Using multimodal data to infer group dynamics in an adversarial group game

Lee Spitzley  
Center for the Management of Information  
The University of Arizona  
 lspitzley@cmi.arizona.edu

Steven Pentland  
Center for the Management of Information  
The University of Arizona  
 spentland@cmi.arizona.edu

Bradley Dorn  
Center for the Management of Information  
The University of Arizona  
 bdorn@cmi.arizona.edu

Bradley Walls  
Center for the Management of Information  
The University of Arizona  
 blwalls@cmi.arizona.edu

Judee Burgoon  
Center for the Management of Information  
The University of Arizona  
 judee@cmi.arizona.edu

Introduction

When military or security personnel interact with a new group of people, they are unsure of who in the group is reliable or trustworthy. The purpose of this study is to use multimodal sensor data (video, audio, language) to determine which members of a group are considered dominant, likeable, trustworthy, or deceptive. We developed a standardized game and environment to collect attitudinal and behavioral data. We collected data from 6 locations (3 in the United States and 3 in other countries). In this paper, we cover the game design, data collection, language processing, and next steps.

Keywords: multimodal data analysis, deception detection, sociocultural, experimental, linguistics, vocalics

Topic Numbers: 1, 2, 4, 8, 18

1. Abstract

When military or security personnel interact with a new group of people, they are unsure of who in the group is reliable or trustworthy. This challenge is compounded when interacting with groups from other cultures. The behaviors associated with dominance in an individualistic society may be very different from those in a collectivist society [1]. Reliance upon intuition when understanding group behavior opens the possibility of bias leading to incorrect decisions. Automated decision models, based on relational communication theories like Dyadic Power Theory (DPT; [2]) and Interpersonal Deception Theory (IDT; [3]), can use multimodal signals to make timely, automated inferences on the status of group members.

1.1. Game Design

We developed a modified version of the Mafia party game [4] for this study. The physical game setup is a portable assembly of tables, laptop computers, and video cameras. On the laptops is a custom game facilitation software (Figure 1) that allows players to vote during the game and answer survey questions between game rounds.

Figure 1: Custom Experiment Software

The laptop computers and video cameras capture video and audio data. See Figure 2 for the physical layout of the game environment.
We know ground truth of deception vs. truth by randomly assigning players as a “Spy” (deceptive) or “Villager” (truthful). The games last for one hour and consist of 5-8 players. Spies try to hide their identity and villagers try to discover the spies.

In each game round, the group first selects a mission leader. The group votes on leader selection through the game app. Once the group successfully choses a leader, he/she will choose a team to participate in a mission. The group then votes on the proposed mission team. The mission consists of a secret vote to “succeed” or “fail”. The villagers win the game if they have more successful missions than the spies have failed missions. Villagers always choose to succeed a mission, and spies must choose whether to succeed or fail a mission.

Our version of the game does not eliminate players. A facilitator uses a script for experimental control and to encourage group discussion. During the game, participants rate other members on dominance, likeability, and trustworthiness. The villagers indicate who they believe to be the spy.

Between game rounds, the participants rate the other players in the game on dimensions of liking, dominance, trust, and nervousness, and identify which players they believe are spies. These measures will be the dependent variables in our analysis.

1.2. Data Collection
We collected data from six locations (three in the United States, and one each in Israel, Singapore (See Figure 2), and Fiji). We have completed 70 games, with more than 450 participants, nearly 600 hours of audio/video recordings, and approximately 6TB of total data.

1.3. Data processing and decision algorithms
At the beginning of each round, the game facilitator plays a pre-recorded “ding” sound through a speaker. The microphones on all devices capture this sound, and we use this to synchronize the audio/video signal from all sources.

The data processing consists of several tasks: round segmentation, audio/video separation, audio transcription, speaker identification, audio segmentation, and joining with in-game data.

Audio/video separation is done using ffmpeg [5]. For audio transcription, the audio from each player’s laptop is sent to the IBM Watson Speech-to-text service [6]. Each microphone captures the speech from the player at the station, but they also capture speech from the facilitator and other players in the game. To merge the transcripts from 5-8 players (depending on game), we use the Recognizer Output Voting Error Reduction (ROVER, [7]). While originally intended to merge the results of multiple transcription algorithms over a single audio source, we use it to merge the transcripts from multiple audio sources. Research assistants then annotate speakers and game meta-data. The transcripts contain timestamp data to 1/100 of a second accuracy. The audio is segmented based on these timestamps and used for vocalic analysis.

1.4. Before the Conference
For the conference, we expect to have early results on linguistic and vocalic correlates of dominance, liking, trust, and deception. The language will be quantified using sentiment analysis and psychometric dictionaries (LIWC [8] and SPLICE [9]). We predict that deceivers will start games using low dominance in their language, and increase dominance as the game progresses, consistent with [10]. Deceivers may also speak using a similar linguistic style to the truthtellers [11].

2. References


