

Canal9: A Database of Political Debates for Analysis of Social Interactions

A. Vinciarelli, A. Dielmann, S. Favre and H. Salamin,
Idiap Research Institute, CP592 - 1920 Martigny (Switzerland)
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1050 Lausanne (Switzerland)
{vincia, adielman, sfavre, hsalamin}@idiap.ch

Abstract

Automatic analysis of social interactions attracts major attention in the computing community, but relatively few benchmarks are available to researchers active in the domain. This paper presents a new, publicly available, corpus of political debates including not only raw data, but a rich set of socially relevant annotations such as turn-taking (who speaks when and how much), agreement and disagreement between participants, and role played by people involved in each debate. The collection includes 70 debates for a total of 43 hours and 10 minutes of material.

1. Introduction

As automatic analysis of social interactions attracts increasingly more attention in the computing community [3] [8], publicly available benchmarks become a crucial element for the progress of the domain. Benchmarks allow different researchers to apply the same experimental protocols over the same data and this is the only way to perform rigorous comparisons between results achieved by different researchers and using different techniques.

This paper presents a corpus of political debates allowing the analysis of important social phenomena like roles (functional and social), conflicts, dominance, agreement and disagreement, status display, communication effectiveness, personality, persuasion, etc. From a social interaction analysis point of view, political debates represent an excellent resource for two important reasons:

- **Realism.** In contrast with most benchmarks (for example [1] [5]), political debates are real-world data. Debate participants do not *act* in a simulated social context, but participate in an event that has a major impact on their real life (for example, in terms of results at the elections). Thus, even if the debate format imposes some constraints, the participants are moved by real motivations leading to highly spontaneous social behavior.

- **Privacy issues.** Social interaction recordings are collected, in general, applying the *Informed Consent* principle [2]: subjects must know that they are recorded and must have the right of destroying, partially or totally, the data where they are portrayed. The result is that the subjects tend to be less spontaneous and to eliminate data showing attitudes they do not consider appropriate. As debates are public events, participants know that they are recorded (the principle is respected), but at the same time they are encouraged to be fully spontaneous because this is the only way to be successful in the debates. Furthermore, they cannot destroy the data because these are typically broadcasted live.

The corpus presented in this work includes 70 recordings for a total of 43 hours and 10 minutes of material. Each debate revolves around a *yes/no* question like “*Are you favorable to new laws on scientific research?*”. The participants state their answer (*yes* or *no*) at the beginning of the debate and do not change it during the discussion. Each debate involves a moderator that tries to give the same space to all participants (or at least to the two fronts corresponding to *yes* and *no* supporters). Furthermore, the moderator tends to reduce tensions when the discussion becomes too heated.

While including a rich set of annotations, the current version of the dataset is only a first release that will be further enriched in the years to come. Indeed, the Canal9 database is currently used in the core activities of the *Social Signal Processing Network (SSPNet)*, a European Network of Excellence aimed at studying Social Signal Processing, and further socially relevant annotations will be added in the framework of this project. The database (including the annotations) will be made publicly available through the web-portal of the SSPNet, at <http://www.sspnet.eu>. The data will be available to any academic and research institution upon signature of a End User Licence Agreement (EULA).

The rest of this article describes the data in terms of media format (Section 2), group composition (Section 3), duration statistics (Section 4), and available annotations (Sec-



Full Group



Personal Shot



Multiple Participants

Figure 1. Most frequent camera views.

tion 5).

2. Format and Structure

The recordings are available as high-quality full-frame (720×576 pixels) DV compressed PAL recordings, along with an uncompressed audio stream sampled at 48 kHz . They have been *live edited* and, in contrast with corpora collected in laboratory settings, not all the participants are visible all the time. All debates took place in the same recording studio (with no audience) and Figure 1 shows some of the most frequent camera views: *full group* (19.7% of data time), *personal shots* (66.1% of data time), and *multiple participants* (11.0% of data time). The remaining 3.2% corresponds to short reports (typically at the beginning of the debate) and credits shown at both beginning and end of each debate¹.

3. Group Composition

One of the most important aspects in any group of interacting individuals is the *composition*, that is number and type of people involved [4]. Political debates include two main roles: *moderator* and *participant*.

3.1. The Participants

Each debate revolves around a central question with a *yes/no* answer like “Are you favorable to the new laws on scientific research?”. Debate participants state explicitly their answer (*yes* or *no*) at the beginning of the discussion and this determines two factions expected to oppose one another during the entire discussion. The spatial arrangement of the participants reflects this situation (see *full group* view in Figure 1). The two factions physically oppose one another in a spatial arrangement that has been shown to elicit

¹The statistics have been extracted from a sample of 10 randomly selected debates.

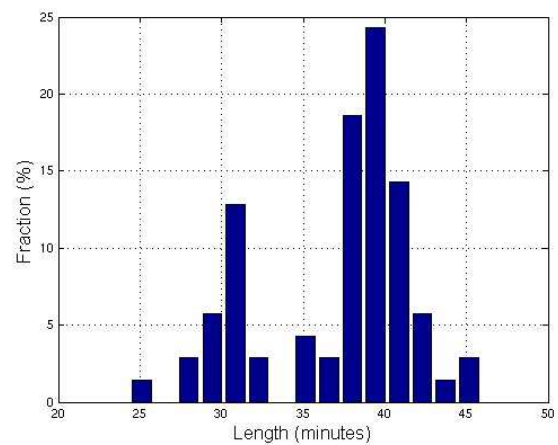


Figure 2. Length Distribution.

agreement between people on the same side and disagreement between people on opposite sides [6]. Overall there are 190 unique participants, 154 participate only in one debate, 25 participate in two debates, and the remaining 11 participate in three. In terms of gender, the set of the participants includes 25 women and 165 men.

3.2. The Moderator

All debates include one moderator expected to ensure that all participants have at disposition the same amount of time for expressing their opinion. Furthermore, the moderator intervenes whenever the debate becomes too heated and people tend to interrupt one another or to talk together. Overall, there are five different moderators, 1 woman and 4 men. The woman moderates 28 debates, while the men moderate 24, 9, 8 and 1 debates, respectively.

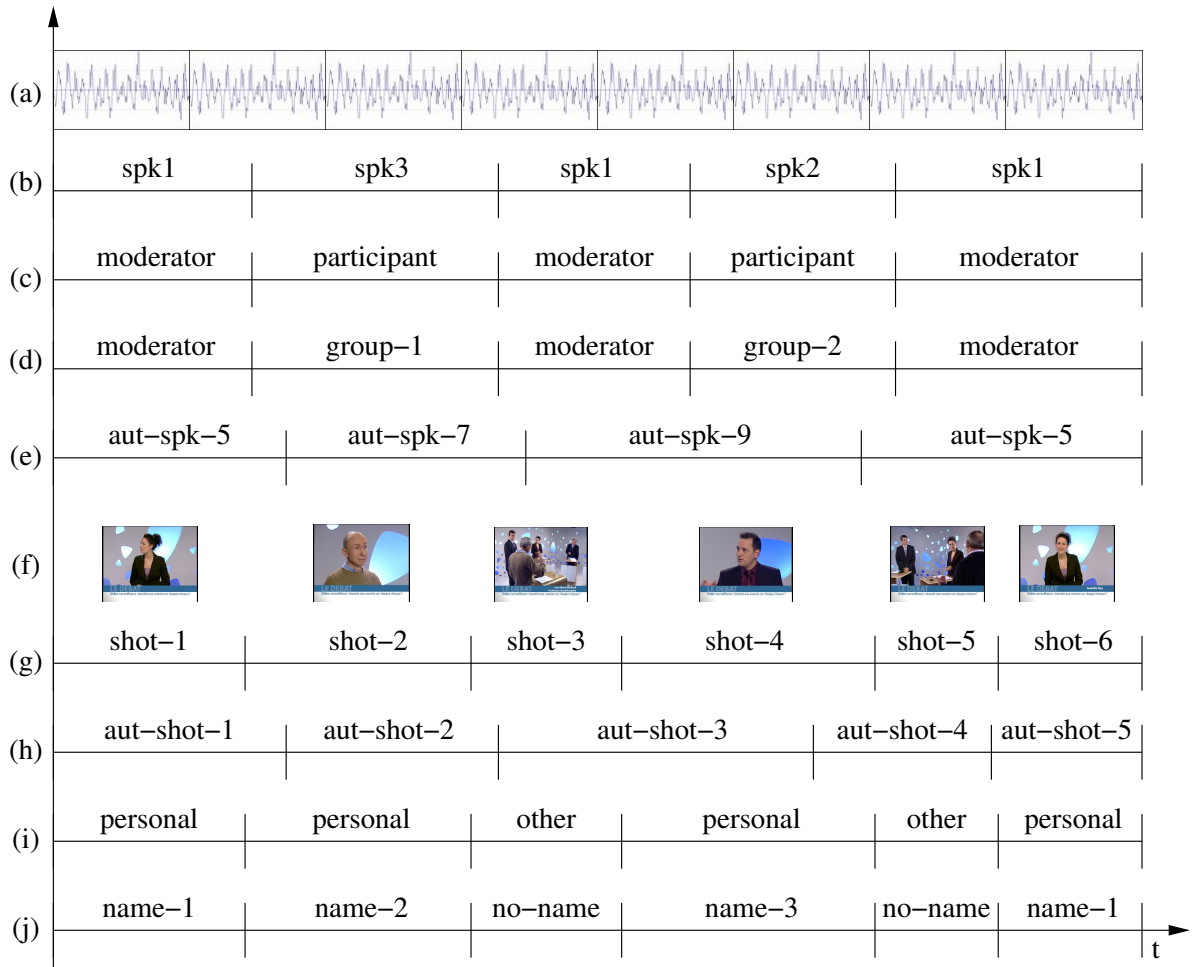


Figure 3. Annotations. The figure shows how the audio channel (a) is annotated in terms of manual speaker segmentation (b), role (c), agreement and disagreement (d), automatic speaker segmentation (e). Furthermore, the figure shows how the video channel (f) is annotated in terms of manual shot segmentation (g), automatic shot segmentation (h), manual shot classification (i), manual identification of participants in personal shots (j).

4. Duration Distribution

In total, the 70 debates of the corpus correspond to 43 hours, 10 minutes and 48 seconds. Of these, 41 hours 50 minutes and 40 seconds (96.9% of the total) correspond to actual discussions, while the remaining time includes reports and credits shown at beginning and end of each debate. The duration changes at each debate and the corresponding distribution is available in Figure 2.

5. Annotations

The political debates are correlated with a wide spectrum of annotations:

- **Manual Speaker Segmentation.** The audio of each debate (see Figure 3a) has been manually segmented into single speaker intervals (see Figure 3b). Speakers

are identified with a label that does not correspond to their names, and all the turns (single speaker segments) where the same person talks hold the same label. The segmentations are stored as `trs` files, an XML format used by the publicly available *transcriber* annotation tool².

- **Role.** The annotations report the *role* played by each person involved in the debates (see Figure 3c), i.e. *moderator* (the journalist expected to guarantee that all persons have enough time to express their opinion and that tries to inhibit aggressive and impolite behaviors) or *participant* (the persons that support one of the two answers to the question around which the debate revolves).

²Available at trans.sourceforge.net/en/presentation.php.

- **Agreement and Disagreement.** The participants (see point *Role*) are labeled in terms of *group-1* and *group-2* according to how they answer to the central question of the debate (see Figure 3d). Participants belonging to the same group agree with one another, while participants belonging to different groups disagree with one another.
- **Automatic Speaker Segmentation.** The output of an automatic speaker diarization system (see Figure 3e) is available for the audio channel of each debate. This allows one to perform experiments where the speaker segmentation is supposed to be performed automatically. Furthermore, the availability of both manual and automatic speaker segmentations allows one to estimate the effect of speaker segmentation errors. The segmentations are available as `trs` files (see *Manual Speaker Segmentation* point).
- **Manual Shot Segmentation.** The video channel of each debate (see Figure 3f) is manually segmented into shots (see Figure 3g), i.e. time intervals between two changes of camera. The shot segmentation is available as a list of shot boundaries, i.e. time instants where the camera changes. The boundaries are stored in ASCII files.
- **Automatic Shot Segmentation.** The output of an automatic shot segmentation system is available for the video channel of each debate (see Figure 3h). This allows one to perform experiments where the shot segmentation is expected to be performed automatically. The availability of both manual and automatic shot segmentations allows one to assess the effect of shot segmentation errors. The format of the automatic shot segmentations is the same as the one of the manual ones.
- **Manual Shot Classification.** Each shot is annotated in terms of two classes (see Figure 3i): *personal shot* (see Figure 1) and *other*. This allows one to identify those segments that are particularly suitable for behavior analysis as they clearly show a single person. No automatic classification is available.
- **Manual Identification of Participants in Personal Shots.** All personal shots showing a given *participant* are annotated with her/his identity (see Figure 3j). This allows one to select only those personal shots where a given participant appears. No automatic version of this annotation is available.

6. Conclusions

This paper has described the first release of the *Canal9* collection of political debates, a corpus aimed at the anal-

ysis of social phenomena taking place in competitive discussions. The corpus includes more than 40 hours of videos fully annotated in terms of a rich set of socially relevant features (turn-taking, agreement-disagreement, role) as well as low level descriptors (speaker segmentation, shot segmentation, identity of people appearing in personal shots, shot classes).

The corpus is publicly available through the web-portal of the *Social Signal Processing Network* (www.sspnet.eu) upon signature of an appropriate End User Licence Agreement. In its present form, the collection has been used in at least two works recently published in the literature [7] [9]. Further releases will be available in the next years and will include benchmarking procedures allowing rigorous comparisons of different results.

Acknowledgments. This work has been supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet). The authors wish to thank Canal9 for kindly allowing the diffusion of the data.

References

- [1] J. Carletta. Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation*, 41(2):181–190, 2007.
- [2] R. Faden, T. Beauchamp, and N. King. *A History and Theory of Informed Consent*. Oxford University Press, 1986.
- [3] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: a review. *Image and Vision Computing, to appear*, 2009.
- [4] J. Levine and R. Moreland. Small groups. In D. Gilbert and G. Lindzey, editors, *The handbook of social psychology*, volume 2, pages 415–469. Oxford University Press, 1998.
- [5] F. Pianesi, M. Zancanaro, E. Not, C. Leonardi, V. Falcon, and B. Lepri. A multimodal annotated corpus of consensus decision making meetings. *The Journal of Language Resources and Evaluation*, 41(3-4):409–429, 2008.
- [6] N. Russo. Connotation of seating arrangements. *The Cornell Journal of Social Relations*, 2:37–44, 1967.
- [7] A. Vinciarelli. Capturing order in social interactions. *IEEE Signal Processing Magazine, to appear*, 2009.
- [8] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing, to appear*, 2009.
- [9] A. Vinciarelli, H. Salamin, and M. Pantic. Social signal processing: Understanding social interaction through nonverbal behavior analysis. In *Proceedings of the International Workshop on Computer Vision and Pattern Recognition for Human Behavior*, 2009.