YOU ARE FIRED! NONVERBAL ROLE ANALYSIS IN COMPETITIVE MEETINGS

*Bogdan Raducanu, Jordi Vitrià**

Computer Vision Center
Edifici O - Campus UAB
08193 Bellaterra - Barcelona
* Dept. of Applied Mathematics and Analysis
University of Barcelona
08003 Barcelona
Spain
E-mail: {bogdan,jordi}@cvc.uab.es

Daniel Gatica-Perez

Idiap Research Institute
École Polytechnique Fédérale de Lausanne
(EPFL)
Switzerland
E-mail: gatica@idiap.ch

ABSTRACT

This paper addresses the problem of social interaction analysis in competitive meetings, using nonverbal cues. For our study, we made use of ”The Apprentice” reality TV show, which features a competition for a real, highly paid corporate job. Our analysis is centered around two tasks regarding a person’s role in a meeting: predicting the person with the highest status and predicting the fired candidates. The current study was carried out using nonverbal audio cues. Results obtained from the analysis of a full season of the show, representing around 90 minutes of audio data, are very promising (up to 85.7% of accuracy in the first case and up to 92.8% in the second case). Our approach is based only on the nonverbal interaction dynamics during the meeting without relying on the spoken words.

Index Terms— Social interaction, competitive meetings, role analysis, nonverbal cues

1. INTRODUCTION

We witness an increasing interest of the computer science community in automatic analysis of social interaction. The understanding of fundamental principles that govern a person’s status in groups is of primary relevance for several social sciences and would pave the way to create tools to support research in social and organizational psychology [2, 9]. As stated in [11], social interaction can be addressed in two frameworks. One of them, comes from linguistics and addresses the problem of social interaction from the perspective of dialog understanding. The other one comes from the nonverbal communication interpretation perspective. Within this framework, speech prosody and body gestures are used in order to get hints about personal behavior. Facial expression, visual focus of attention, dialog structure, back-channels could provide powerful cues regarding engagement (interest level), persuasion, mirroring, dominance, etc.

The automatic analysis of group interactions has mainly focused on informal meetings [8, 3, 14, 10]. In some cases [8], meetings follow a scenario and so people behave in a somewhat controlled manner. In other cases, meetings are task-oriented [12] or driven by a topic of discussion [10], but the implicit degree of antagonism or controversy is not very high, thus resulting in essentially non-competitive conversations. Political debates [5] are example of competitive discussions, under fairly controlled conditions.

In this paper we add a novel dimension to the automatic analysis of social interactions, by studying the openly competitive scenario. More concretely, we address the problem of role analysis in competitive meetings using nonverbal cues. ”The Apprentice” US TV series offers an attractive scenario for our purpose. ”The Apprentice” is a NBC reality television show hosted by magnate Donald Trump [15]. Dubbed as ”The Ultimate Job Interview”, the show features sixteen to eighteen business people participating in an elimination-style competition for a one-year, $250,000 salary to run one of Trump’s companies. The winner of the competition is called ”The Apprentice”. The show represents a unique data set for the study of social interaction in competitive meetings. Being a reality-show, the behavior and reactions of the participants are natural, displaying a high degree of involvement. Participants are real people (not actors), fighting for a real goal. In the elimination process, nothing is arranged before-hand - everything happens ”now” - and the outcome is not known a priori.

The paper is organized as follows. In section 2, we present the data set. Section 3 explains the method used for nonverbal audio cue extraction needed for our analysis. In Section 4 we define the research tasks. Section 5 is dedicated to the presentation of the experimental results followed by some discussion. Finally, in section 6, we draw our conclusions and offer the guidelines for future work.

2. ”THE APPRENTICE” DATA SET

The show has a season-based periodicity with the first season being broadcasted in 2004, and the last one finished in spring 2008. Each season starts with a group of candidates having different backgrounds, including real estate, political consulting, sales, management, and marketing. People are placed in two teams, and each week they are assigned a task to be performed and asked to select a project manager for the task. The decision of what team wins/loses is made based on the teams’ performance with respect to the task assigned. The winning team receives a reward, while the losing team faces a ”boardroom showdown” in order to determine which team member should be fired (eliminated from the show). Elimination proceeds in two stages. In the first one, all of the losing team’s members are confronted. The project manager is asked to select some of the team members who are believed to be most responsible for the loss. In the second stage, which takes place in the boardroom meeting, the rest of the team is dismissed, and the project manager and the selected members face a final confrontation in which at least one is fired by Trump at the end of the meeting. So, on one side we have the ’candidates board’ and on the other side we have the ’executive board’. The ’executive board’ is formed by Trump together with other per-

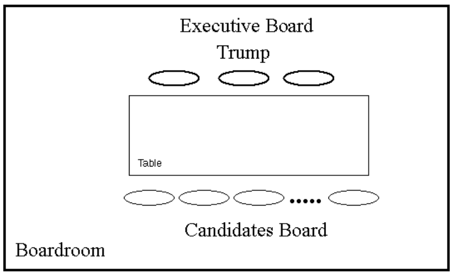

Fig. 1. ”The Apprentice” boardroom meeting scenario

sons (usually two) which will help him to make the decision of who from the losing team gets fired. Figure 1 presents a sketch of this scenario.

The data collected for our study belongs to the 6th season of the show, which took place during 14 weeks between January 7th - April 22nd, 2007. The number of initial candidates is 18. The following assumptions have been taken based on the exceptions mentioned below:

- in episode 3 (third week), one of the candidates resigns. Although this is a voluntary act, we consider it as a firing;
- in episode 13, Trump made it clear from the beginning that there would be no winners or losers for the assigned task; for this reason, we removed it from our study;
- episode 14, the final one, consists of two stages: the ’semi-final’ and ’final’. In the ’semi-final’, two persons are chosen (from a total of 4) to become the finalists; in the final, the hired person will be declared. For this reason, we treat episode 14 as two separate meetings.

In conclusion, our data set is formed of 14 valid meetings. From each episode, we segmented the second stage of the elimination process (i.e. ”the boardroom meeting”). These meetings have a duration between 2.20 minutes and 9.5 minutes and the number of participants varies between 5 and 11. Overall, we processed around 90 minutes of audio data.

3. FEATURE EXTRACTION AND REPRESENTATION

The data we had access to was the TV broadcast, so we had only one audio channel available. Due to the recording conditions (strong background music for the whole duration of each meeting), for our study we decided to manually produce the speaker segmentation in order to assure an optimal analysis of the data. Speaking segments are a state vector indicating the status of a person (speaking/non-speaking). Afterwards, for each participant, we extracted automatically the speech features using the library developed at MIT [13]. Based on these speaking segments, we define two types of data that were used as meeting-wise descriptors.

One type is represented by the class of individual descriptors, that are person characteristic:

- TST - Total number of Speaking Turns: how many times a person takes the speaking turn during the meeting;
- TSI - Total number of Successful Interruptions: how many times a person successfully interrupts the others. ’A’ interrupts successfully ’B’ if ’A’ was talking when ’B’ started talking and ’A’ stopped talking before ’B’ does;

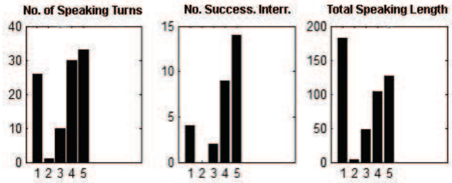

Fig. 2. Histogram of individual descriptors: TST, TSI and TSL, respectively in a meeting consisting of 5 participants. TSL is expressed in seconds

- TSL - Total Speaking Length: the total time a person speaks during the meeting.

In Figure 2, we depict the histogram of these descriptors corresponding to the the first episode. The number of persons who participated in this meeting is 5. These descriptors have been used with relative success also in [7].

The other type of data is formed by the class of relational descriptors, which characterize the interaction between persons:

- IM - Interruption Matrix. It contains the information regarding ’who interrupts who’ (column ’j’ interrupts line ’i’). Its size is NxN, N being the number of participants in the meeting;
- TTM - Turn Taking Matrix. It contains the information regarding ’who is talking after who’ (column ’j’ talks after line ’i’). It can be also roughly interpreted as ’who answers to who’. Its size is the same as IM.

In the case of individual descriptors, each of them represented also a measure for personal characterization. Following the same approach, we needed also to define an equivalent measure for the relational descriptors. Within social network analysis, a very common approach to assess a person’s position in a group is ’centrality’ [16]. From the graph theory perspective, if we consider the nodes corresponding to persons, then the arcs represents the ’strength’ of a person with respect with the other persons in the group (how a person relates to others). The reason we decided to use this measure in order to predict the ’fired’ person is because the person who feels his position is ’threatened’ try to become more involved. In other words, he/she might manifest a high degree of engagement in the meeting, by trying to persuade the others. Intuitively, we also expect that the person with the highest status tends to occupy a central position in the group.

Centrality measure can be expressed in several ways. We chose for our study the following definitions [6]:

- degree centrality: it is defined as the number of links incident upon a node (i.e., the number of ties that a node has). If the network is directed (meaning that ties have direction), then we usually define two separate measures of degree centrality, namely indegree (IC) and outdegree (OC). Indegree is a count of the number of ties directed to the node, and outdegree is the number of ties that the node directs to others.
- closeness centrality (CC): is a centrality measure of a node within a graph. Nodes that have short geodesic distances to other nodes in the graph have higher closeness. In the context of group meetings, we can say that the smaller the distance between people corresponds to higher interaction between them.

Fig. 3. Snapshots from the boardroom meeting; the right image corresponds to the moment when Trump announces the fired person

4. ROLE ANALYSIS

The particular character of this data set offers us the possibility to study and model roles in a group using nonverbal cues. More concrete, we want to study the following aspects regarding role analysis:

- which measures are good in predicting the person with highest status (Trump) (task T1);
- which measures are good in predicting the person who is fired (task T2).

More precisely, we would like to see which measures are most relevant for both cases. In Figure 3, we depict two instances from the show.

At this point, some clarifications are required. The relationship between status and other characteristics (like dominance, for instance) has been studied in-depth by psychological research [4]. Status is seen as a quality which implies respect and privilege, but not necessarily the ability to control others. On the other hand, dominance represents the quality to exert power and influence. Although they are different concepts, they are strongly inter-twinned: dominant persons usually occupy higher status in a group; the other way round, people with higher status tend to make use of their influence and power over their subordinates. A number of nonverbal cues have been found to be correlated with both status and dominance and some of them have been used to predict dominant people [7, 12].

5. EXPERIMENTAL RESULTS

We adopted a common framework for the analysis of both T1 and T2. based on the individual and relational descriptors.

For each of the types of measures previously defined, we performed a rank-based classification, taking into account the first two persons with highest measure values. This decision was motivated by the assumption that the person with the highest status and the fired person are the ones who interact the most. In case of the task T1, the following rule was established: ”the person with the highest status is the first ranked person”. Regarding T2, a similar rule was applied: ”consider the second-ranked position, except for those situations when this position is occupied by Trump, in which cases consider the first-ranked person”. This rule arises from the obvious fact that Trump cannot be fired. Although more than one person can be fired in the boardroom meeting, in our study we consider to have made a good prediction, if we are able to make one positive identification.

From this rank-based classification, and after applying the aforementioned rules, we generated the following tables containing the prediction accuracy. In Table 1, we present the values from the estimation based on the individual descriptors. We can appreciate that, in general, TSI and TSL are very good measures for both T1 and T2. In change, TST is a good measure only for T2. In Tables 2 and 3, we present the values from the estimation of relational descriptors IM and TTM, respectively. From these results, we could see that some of the centrality measures are much better in predicting T2 than T1. Between them, predicting T2 based on TTM is more reliable than on IM. From the results obtained so far, we could remark that centrality measures based on successful interruptions and turn taking descriptors seem to provide a good characterization of interaction dynamics. They contain implicit information that could be exploited for role analysis. The results we obtained come as a confirmation to the evidence according to which, 'thin-slices' of behavioral data, based exclusively on nonverbal cues, are enough in order to predict the outcome of an interaction [1].

Tasks	TST (%)	TSI (%)	TSL (%)
T1	50.0 (7/14)	85.7 (12/14)	64.2 (9/14)
T2	92.8 (13/14)	85.7 (12/14)	78.5 (11/14)

Table 1. Individual measures used to predict the highest-status person and the fired candidate

Tasks	IC (%)	OC (%)	CC (%)
T1	21.4 (3/14)	85.7 (12/14)	42.8 (6/14)
T2	78.5 (11/14)	71.4 (10/14)	64.2 (9/14)

Table 2. Relational measures based on the successful interruption matrix (IM) used to predict the highest-status person and the fired candidate

From the three tables, we observe a large variation in the performance of measures we used, which suggest that some measures are more suitable than others to characterize the addressed tasks. It is worth to notice that for the current case of competitive meetings, TSI seems to be much better measure than TSL for predicting the person with the highest status (note that the average percentage of overlapping speaking time is about 14.3%, which shows that interruptions seems to play an important role). This finding comes in contrast to previous research on non-competitive meetings [7], which found TSL to be the best measure to characterize high status. As a consequence of this observation, we performed an online analysis of predicting the person with the highest status based on the TSI measure, to see from which point during predictions were correct. The results of this analysis are depicted in Figure 4.

The horizontal axis corresponds to the percentage of accumulated meeting duration (this way we have a normalized representation of the data, irrespective to each meeting duration). We started our analysis at 30%. The vertical axis corresponds to the prediction accuracy (in percentage) a given stage. When 100% of the data has been processed, the curve converges towards the result shown in Table 1. From this figure we could appreciate that the person with the highest status can be identified in the early and late stages of the meeting (when the conclusions are made and the final decision is announced). The drop suffered by the curve (corresponding more or less to the middle of the meeting) might be explained by the fact that

Tasks	IC (%)	OC (%)	CC (%)
T1	57.1 (8/14)	64.2 (9/14)	42.8 (6/14)
T2	92.8 (13/14)	85.7 (12/14)	64.2 (9/14)

Table 3. Relational measures based on the turn taking matrix (TTM) used to predict the highest-status person and the fired candidate

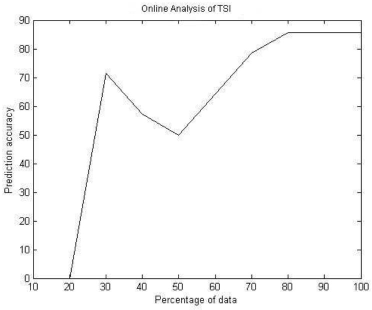

Fig. 4. Online analysis based on TSI for predicting the person with the highest status. See text for more details.

at some point during the debate, the person with the highest status 'passes on' the protagonism to the other participants and withdraws a bit. Starting with 60% of the meeting time, the curve recovers its ascending trend.

We would like to make some general considerations regarding the challenges we confronted in our work. One of the main limitations is represented by the reduced size of the current data set. Due to small number of meetings, we did not build a statistical model to be used for our study.

Another limitation is represented by the recording conditions. Being a TV-show, we had to adapt the analysis modality to the existing conditions. The only source of information that was consistent during the meeting and valid for analysis was the audio channel. The audio processing library proved to be very robust and the extracted speaking segments were not affected by noisy circumstances. It would have been very interesting to analyze also the visual channel. Unfortunately, the data was not consistent, since cameras were moving from one participant to the other. Extracting additional characteristics (like, for instance, the visual focus of attention), would have provided additional cues that cannot be studied for this data set.

6. CONCLUSIONS AND FUTURE WORK

In this paper we addressed the problem of role analysis in competitive meetings using non-verbal cues. Our study was based on the "The Apprentice" TV-reality show which offered an adequate data set. We addressed the following questions: (i) which measures are good to predict Trump as the person with the highest-status and (ii) which measures are good to predict the fired person. We presented two types of results: based on individual measures and relational measures. We found that our approach served as a good predictor for role analysis. Although we performed our analysis on a small data set, consisting of 14 meetings, the preliminary results obtained so far are promising. The methodology presented in this case-study would have to be validated in other types of competitive meetings to clarify whether the investigated features are good predictors of role-related behavioral outcomes'.

In the future we are planning to extend the results presented here, by incorporating data from other seasons of the show. Having more data available, it will also allow us to perform analysis based on statistical models. Another direction of research is represented by an extended and systematic study of the online processing of the data and the automatic speaker segmentation.

ACKNOWLEDGEMENTS

This work was done while B. Raducanu visited IDIAP as an AMIDA project trainee. D. Gatica-Perez thanks the support of the AMIDA and IM2 projects. B. Raducanu is also supported by MEC Grant TIN2006-15308-C02, CONSOLIDER-INGENIO 2010 (CSD2007-00018) and the Ramon y Cajal research program, Ministerio de Educación y Ciencia, Spain.

7. REFERENCES

[1] N. Ambady, F. Bernieri, and J. Richeson, "Towards a histology of social behavior: Judgmental accuracy from thin slices of behavior," in *Advances in Experimental Social Psychology*, P. Zanna, Ed., pp. 201–272. 2000.

[2] R.F. Bales, *Interaction Process Analysis: a method for the study of small groups*, Addison–Wesley, 1951.

[3] S. Basu, T. Choudhury, B. Clarkson, and A. Pentland, "Towards measuring human interactions in conversational settings," in *Proc. IEEE Int'l Conf. on Computer Vision, Workshop on Cues in Communication (CVPR-CUES)*, Kauai, Hawaii, USA, December 2001.

[4] J.K. Burgoon and N.E. Dunbar, "Nonverbal expressions of dominance and power in human relationships," in *The Sage Handbook of Nonverbal Communication*, V. et al. Manusov, Ed., pp. 279–297. Sage, 2006.

[5] S.W. Gregory Jr. and T.J. Gallagher, "Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes," *Social Psychology Quarterly*, vol. 65, no. 3, pp. 298–308, 2002.

[6] R. A. Hanneman and M. Riddle, *Introduction to social network methods*, Riverside, CA: University of California, Riverside, 2005. Retrieved from http://faculty.ucr.edu/˜hanneman/.

[7] D. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez, "Modeling dominance in group conversations from nonverbal activity cues," *To appear in: IEEE Trans. on Audio, Speech and Language Processing, Special Issue on Multimodal Processing for Speech-based Interactions*, January 2009.

[8] I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang, "Automatic analysis of multimodal group actions in meetings," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 27, no. 3, pp. 305–317, 2005.

[9] J.E. McGrath, *Groups: Interaction and Performance*, Prentice Hall, 1984.

[10] K. Otsuka, H. Sawada, and J. Yamato, "Automatic inference of cross-modal nonverbal interactions in multiparty conversations," in *Proc. ACM 9th Int'l Conf. on Multimodal Interfaces (ICMI)*, Nagoya, Japan, November 2007, pp. 255–262.

[11] A. Pentland and A. Madan, "Perception of social interest," in *Proc. IEEE Intl. Conf. on Computer Vision, Workshop on Modeling People and Human Interaction (ICCV-PHI)*, Beijing, China, October 2005.

[12] R. Rienks, D. Zhang, D. Gatica-Perez, and W. Post, "Detection and application of influence rankings in small group meetings," in *Proc. ACM 8th Int'l. Conf. on Multimodal Interfaces (ICMI)*, New York, US, November 2006, pp. 257–264.

[13] Speech feature extraction library, "http://groupmedia.media.mit.edu/" .

[14] R. Stiefelhagen, S. Chen, and J. Yang, "Capturing interactions in meetings using omnidirectional cameras," *Int'l. Journal of Distance Education Technologies*, vol. 3, no. 3, pp. 34–37, 2005.

[15] The Apprentice, "http://www.nbc.com/the_apprentice/" .

[16] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge University Press, 1994.