on a pre-specified topic for one minute before the floor was opened. A combination of manually and automatically extracted audio-visual features were used such as speaking status, turns, and visual activity patterns from skin-color blob-tracking.

Zhang et al. [18] proposed the team-player influence model (TPIM) which used only automatically extracted audio cues like speaking activity features from a microphone array and headset microphones, and also manual speech transcripts of the meetings for automatic topic analysis. They tested on 2.5 hours of meetings where discussions and monologues were encouraged using pre-defined discussion topics and an action agenda. The TPIM represents explicitly the states of the group and its influence on the state transitions of individuals using a two-layer dynamic Bayesian network (DBN) so that an influence parameter could be estimated for each participant. This was evaluated qualitatively by comparison with ground truth annotations where a proportionate ranking of dominance was distributed to each participant such that the total summed to 1. There was no systematic nor quantitative evaluation.

Concurrently, Rienks et al. [12] used audio cues to estimate dominance. They used more varied corpus consisting of 1.5 hours of audio-visual data of 8 meetings from the MultiModal Meeting Manager (M4) corpus (also used by [18]) and the Augmented Multi-party Interaction (AMI) corpus [4]. A support vector machine (SVM) was used to estimate the dominance of the participants which was ranked manually according to their perceived dominance by 10 annotators. The rankings were distributed into three bins which represented high, normal, and low perceived dominance. All the features were annotated manually and included non-verbal (e.g. speaking turns, speaking length and floor grabs) and verbal cues (e.g. number of words spoken, number of questions asked).

Soon after, Rienks et al. [13] conducted a comparative study of both [12] and [18]. They used the same three-point dominance scale created from absolute rankings but the annotations were provided by the meeting participants themselves. The same audio features in [12] were used again and the SVM model outperformed the TPIM. For the absolute dominance rankings, since the meetings lasted from 5 to 35 minutes, it is likely that longer meetings would be more difficult to annotate. This could lead to lower annotator confidence and an increased likelihood of variability in the annotations.

Otsuka et al. [11] used non-verbal cues based on automatically extracted gaze patterns, to explain pair-wise influence in group discussions. They used 10 minutes of conversational data of pre-defined topics collected from two 4-participant groups. The participants were asked to come to a conclusion on each topic after 5 minutes. There was no quantitative evaluation of their method.

The discussions above highlight four issues; (i) audio-visual feature extraction particularly from non-verbal cues, (ii) the nature of the data, (iii) the annotation and evaluation procedure iv and possible methods of modeling. Table 1 summarizes the differences between these works. For audio-visual feature extraction, we cover only non-verbal cues since there is much to study using simple automatically extracted features before moving onto features which are more computationally expensive to extract. While many different audio and visual feature extraction methods were used, there was no systematic study of the benefits of each feature for dominance.

The variety of corpora indicates that dominance can be inferred in both conversational and meeting environments.

Reference	Data	Features	Manual/Automatic	Dominance model	Static/Dynamic	Task
[2]	Debating games (2 hrs)	A,V	Automatic+Manual	IM	Dynamic	Predict influence
[18]	Scripted (M4) (2.5 hrs)	A	Automatic+Manual	TPIM	Dynamic	Predict influence
[12]	M4 and AMI (1.5 hrs)	A	Manual	SVM	Static	Dominance (high, normal, low)
[13]	M4 and AMI (1.5 hrs)	A	Manual	SVM and TPIM	Static,Dynamic	Dominance (high, normal, low)
[11]	Scripted (10 mins)	A	Manual	Bilateral influence/Influence balance	Static	Dominance
[8, 7, 9]	AMI (3-5 hrs)	A,V,A/V	Automatic	SVM, Max/Min	Static	Most and Least Dominant

Table 1. Summary of literature in automatic dominance estimation (A:Audio,V:Video,A/V:Audio-Visual).

We distinguish conversations and meetings since the latter can involve more than just debating which leads to more challenging data where people are able to move freely and may walk to items such as a slide screen or whiteboard. Also, the meeting length can affect participant behavior as in general, shorter discussions can lead to higher levels of engagement and observable behavior [11]. In real meetings, participants may not maintain such interest levels, leading to more subtle group dynamics [1].

In terms of the annotation, analyzing perceived or self-reported dominance levels is not straightforward due to the variability of human judgments. However, a full analysis of annotator variability would be useful to highlight ways in which automated dominance analysis could be solved more systematically. Our own work has tried to address these issues and the next section provides an overview of the work.

4. Overview of our work

In this section, we summarize our recent work [8, 7, 9] and highlight the main findings. We focus on non-verbal cues since we felt that it was important to study systematically the impact of simple audio and visual activity features on estimating dominance automatically, where larger amounts of audio or visual activity were found to indicate more dominant behavior [6].

We used a subset of the publicly available AMI meeting corpus [4], containing audio and visual data of 5 different teams of 4 participants who met on several occasions to complete a task through role-play. 12 meeting sessions were selected for our experiments, from which 59 non-overlapping 5-minute meeting segments were created. 21 annotators were grouped so the same 3 individuals annotated common segments, enabling a majority consensus.

For each segment, annotators were asked to rank the participants in order of dominance from 1 (resp. most) to 4 (resp. least). The annotators were not given a definition for dominance and provided their own in free-form on completion of the annotations; over half reported using speaking time or talkativeness as a cue. From the annotations, 34 (resp. 31) meetings had full agreement for all 3 annotators on the most (least) dominant person. The annotators reported their level of confidence about their annotations on a 7-point scale where 1 represented high confidence. The average annotator confidence was 1.74 and 2.11 for the most and least dominant person labels respectively, suggesting the increased difficulty of labeling the least dominant person (reflected also in the free-form descriptions). Further details can be found in [9].

4.1. Audio Features from Individual Microphones

Audio activity features were generated by extracting speech from individual headset microphones for each participant. From this signal, a binary and a real-valued speech signal were generated by firstly extracting the energy from the signal and then thresholding this to form a binary signal that represented speaking status as 1 (speaking) and 0 (non-speech). We used the total of the energy (TSE) and the speaking length (TSL) to represent audio activity for each participant. Derived audio features were also used to represent speech activity such as total speaker turns (TST) and total turns without short turns (which could be back-channels) (TSTwoBC). We also used the total number of successful interruptions (TSI). In addition, a histogram was created to characterize the distribution of the turn durations of each person in the meeting (SDHist) to capture the frequency of longer and shorter turns.

4.2. Audio Features from a Single Source

In addition to extracting audio features from individual headset microphones, we also experimented with different single-source scenarios where speaker diarization was applied to the signal to discover who the most dominant person was [7]. The task of speaker diarization is to identify speakers and when they spoke from a single source. The diarization method that we used involved applying an agglomerative clustering method which iteratively merged clusters according to a pair-wise Bayesian information criterion (BIC) score. Calculating the BIC score for each potential cluster pair is a time consuming process and through some faster pre-selection steps to prune the hypothesis space, the computation time could be decreased without serious degradation in performance. With these speed-based improvements, we extracted speaker diarization outputs using increasingly faster versions of the algorithm. We also performed robustness testing by studying different distant microphone sources with decreasing signal to noise ratio.

4.3. Video Features from Individual Cameras

Computationally efficient visual activity features were extracted by taking advantage of the features that are already computed for video compression. We were able to extract visual activity features taken from motion vectors

Figure 1. Example screen-shots from the close-view cameras.

and the residual coding bitrate from MPEG-4 video of each person using close-up cameras in the meeting, as shown in Figure 1. Then, 3 different visual activity features were generated that represented the average motion vector magnitude (Vector), residual coding bitrate (Residue) of the visual activity that could be not be associated with specific motion vectors, and the average of both features (Combo). More details about these features can be found in [8]. Again, a real-valued and a thresholded binary visual activity signal were extracted for each participant. Using the binary visual activity values, we accumulated the total visual activity (TVL). We also extracted total visual activity turns (TVT), where a turn is when someone is active for a continuous period. Finally, the visual activity turns were accumulated into a histogram of their durations (VDHist).

4.4. Unsupervised Dominance Estimation

Our initial experiments on dominance estimation hypothesized that dominant people move and talk more [6] so the person with the highest or lowest total feature value was selected as the most or least dominant person, respectively.

4.4.1 Audio Activity Cues

Using our audio cues we found the highest/lowest total value of each feature to indicate the most/least dominant person well. Table 2 shows a summary of the results. The best performing cue for each dominance task is highlighted in bold. It was interesting to observe that both the total speaking length (TSL) and total speaker turns without short turns (TWTwoBC) performed the best for both dominance tasks. There was a slight drop in performance for the least dominant person task, which could be an indication of the difficulty of identifying passive people. This is observed further in the difference in performance for TSE, which could indicate that noise levels in the energy signal is much higher for the passive participants compared to the more active ones. This difficulty in finding the least dominant person was also reflected in the self-reported annotator confidence. Another interesting observation was the marked improvement in performance of both dominance tasks when the shorter turns were removed from TST to form TSTwoBC indicating that the shorter turns are less correlated with dominance.

Features	Most Dom. Class. Acc.(%)	Least Dom. Class. Acc.(%)
TSL	**85.3**	**83.9**
TSE	82.4	67.7
TST	61.8	71.0
TSTwoBC	**85.3**	**83.9**
TSL(SDM)	77.0	not available
Random	25.0	

Table 2. Performance of audio cues for both dominance tasks using the unsupervised model. Results taken from [9, 7].

Dominance estimation from a single audio source. In addition to estimating the most dominant person from speaking activity levels extracted from headset microphones, we performed some experiments based on the assumption that there was only a single audio source in the meeting [7]. The resuls are shown in Table 2. The single sources were taken from a single distant microphone (SDM) located on either the table or ceiling of the meeting room. In addition, synthesized audio signals were created from performing a delay-sum on the individual microphones of which there were two types; headset and lapel. While the diarization error rate increased with a lower signal-to-noise ratio (SNR), there was not a clear decrease in performance for the dominance estimation task. This was also observed for the different diarization strategies that were used to decrease computation time.

4.4.2 Visual Activity Cues

The visual activity features were less effective for predicting dominance but performed surprisingly well. Similar to audio, the total visual activity length (TVL) and the visual activity turns (TVT), were the most effective single features for decoding dominance. A summary of the results are shown in Table 3 where TVT included longer turns. Similar to the audio activity features, there was a decrease in performance between the most and least dominant tasks but for the visual activity features, this decrease was more pronounced, highlighting that these features are less well correlated with less dominant behavior. Also, the TVT feature performed better than TVL alone for the least dominant person task, which highlights that the shorter turns are not discriminative. Overall, single audio features performed the best and reflects the findings in [10].

Features	Most Dom. Class. Acc.(%)	Least Dom. Class. Acc.(%)
TVL(Residue)	**76.5**	45.2
TVT(Combo)	**76.5**	**64.5**

Table 3. Results using visual cues and unsupervised model for both dominance tasks.

4.5. Feature Fusion for Dominance Estimation

In [9], we conducted experiments to observe how both the audio and video modalities affected the performance of the dominance estimation task using SVMs.

Audio Feature Fusion: Our results from fusing audio activity features only found that there was some complementary nature to the features which led to a 6% absolute increase in performance for the most-dominant person classification task as shown in Tables 4. Also, while the total speaker interruptions (TSI) did not perform so well as an individual feature, it appeared often as a good complementary feature for other audio cues. This is supported by [17] who stated that interruptions could be “a device for exercising power and control in conversation” and also by [16] since interruptions do not always correspond to dominant behavior but to an individual’s level of engagement. For the

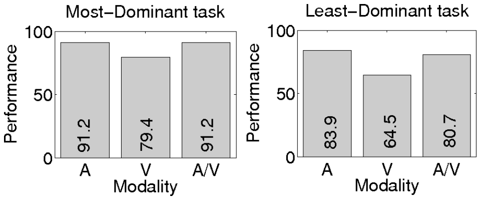

Figure 2. Comparison of the best results for each dominance task using (A)udio, (V)ideo, audio-visual (A/V) modalities from [9].

least dominant person task, we did not observe any increase in performance when audio features were combined.

Features	Class. Acc.(%)
TSL, TSE, TST	88.2
SDHist, TSE, TST, TSI	**91.2**
VDHist, TVL (Residue)	73.5
VDHist, TVT (Residue)	76.5
VDHist, TVL, TVT (Residue)	**79.4**
SDHist, TSE, TST, TSI, TVL	**91.2**
SDHist, TSE, TST, TSI, VDHist	**91.2**
SDHist, TSE, TST, TSI, VDHist, TVL	82.4

Table 4. Results from fusing audio, visual and audio-visual cues for most-dominant person with a supervised model (from [9]).

Video Feature Fusion: We conducted similar experiments to measure the complementary nature of the visual activity features. Results for estimating the most dominant person are shown in Table 4. After feature fusion, approximately 3% in improvement was possible. For the least dominant person task, fusing video features reduced performance.

Audio-Visual Feature Fusion: Selected results for estimating the most dominant person fusing the speech and visual activity features are shown in Table 4. The audio-visual feature combinations did not outperform the audio-only combinations for the most dominant person task. For the least dominant person task, audio-visual fusion did not improved over audio-only cues.

A summary of the best results for each dominance task using audio, visual and audio-visual cues are shown in Figure 2. The visual activity features performed worse than the audio features and also worst for each dominance task. Also, the audio-visual activity features could not outperform the audio-only features. Estimating the least dominant person was more difficult, resulting in lower performance in all cases, highlighting the increase in noise of the features for people who have a low activity levels.

4.6. Beyond Simple Single-Modality Activity Cues

So far, we have described our work using simple single-modality features where a person's non-verbal behavior could be characterized in terms of audio and/or visual activity levels. A brief observation of the annotators' free-form definitions of dominance found that the dominant person tended to receive more visual attention from the others when they spoke. This is similar to Dovidio and Ellyson's idea of the dyadic visual dominance ratio (VDR) [5] which is defined as the ratio of time spent looking while speaking over looking while listening. That is, dominant people tend to address the other more and listen to them less.

To use the VDR to infer dominance in larger groups, we redefined the VDR for multi-party conversations (MVDR) so the ratio quantifies the time each participant looks at others while speaking (TLWS) compared to the time they spend looking at other speakers (TLWL). We used human annotations of the visual focus of attention (VFOA) of each participant. The results are summarized in Table 5. Interestingly, the MVDR performed worse since the TLWL was not very discriminative compared to the TLWS which highlights again the problem of detecting passive behavior, or listening in this case.

Features	Most Dominant Person Class. Acc. (%)
MVDR	73.5
TLWS	**79.4**
TLWL	41.2

Table 5. Results for the dominant person task using the MVDR.

5. Challenges

Following the discussions and overview, several open issues remain. Our studies found that there were ambiguous cases where the most dominant person was estimated inaccurately because there was variability in the annotation of two of the participants. In such cases, dominant cliques may exist, which could be cooperative or competitive behavior, representing more subtle aspects of group hierarchy.

Experiments using a single distant microphone to estimate dominance showed that the performance was not particularly sensitive to the SNR. Distant cameras could also be used to extract visual features such as the VFOA in meetings where people could look at others and objects in the environment. However, estimating the VFOA robustly is challenging; the current performance is around 50% [1] in realistic meeting scenarios. VFOA can also be used to estimate when someone is addressing others or listening to someone. However, detecting passive behavior robustly, such as the act of listening, remains challenging. Also, while speaker interruptions have been addressed, the extraction method is crude and could be improved by analyzing the quality of the interruption, as suggested by Tannen [16].

Most of the work we presented have used static measures of dominance. For those which were dynamic, and used time-varying interactions to estimate influence, their performance tended to be worse [13]. This could be viewed as counter-intuitive since Millar and Millar [14] already defined dominance in terms of 'one-up' and 'one-down' interactions. While these described dyadic interactions, group exchanges may have a different dynamic where individuals can have influence on more than one person at a time. In addition, the proposed dynamic models encoded either dyadic interactions or group interactions but not both together. Identifying overall relative dominance in groups through dyadic relationships is difficult since the overall rankings could be cyclic.

The influence of situational factors should not be underestimated when observing the relationship between dominance and speaking time [10]. Recording natural data where individuals are strongly driven to dominate others for their own goals might be difficult. The AMI data captures natural meetings but the participants volunteered and did not have a vested interest in the outcome. Also corporate meetings may involve more than talking, e.g. the use of a whiteboard. Extracting contextual features about the meeting activities themselves would be challenging but could be beneficial.

6. Conclusion

While some work in the area of automated dominance estimation has relied on complex models, we have shown that in challenging meeting scenarios where the participants are able to behave naturally, simpler methods had superior performance. In addition, our detailed studies of the limitations with working with single modalities as well as the benefits of fusion shows what can already be achieved and where focus could be in the future.

The AMI meeting data used natural interactive activity but did not capture people who were extremely driven to attain their own goals or those of the team. Recording data which captures the everyday dynamic of employees in a company, for example, would provide a richer framework for analyzing dominant behavior in individuals and cliques. It remains to be seen whether this data could be captured accurately, while overcoming sensitivity to privacy.

Acknowledgments

This research was partly funded by US VACE, the EU project AMIDA , Swiss NCCR IM2, the German Academic Exchange Service (DAAD), and the Singapore (A*STAR). The work described is the result of ongoing collaborations with Dinesh Babu Jayagopi, Chuohao Yeo, Yan Huang, Gerald Friedland, Sileye Ba, and Jean-Marc Odobez to whom we thank.

References

[1] S. Ba and J.-M. Odobez. Multi-party focus of attention recognition in meetings from head pose and multi-modal contextual cues. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, Mar. 2008. 3, 5

[2] S. Basu, T. Choudhury, B. Clarkson, and A. Pentland. Learning human interactions with the influence model. In *NIPS*, 2001. 2, 3

[3] J. K. Burgoon and N. E. Dunbar. Nonverbal expressions of dominance and power in human relationships. In V. Manusov and M. Patterson, editors, *The Sage Handbook of Nonverbal Communication*. Sage, 2006. 2

[4] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraiij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, M. McCowan, W. Post, D. Reidsma, and P. Wellner. The ami meeting corpus: A pre-announcement. In *Proc. MLMI*, 2005. 2, 3

[5] J. F. Dovidio and S. L. Ellyson. Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly*, 45(2):106–113, June 1982. 2, 5

[6] N. E. Dunbar and J. K. Burgoon. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, 22(2):207–233, 2005. 1, 2, 3, 4

[7] H. Hung, Y. Huang, G. Friedland, and D. Gatica-Perez. Estimating the dominant person in multi-party conversations using speaker diarization strategies. In *International Conference on Acoustics, Speech and Signal Processing*, 2008. 3, 4

[8] H. Hung, D. Jayagopi, C. Yeo, G. Friedland, S. Ba, J.-M. Odobez, K. Ramchandran, N. Mirghafori, and D. Gatica-Perez. Using audio and video features to classify the most dominant person in a group meeting. In *ACM Multimedia*, 2007. 3, 4

[9] D. Jayagopi, H. Hung, C. Yeo, and D. GaticaPerez. Modeling dominance in group conversations from non-verbal activity cues. *Special issue on Multimedia in IEEE Transactions on Audio, Speech and Language Processing*. 3, 4, 5

[10] M. S. Mast. Dominance as expressed and inferred through speaking time. *Human Communication Research*, (3):420–450, July 2002. 1, 2, 4, 6

[11] K. Otsuka, J. Yamato, Y. Takemae, and H. Murase. Quantifying interpersonal influence in face-to-face conversations based on visual attention patterns. In *Proc. ACM CHI Extended Abstract*, Montreal, Apr. 2006. 2, 3

[12] R. Rienks and D. Heylen. Automatic dominance detection in meetings using easily detectable features. In *Proc. Workshop on Machine Learning for Multimodal Interaction (MLMI)*, Edinburgh, Jul. 2005. 2, 3

[13] R. Rienks, D. Zhang, D. Gatica-Perez, and W. Post. Detection and application of influence rankings in small group meetings. In *Proceedings of the 8th International Conference on Multimodal interfaces*. ACM Press, 2006. 2, 3, 5

[14] E. Rogers-Millar and F. M. III. Domineeringness and dominance: A transactional view. *Human Communication Research*, 5(3):238–246, 1979. 1, 5

[15] E. Rosa and A. Mazur. Incipient status in small groups. *Social Forces*, 58(1):18–37, September 1979. 1

[16] D. Tannen. *Gender and Discourse*, chapter Interpreting Interruption in Conversation, pages 53–83. Oxford Univesrity Press, 1993. 4, 5

[17] C. West and D. H. Zimmerman. *Language, Gender, and Society*, chapter Small Insults: A study of interruptions in cross-sex conversations between unaquainted persons, pages 103–117. Newbury House, 1983. 2, 4

[18] D. Zhang, D. Gatica-Perez, S. Bengio, and D. Roy. Learning influence among interacting Markov chains. In *NIPS*, 2005. 2, 3