The KTH Games Corpora: How to Catch a Werewolf

Catharine Oertel[1], Giampiero Salvi[1], Jana Götze[1], Jens Edlund[1], Joakim Gustafson[1], and Mattias Heldner[2]

[1] KTH Royal Institute of Technology, Stockholm, Sweden
[2] Stockholm University, Stockholm, Sweden

Abstract. This paper introduces the Stockholm Werewolf Corpus, an eight-party, multimodal corpus. It is based on a role playing game called "Werewolves" and features speech and gaze annotations as well as electrodermal activity data. The corpus is also annotated for individual engagement as well as group involvement both by the participants in the game and external annotators. The corpus will be of interest to those who are interested in argumentative speech or deceptive speech as well as varied conversational dynamics.

Keywords: eight-party multimodal corpus, gaze, speech, electrodermal activity

1 Introduction

In the last decade, more and more multimodal corpora have been recorded so that nowadays many corpora are available to the public. One category of corpora is for example meeting corpora [1] another category are completely spontaneous corpora such as D64 [2] and Spontal [3]. While meeting corpora offer a higher amount of control over issues such as topic and social hierarchies within the group, they might not allow participants to relax as much and show as informal behavior as is typical of gatherings between friends. In contrast, completely spontaneous and unscripted corpora are hard to control for in terms of both topic and conversational dynamics.

With the Stockholm Werewolf Corpus we are trying to introduce a compromise between those two kinds of corpora with the aim to provide a more thorough understanding of conversational involvement. For this we aimed to record a multimodal corpus which is rich in conversational dynamics but also gives the subjects the possibility for self-reflection on their involvement and evaluate their peers and the group as a whole. One aspect we were also interested in was the physiological aspect of involvement therefore we included electrodermal activity (EDA) [4] of participants.

2 The Werewolf Corpus

The Stockholm Werewolf Corpus (SWC) is part of a bigger collection of game-based interactions called the KTH Games Corpora [5]. The SWC currently con-

Fig. 1. Illustration of the recordings

sists of a set of three rounds in which participants are engaged in playing the "Werewolf" game (see Section 2.1). One round varies between 15-20 minutes in length depending on the dynamics and involvement of the group of participants. Up until now a total of one hour of game sessions has been recorded but further recordings are underway.

Eight subjects participated in the recordings of which two were female and six male. There were four native Swedish speakers, one Hindi, one German, one Finnish and one Bulgarian. The recordings were carried out in English, in which all participants had a high proficiency. The degree of acquaintance prior to the recordings varied across participants, e.g. some were friends, some colleagues and some had never met each other before. Yet everybody knew at least two more participants before commencing with the recordings. Each participant took part in each session but with different roles in the game. Participants were seated around a table and were recorded by 4 cameras. Each camera captured two participants each (see Figure 1). The cameras were recording at a frame rate of 25 frames per second.

2.1 The Werewolf Game

The Werewolf Game is a role-playing game for eight or more participants, one of them takes on the role of game master who leads the participants through the different rounds of the game, each consisting of a night phase, in which the participants are asked to close their eyes, and a day phase. Each of the other participants is assigned a role that is only known to him/herself. Two participants are werewolves, all others are simple villagers.

Each of these two groups is trying to win the game by eliminating participants from the game. The werewolves by killing villagers during the night phase of the game, the villagers by trying to guess who is a werewolf and killing their choice during the day phase of the game. When a person is killed he is excluded from voting of who to kill and is not being counted in the number of remaining participants.

2.2 Annotations

Sessions of SWC have been annotated for both speech activity as well as eye-gaze. Eye-gaze annotations were carried out on a frame-by-frame basis. For each participant and for each frame in time the annotations indicate the target of the gaze. Possible targets are any of the other participants as well as other objects in the visual vicinity such as a sheet of paper with the game instructions. Other targets are "eyes closed", "looking down", and "looking at the sheet".

The SWC has also been annotated for group involvement. A distinction between the following four different classes was made: Class "high" is annotated when the group displays a high level of involvement. This can be compared to the more verbal involvement levels 7 to 9 in the annotation scheme described in [6]. Class "low" is annotated when the group displays a low level of involvement. This can be mapped to involvement levels 1 to 3 in [6]. Class "org" is annotated when the game leader is steering the conversation. This should be seen as a minimal category which is only chosen when participants are not actively taking initiative themselves. This category can be mapped to involvement level 4. Category "lead" is annotated when the group is forming itself. It is the period before the actual start of the game in which participants can get to know each other and questions can be asked. It was included in the corpus and subsequently also into the analysis in order to have an example of more non-task directed conversation. Class "lead" is only present in the first round of the game.

Group involvement annotations were carried out by three annotators. Each annotator annotated the corpus first separately. Afterwards annotations were compared and a consensus on which label to use was reached. There are only two instances in the corpus were no consensus could be reached (the 4th "org" in Round 1 and the first "low" in Round 3). In this case the annotation of the most experienced annotator was followed. In addition to the group involvement rating, participants in the game also rated themselves and each other in terms of individual engagement. In order to make these ratings as unobtrusive as possible, participants were asked to carry out the rating after each day phase.

In this study, we also wanted to explore the possibility of measuring EDA (electrodermal activity) as a physiological measure for individual engagement. For measuring this, we used a wearable EDA device, which exerts a direct current on the skin of the subject in order to measure skin conductance responses. For these measurements as well as the logging of the data the Q-Sensor developed by Affectiva was used. The measurements were taken from the fingertips of the subjects. The sampling rate was 8 Hz.

3 Corpus Analysis

In this section we present some descriptive statistics on the corpus. The gaze and speech annotations are used to generate the statistics. We consider the first sessions of Round 1 and Round 3 of the game because the game master was played by different participants.

The analysis in this section is performed on a frame-by-frame basis, where the frame rate is the same as the video frame rate (25 fps). Figure 2 shows cumulative statistics for each game round. The left plots show the fraction of the total number of frames in which each subject was gazing at each target. The first eight targets are the subjects themselves and the last four are other targets. The special target "closed" corresponds to the time when subjects are forced to close their eyes due to the game rules. Note also that the diagonal indicated by the dashed line is zero because each subject cannot gaze at him/herself.

The plot on the right is the fraction of frames in which each subject was speaking. It can be noted that BG in Round 1 and RM in Round 3 show different behaviour compared to the other subjects because they are the game masters in the corresponding rounds.

The ratios displayed in Figure 2 can be interpreted as frequentist estimation of probabilities. For the gaze data it is the probability that each subject would look at a particular target given the game round and for the speech data, the probability that each subject would speak in that round.

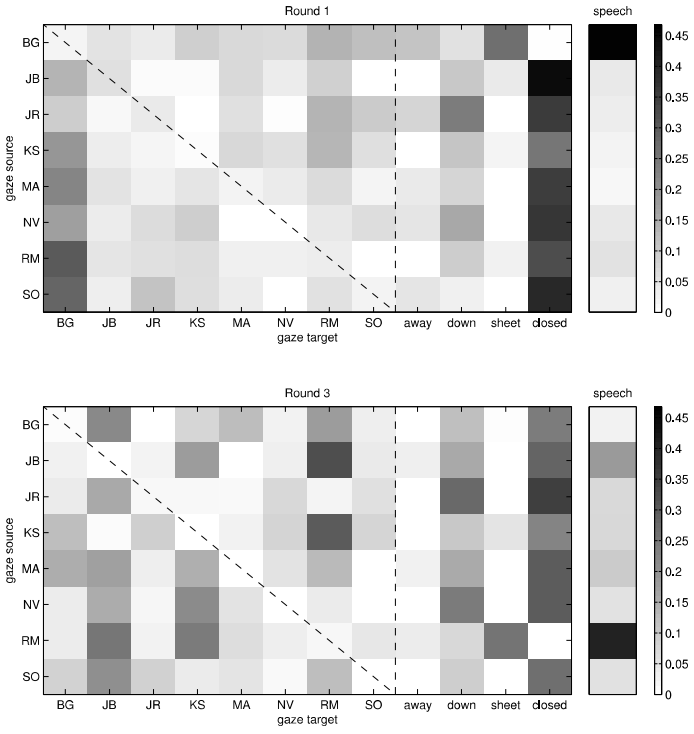

Fig. 2. Cumulative statistics for Round 1 and 3. Left: fraction of total number of frames in which each participant (y-axis) looked at each target (x-axis). Right ("speech"): fraction of frames in which each participant talked.

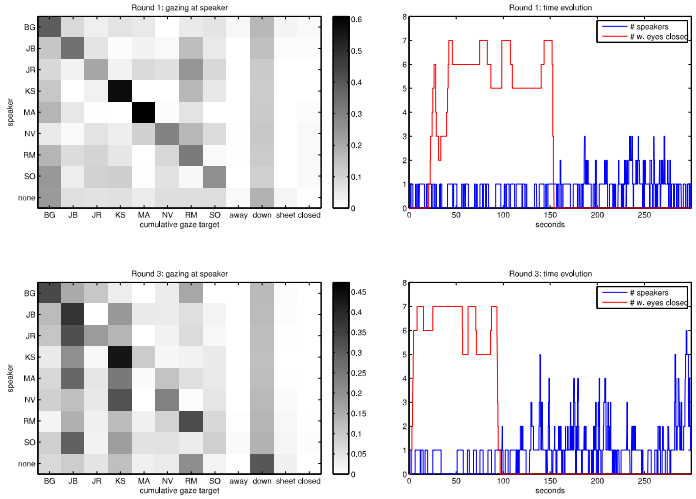

Fig. 3. Relating speech and gaze annotation. Left: fraction of frames in which participants were looking at a particular target (x-axis) when a particular participant was talking (y-axis). Right: number of simultaneous speakers and participants with the eyes closed per frame

Figure 3 displays the relation between speech and gaze data. The left plots show the fraction of frames in which the subjects are looking at particular targets when a specific subject is speaking. The right plots show how many speakers are talking simultaneously for each frame and therefore illustrate how much they overlap in the conversation. The red line shows how many subjects had their eyes closed because of the rules of the game. Those frames were removed when calculating the frequencies in the left plots.

In Round 1 we can observe that the subjects are principally looking at the person that is currently talking. There is a similar trend in Round 3, but it is less pronounced, probably because there is more overlap between the speakers. Also, in the frames when nobody is talking, the subjects tend to look at the game master (BG for Round 1 and RM for Round 3).

The numbers in the left plots of Figure 3 can be interpreted as probabilities of any subject looking at each target given that a specific subject is talking and given the game round.

4 Discussion and Conclusion

In this paper we gave an overview over the Stockholm Werewolf Corpus. We gave results on some preliminary analysis of the corpus in terms of gaze, speech and involvement annotations which are designed to understand group dynamics in spontaneous multiparty conversations.

To our knowledge only one further study [7] used the "Werewolf" game paradigm. The corpus in the current paper is different in that it a) introduces a new data stream namely electrodermal activity and b) is designed with the study of personal engagement and group involvement in mind, so that participants annotated engagement and involvement already during the duration of the game. It is also slightly different in the way the rules were applied. In [7] a time constraint was imposed on the participants to decide on which person to kill during the day phase. No such time constraint was imposed in the case of this study.

The Stockholm Werewolf Corpus is novel as it combines speech, gaze and EDA data as well as subjective engagement and group involvement annotations for the study of conversational dynamics. We tried to find a comprise between spontaneous, unrestricted, conversational corpora on the one hand and more controlled corpora on the other hand by using a game paradigm. The game "Werewolves" was found to be quite engaging and allows for a large number of people to contribute to the conversation at the same time. It provides the possibility to control for different roles without taking away from the spontaneity in the conversations of people. The corpus is of particular interest to those who are concerned with the study of conversational dynamics such as individual engagement and group involvement as well as argumentative and deceptive speech. We plan to release the corpus to the public over the course of the summer.

5 Acknowledgments

This work is supported by GetHomeSafe (EU 7th Framework STREP 288667).

References

1. Carletta, J.: Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus. Language Resources and Evaluation **4**(2) (2007) 181–190
2. Oertel, C., Cummins, F., Edlund, J., Wagner, P., Campbell, N.: D64: a corpus of richly recorded conversational interaction. Journal on Multimodal User Interfaces **7** (2013) 19–28
3. Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strombergsson, S., House, D.: Spontal: a swedish spontaneous dialogue corpus of audio, video and motion capture. In: Proc. of LREC 2010. Valetta. (2010) 2992 – 2995
4. Bouscein, W.: Electrodermal activity (2nd ed.). Springer, New York (2012)
5. Al Moubayed, S., Edlund, J., Gustafson, J.: Analysis of gaze and speech patterns in three-party quiz game interaction. In: To be published in Interspeech 2013. (2013)
6. Oertel, C., de Looze, C., Windmann, A., Wagner, P., Campbell, N.: Towards the automatic detection of involvement in conversation. In: Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues, Lecture Notes in Computer Science. (2011)
7. Hung, H., Chittaranjano, G.: The idiap wolf corpus: Exploring group behaviour in a competitive role-playing games. In: International Conference on Multimedia. (2010) 879–882